JMIR Infodemiology

Journal Impact Factor (JIF) (2023): 3.5
Volume 6 (2026) ISSN 2564-1891 Editor-in-Chief: Tim Ken Mackey, MAS, PhD

Contents

Viewpoints

Multilayered Epistemic Disruption in AI-Driven Health Misinformation: Conceptual Framework and Viewpoint (e96664)
Muzaffer Malkoç. . . . 6

The Information Void in Asymptomatic Chronic Disease: A Digital Health Framework for Understanding Social Media Health Information Seeking in Young Adults (e86489)
Victoria Ekstrom. . . . 19

Reviews

Online COVID-19 Misinformation and Infodemic Governance During the Peak Pandemic Period: Scoping Review of Social Media Research From 2020 to 2022 (e79434)
Yang Liu, Zhiying Yue. . . . 28

The Impact of Social Media Videos on Quantitative Health Outcomes: Systematic Review (e77578)
Fangyue Chen, Tricia Tay, Hannah Thould, Chatchaya Nangsue, Simon Dryden, Amish Acharya, Ara Darzi, Kate Grailey. . . . 52

Original Papers

Identifying Public Response Topics to the Centers for Disease Control and Prevention's COVID-19 Communications on Social Media: Infoveillance Study Using Large Language Model–Based Rephrasing (e98319)
Wangjiaxuan Xin, Shuhua Yin, Albert Park, Yaorong Ge, Shi Chen. . . . 67

Temporal Dynamics of Influenza-Associated Anxiety Symptom Linguistic Markers on Weibo (2023-2024): Observational Study (e88849)
Yifan Ou, Gert-Jan Bruijn, Peter Schulz. . . . 90

Mpox on Instagram: Content Analytic Study (e85379)
Elizabeth Havron, Kylee Chenault, Danny Valdez, Rebecca Houghton, Eric Walsh-Buhi. . . . 114

Tetanus Health Information on YouTube and TikTok: Cross-Sectional Analysis of Quality, Reliability, and Public Health Implications (e85397)
Rui Huang, Yajing Shen, Guangqing Zou, Jun Zhou, Xiaolu Deng, Qian Li, Xiaoxiong Chen. . . . 128

XSL•FO
RenderX

Understanding Vaccine Hesitancy in Louisiana Through Social Media Listening and Community Feedback: Cross-Sectional Study (e76827)
Angie Sanchez, Sadie Smith, Sahithya Sakhamuri, Julia Ardoin, Henry Chu. 140

Marketing Strategies and Factors Influencing the Popularity of Alcohol Videos from Official Brand Accounts on Douyin: Content Analysis Study (e74221)
Yuchen Zhao, Lingyun Zhang, Chenyu Qian, Wenjie Guo, Weiyun Zhu, Pinpin Zheng. 155

Using Artificial Intelligence Methods to Evaluate the Effect of the National Cytomegalovirus Awareness Month on the Content and Sentiment of Social Media Posts: Infodemiology Study (e80922)
Tracy Rosebrock, Zhen Yang, Lauren D'Arco, Tapan Pathak, Rebecca Vislay-Wade, Karen Fowler, John Diaz-Decaro, Colin Kunzweiler. . . . .
1 7 2

Automated Risk Assessment of Opioid Use: Analysis Using Pre-Trained Transformers on Social Media Data (e77783)
Muhammad Ahmad, Rita Orji, Maaz Amjad, Abubakar Siddique, Nailya Kubysheva, Ildar Batyrshin, Grigori Sidorov. 186

Understanding Social Support and Opinion Leaders in a Tuberculosis-Related Online Community in China: Content and Network Analyses (e79140)
Xiaojun Fan, Jueman Zhang, Xiuli Wang. 204

Evaluating Public Sentiment on Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder Compared With Other Mental Health Disorders From Posts on X (Formerly Known as Twitter): Longitudinal Analysis (e74440)
Vitor Tiosso Batistetti, Facundo Sanchez, Andrea Varaona, Francisco Lara-Abelenda, Mariana Pinto da Costa, Juan Chart-Pascual, Alberto Rodriguez-Quiroga, Javier Quintero, Miguel Alvarez-Mon. 223

Early Depression Detection in Social Media: Monitoring of Individual Nighttime Dynamics and Large Language Model Analysis (e87138)
Bicheng Yu, Zhichang Zhang, Lulu Ma, Jiongfu Cai, Yuanyuan Zhang. 238

Applying the Vertically Integrated Project Model to Youth-Inclusive Social Media Coding of Zyn-Related YouTube Shorts: Methodological Case Study (e95363)
Julia (Pengyue) Dou, Beth Hoffman, Piper Narendorf, Tatiana Limoncic, Grace Carver, Natalia Connor, Christine Larkin, Jaime Sidani. 255

Characterizing Family Abuse in Suicidal Ideation Posts on Reddit: Large Language Model–Assisted Content Analysis (e90340)
Salim Sazzed. 268

Monitoring Canada's Health Care Polycrisis Through News Headline Analysis: Infodemiology Study (e99655)
John Muteba Mwamba, Vijay Mago. 281

Parents' Multiperspective Views and Misinformation About COVID-19 Mitigation Measures During Tennessee School Board Meetings: Qualitative Content Analysis Using YouTube (e75691)
Olufunto Olusanya, Brianna White, Brenda Amuchi, Chad Melton, Arash Shaban-Nejad. 297

Implementation and User Evaluation of the SANGYAN Digital Health Platform to Enhance Knowledge About COVID-19 and Other Health Conditions: Quasi-Experimental Study (e67504)
Ashish Joshi, Ashruti Bhatt, Surapaneni Mohan, Gunjan Malhotra, Doilyn Oliveira, Saravanavel Revathi, Ashoo Grover. 309

Analyzing Misinformation and Disinformation: Understanding Swiss COVID-19 Narratives Through Natural Language Processing Analysis (e76441)
Federico Germani, Giovanni Spitale, Franc Fritschi, Sonja Merten, Nikola Biller-Andorno. 326

XSL•FO
RenderX

Misinformation Through the Lens of Dutch News Media and Public Health Authorities: Content Analysis of Topics and Narratives Present in Pandemic-Related Misinformation During COVID-19 (e91743)
David Blanco-Herrero, Bas van den Putte, Toni van der Meer. . . . 343

YouTube Videos as Educational Resources for Seizure Clusters and Treatment Options: Cross-Sectional Observational Study (e98768)
Taylor Elliott, Andrew Gienapp, James Wheless. . . . 360

Impact of Repeated Exposure to Polarized Health-Related News on Explicit and Implicit Attitudes Toward Dietary Supplements: Online Experimental Study (e88632)
Eugen-C Iin Secar , Nicolae-Adrian Opre. . . . 372

Quality, Reliability, and Dissemination of In Vitro Fertilization–Related Videos on Chinese Social Media: Cross-Sectional Analysis of 300 Short Videos (e83900)
Xueyan Bai, Feng Guo, Dapeng Chu. . . . 387

Japanese-Language AI Agent System for Human Papillomavirus Vaccine Infoveillance and Public Communication: Development and Feasibility Evaluation (e90295)
Junyu Liu, Siwen Yang, Dexiu Ma, Qian Niu, Zequn Zhang, Momoko Nagai-Tanima, Tomoki Aoyama. . . . 401

Review of the Quality and Reliability of Online Arabic Content on Diabetic Retinopathy: Infodemiological Study (e70514)
Abdullah Alabdrabulridha, Dalal Alabdulmohsen, Maryam AlNajjar, Ibtisam Algouf, Abdullah Al-Omair, Omaima Alyahya, Mesa Almahmudi, Abdulaziz Al Taisan. . . . 414

Shifting Public Discourse on Life-Sustaining Treatment in South Korea From 2014 to 2025: Longitudinal Social Media Analysis (e95918)
Miryoung Kim, Hae Suh. . . . 426

How Social Media Analysis Offers an Opportunity to Understand the Reality of People Living With Multiple Sclerosis: Descriptive French Study (e88763)
Emmanuelle Leray, Stéphane Schück, Pamela Voillot, Nathalie Texier. . . . 437

Investigating Online Discussions About Cancer Screening on Twitter (Subsequently Rebranded as X): Corpus Analysis (e90916)
Martin-Pieter Jansen, Hanneke Hendriks, Suzan Verberne, Gert-Jan de Bruijn, Enny Das. . . . 447

Exploring the Discourse Around Zyn Nicotine Pouches on Instagram and TikTok: Content Analysis (e88825)
Arpita Tripathi, Beth Hoffman, Christine Larkin, Piper Narendorf, Chaim Kittredge, Coltin Kunz, Jaime Sidani. . . . 464

Investigation of Differences in Tobacco Use Language Between Groups: Corpus-Assisted Analysis (e86593)
Iona Fitzpatrick, Xinmei Sun. . . . 478

Portrayals of Depression on TikTok: Content Analysis of Diagnostic Accuracy, Creator Type, and Stylistic Features (e85323)
Elena Rainer, Amber van der Wal, Ine Beyens. . . . 492

Nonnegotiable Symbolic Value and Sugar-Driven Food Habits in Indonesia: Mixed Methods Study Using a Digital Sociological Approach (e77261)
Ewina Manik, Sudarsono Hardjosoekarto, Ricardi Adnan, Radhiatmoko Radhiatmoko, One Herwantoko, Hakiki Nurmajesty, Darwan Darwan, Farrah Putri, Astuti Pawening. . . . 506

Leveraging AI for Analysis of Digital Health Information on Cancer Prevention Among Arab Youth and Adults: Content Analysis (e77888)
Alia Komsany, Obada Al Zoubi, Laetitia Sebaaly, Gabrielle Harrison, Orysya Soroka, Safa ElKefi, David Scales, Erica Phillips, Laura Pinheiro, Israa Ismail, Perla Chebli. . . . 525

XSL•FO
RenderX

Online Discussions of Men's Mental Health on Reddit and YouTube: Cross-Sectional Mixed-Methods Infodemiological Study (e81315)
Anurag Shekhar, Musawenkosi Saurombe. . . . 540

Measuring Safety Risks in Online Drug Sales: Empirical Study Based on Complex Network Theory (e86876)
Rong Jiang, Bingke Xing, Lixin Shu. . . . 563

Comparative Content Analysis and Expert Physician Risk Assessment of Aesthetic Procedures Promoted on TikTok in Taiwan: Quantitative Study (e88227)
Cheng-Jung Wu, Sheng-Yu Wu, Cheng-Yu Tsai, Arnab Majumdar, Jeffrey Yang, Jinn-Moon Yang, Lok-Yee Li. . . . 578

Emotional Expression and Mental Health Support in BTS Fandom Communities Using Natural Language Processing on YouTube Comments: Cross-Sectional Study (e74397)
Nari Yoo, Aaron Rodwin, Michael Park, Sangpil Youm, Sou Jang. . . . 586

Social Media Constructions of Alcohol as Emotional Coping and Social Connection in Ghana: Qualitative Study (e102546)
Rachel Thompson, Evershed Amuzu, Tricia Thompson, Mark Gyasi, Orlando Harris, Glenn-Milo Santos, Jerry Ouner. . . . 603

Social Media Reactions to Sex Toy Criminalization in Ghana: Cross-Sectional Sentiment and Text Network Analysis (e91422)
Anthony Agbeve, Celestin Niyomugabo, Celestin Hakizimana, Daniel Fiaveh, Amanda Ugboji, Caroline Pukall, Martina Anto-Ocrah. . . . 616

Social Media Perspectives on a Future HIV Vaccine: Mixed Methods Analysis (e82917)
Megan Rabin, Sarah Penuela-Wermers, Neil Sehgal, Teniola Egbe, Criswell Lavery, Sharath Guntuku, Alison Buttenheim. . . . 629

Online Information Behavior Regarding COVID-19 Vaccination and Its Association With Vaccination Behavior Based on Cluster Analysis of User Groups: Cross-Sectional Study (e82221)
Lea Liebner, Birgit Babitsch, Lisa Schmidt. . . . 640

Exploring Weight Loss Medication Discourse: Mixed Methods Analysis of US-Based Facebook Posts (e89732)
Elizabeth Dennard, Katrina Makres, Amrutha Alibilli, Vidur Jain, Anne Wang, Roger Abim-Karmon, Shaniece Criss, Yulin Hswen, Quynh Nguyen, Alexandria Ratzki-Leewing, Rozalina McCoy, Thu Nguyen. . . . 654

Online Information-Seeking Behavior Regarding Vitreous Floaters in the United States: Google Trends Analysis (e94821)
Thanaphat Seeboonruang, Qais Dihan, Jamie Surgent-Nahay, Muhammad Chauhan, Ahmed Shakarchi, Ahmed Sallam. . . . 671

Health Data for Linguistic Minority Group Research in Canada: Proof-of-Concept Centralized Health Care Metadata Repository Development and Usability Study (e77242)
Vincent Martin-Schreiber, Cayden Peixoto, Ricardo Batista, Christopher Belanger, Peter Tanuseputro, Amy Hsu, Lise Bjerre. . . . 679

Audience-Specific Health Communication: Mixed Methods Evaluation of the Maria Ciência AI-Assisted Knowledge Translation Tool (e78843)
Mariana Araújo-Pereira, Klauss Villalva-Serra Jr, Gustavo Pires-Ramos, Beatriz Sousa-Peres, Joanã Conceição-Oliveira, Sarah Maiche, Rebeca Silva, Bruno Andrade. . . . 688

Performance of Large Language Models in the Cognitive Analysis of Misinformation: Evaluation Study (e72524)
Dominika Wojtczak, Cheryl McQuire, Luisa Zuccolo, Claudia Peersman, Ryan McConville. . . . 706

XSL•FO
RenderX

Research Letters

Self-Reported Tianeptine Experiences on Reddit: Natural Language Processing–Assisted Qualitative Study (e86683)
Christopher Counts, Anthony Spadaro, Sahithi Lakamana, Abeed Sarker, Rachel Wightman, Jennifer Love, Diane Calello, Jeanmarie Perrone. 1 0 9

Cross-Sectional Evaluation of Medical Disinformation Safeguards in Consumer-Facing Large Language Model Platforms (e89831)
Natansh Modi, Cyril Alex, Abdulhalim Awaty, Bradley Menz, Stephen Bacchi, Kacper Gradon, Jessica Logan, Andrew Rowland, Lisa Kalisch Ellett, Ross McKinnon, Michael Wiese, Michael Sorich, Ashley Hopkins. 728

Characterizing Online Opinion on Vaping and E-Cigarettes in X (Formerly Twitter) Discourse in Singapore: Infodemiology Study (e103561)
Charles Alba. 733

Corrigenda and Addendas

Correction: Quality, Reliability, and Dissemination of In Vitro Fertilization–Related Videos on Chinese Social Media: Cross-Sectional Analysis of 300 Short Videos (e95354)
Xueyan Bai, Feng Guo, Dapeng Chu. 722

Correction: Japanese-Language AI Agent System for Human Papillomavirus Vaccine Infoveillance and Public Communication: Development and Feasibility Evaluation (e109982)
Junyu Liu, Siwen Yang, Dexiu Ma, Qian Niu, Zequn Zhang, Momoko Nagai-Tanima, Tomoki Aoyama. 723

XSL•FO
RenderX

Multilayered Epistemic Disruption in AI-Driven Health Misinformation: Conceptual Framework and Viewpoint

Muzaffer Malkoç, MA
Istanbul Medipol University, Kavacık Güney Kampüsü, Göztepe Mah., Beykoz, Istanbul, Türkiye

Corresponding Author:
Muzaffer Malkoç, MA
Istanbul Medipol University, Kavacık Güney Kampüsü, Göztepe Mah., Beykoz, Istanbul, Türkiye

Abstract

Generative AI has transformed the health information ecosystem by enabling scalable, sophisticated health misinformation production at near-zero marginal cost. Current literature addresses AI's role in health misinformation predominantly through a binary threat detection framework, systematically overlooking the structural, multilayered mechanisms through which AI simultaneously embeds false claims across intersecting human trust systems. This paper introduces the Multilayered Epistemic Disruption Framework (MEDF), which conceptualizes how AI-driven health misinformation structurally undermines public trust through four interdependent layers of cognitive and institutional disruption: discursive (clinical language shielding: fluent medical terminology and fabricated citations deployed as credibility signals), biometric (embodied authority transfer: deepfake appropriation of real clinicians' faces and voices), temporal (the synthetic chorus effect: near-simultaneous fabrication of apparently independent corroborating sources), and systemic (structural epistemic erosion: cumulative macro-level collapse of trust in medical institutions). Adopting a socioecological and structural epistemic approach, this viewpoint synthesizes empirical findings from communication psychology, medical sociology, and digital infodemiology, and the MEDF is explicitly positioned relative to established health communication frameworks, including the i-frame and s-frame distinction (individual-level vs system-level intervention targets) and socioecological infodemic models, with each construct's novelty defined in relation to adjacent concepts in prior literature. The MEDF proposes that AI-driven health misinformation is distinctively dangerous due to its capacity to exploit variable individual receptivity to medical authority claims and to simultaneously lower epistemic thresholds across multiple trust layers. Population-level data indicate that individuals who frequently encounter health misinformation on social media are 1.66 times more likely to report systemic distrust of health care institutions (odds ratio 1.66, 95% CI 1.11-2.48). Perceptual studies document that listeners correctly identify AI-generated voice clones only about 60% of the time and perceive a cloned voice as identical to its real counterpart in approximately 80% of trials. Existing defenses—including Content Provenance and Authenticity standards, automated deepfake detection (showing area under the curve drops of up to 50% under real-world conditions), and prebunking interventions—are shown to address only subsets of the proposed cascade, leaving temporal and systemic layers substantially unmitigated. Four testable hypotheses are advanced for empirical validation. Addressing AI-driven health misinformation requires moving beyond individual-level i-frame interventions toward structural, s-frame policy responses calibrated to each layer of the MEDF cascade. Policymakers and platforms must implement source identity verification, clinician biometric protection protocols, cross-platform ecosystem governance, and proactive trust infrastructure, with particular urgency in lower- and middle-income country contexts where regulatory capacity and platform oversight are most limited.

(JMIR Infodemiology 2026;6:e96664) doi:10.2196/96664

KEYWORDS

health misinformation; generative artificial intelligence; large language models; deepfakes; infodemiology; epistemic trust; health communication; conceptual framework; public health policy

Introduction: Epistemic Transformation and the Research Gap

The production and consumption of health information has historically been grounded in a robust institutional framework. Medical authority has been constructed through education, licensure, and professional accountability mechanisms, providing an epistemic assurance function for the credibility of health knowledge. The transformation of information ecosystems by algorithmic interfaces, accelerated by the proliferation of generative AI (GenAI), has brought this framework to a qualitative breaking point. The "infodemic" concept formulated by the World Health Organization during the COVID-19 pandemic foreshadowed this rupture; yet GenAI has fundamentally altered both the production logic and the scale of infodemics.

The existing literature largely addresses AI's role in health misinformation across two axes: threat vectors (hallucinations, deepfakes, and propaganda automation) and defensive

XSL•FO
RenderX

opportunities (early warning systems, real-time verification, and health literacy). These axes are sometimes consolidated under the rubric of a "dual role," emphasizing that AI can both generate and detect misinformation through the same technical infrastructure. A recent scoping review theorized this paradox through the concept of "epistemic ambivalence"—AI's potential to simultaneously construct and erode public knowledge [1]. However, this framing leaves a fundamental question unresolved: why is AI-driven health misinformation qualitatively distinct from other forms of misinformation?

This viewpoint's central argument is as follows: the danger of AI in the health information ecosystem lies not in generating false information per se, but in its capacity to embed that false information simultaneously across multiple layers of human trust systems. This capacity operates through clinical language shielding (false claims wrapped in fluent medical terminology and fabricated citations), embodied authority transfer (deepfaked clinician faces and voices carrying claims their real counterparts never made), the synthetic chorus effect (the near-instantaneous fabrication of apparently independent corroborating sources), and structural epistemic erosion (the cumulative collapse of trust in medical institutions). Each layer amplifies the preceding one, producing cumulative harm that exceeds the sum of individual layers.

This framework is specific to the health domain. In political or commercial misinformation, audiences are conditioned to approach sources with some degree of skepticism. In health information, the opposite dynamic obtains: health information seekers frequently exhibit heightened receptivity to claims presented with apparent medical authority—a term preferred over more reductive framings of uniform vulnerability, in line with current National Academies of Sciences, Engineering, and Medicine guidance [2]. This receptivity is not homogeneous: empirical research documents meaningful variation in individual susceptibility across populations and contexts. Chronic disease patients show direct associations between health-related social media use and cyberchondria—excessive anxiety and false health beliefs mediated by negative affect [3]. Heightened receptivity is further documented among older adults [4], individuals experiencing depression [5], and populations experiencing psychological distress or loneliness [6]. While institutional trust in medical authority provides a powerful baseline orientation for health information seeking, this trust is neither universal nor culturally invariant: cross-cultural research documents substantial variation in public deference to medical authority, structured primarily by societal power distance and cultural egalitarianism [7,8]. Under the structural and cultural conditions where such institutional trust does operate, AI systematically targets and exploits it.

This paper is presented as a viewpoint and framework piece. The framework is a conceptual proposal that has not yet undergone empirical validation; testable hypotheses have been developed for each layer and future research agendas identified. The aim of this paper is to introduce the Multilayered Epistemic Disruption Framework (MEDF), which conceptualizes how AI-driven health misinformation structurally undermines public trust through four interdependent layers of cognitive and institutional disruption. Moving beyond binary threat–opportunity framing, the MEDF maps how generative technologies alter the structural conditions of health communication and derives layer-specific policy interventions and empirical research agendas. The structure is as follows: the Core Constructs section provides a glossary of specialized cross-disciplinary terms (Table 1) and formal definitions of the four layers; the Theoretical Framework section addresses conceptual positioning and construct originality; four subsequent sections elaborate each layer in detail; the paper then evaluates existing defense mechanisms, presents policy recommendations, and closes with testable hypotheses, conclusions, and limitations.

XSL•FO
RenderX

Table . Working definitions of specialized cross-disciplinary terms used in this paper.

Term	Working definition
Machine heuristic	The cognitive shortcut by which users automatically attribute objectivity and accuracy to machine-generated content [9].
Truth-default bias	The default tendency to presume that communication is honest unless suspicion is actively triggered, recently extended to human responses to generative AI content [10].
Liar's dividend	The benefit accruing to dishonest actors when public awareness of deepfakes allows authentic evidence to be dismissed as fabricated [11].
Epistemic trust	In the sense of Fonagy and colleagues: an individual's willingness to accept new communicated information as trustworthy, generalizable, and relevant to the self [12].
i-frame/s-frame distinction	The contrast between interventions targeting individual cognition and behavior (i-frame) and interventions targeting the systemic rules and structures within which individuals act (s-frame) [13].
Need for cognitive closure	The motivated desire to reach a firm answer and avoid ambiguity, leading individuals to seize on early explanations and then freeze on them [14].
Dual coding theory	The theory that information encoded simultaneously through verbal and visual channels leaves stronger memory traces than single-channel encoding [15].
Illusory truth effect	The increase in the perceived truth of a statement produced by repeated exposure, largely independent of source credibility [16].
Epistemic ambivalence	Generative AI's paradoxical capacity to simultaneously construct and erode public knowledge [1].
Astroturfing and coordinated inauthentic behavior	The manual orchestration of ostensibly independent accounts and messages to simulate organic grassroots consensus, typically for political or commercial ends.

Core Constructs: Definitions and Architecture of the MEDF

This viewpoint draws on specialized concepts from clinical medicine, public health, informatics, and communication research. Table 1 provides working definitions of these cross-disciplinary terms together with their source literature; the four MEDF constructs themselves are formally defined in the remainder of this section.

The MEDF is formally defined as an analytic framework that conceptualizes AI-driven health misinformation not as a content problem—discrete false claims to be detected and corrected—but as a structural process operating simultaneously across four distinct layers of the trust systems through which people assess health information. Its unit of analysis is the information ecosystem rather than the individual message or user, and its purpose is twofold: to differentiate four mechanisms of epistemic disruption that require different points of intervention, and to articulate a testable hypothesis about how these mechanisms compound one another. The MEDF is not an empirically validated causal model, nor a detection instrument; it is a conceptual lens whose propositions are advanced for empirical evaluation. The four constructs are formally defined here, before their theoretical positioning and elaboration in subsequent sections. They are ordered analytically from micro-level linguistic manipulation to macro-level institutional erosion; this ordering reflects the proposed cascade logic, not an established temporal sequence.

1. Discursive layer—clinical language shielding: the mechanism by which inaccurate, unproven, or fabricated health claims are embedded within sophisticated medical terminology, pseudoclinical reasoning, and structurally plausible but distorted or fabricated citations. The professional linguistic frame itself functions as a deceptive credibility signal, bypassing the user's content-level evaluation capacity.
2. Biometric layer—embodied authority transfer: the digital hijacking and weaponization of a verified medical professional's physical identity—including voice clones, facial expressions, and somatic cues via deepfakes—to disseminate health claims the clinician never authorized or made.
3. Temporal layer—the synthetic chorus effect: the automated, near-instantaneous creation of a dense, cross-channel information infrastructure—including clone websites, fabricated patient testimonials, synthetic social media profiles, and video content—centered around a single false claim, creating a persuasive illusion of independent consensus within hours.
4. Systemic layer—structural epistemic erosion: the macro-level degradation of public trust in real medical authorities and institutions, occurring as the pervasive presence of hyper-realistic health deepfakes produces cognitive fatigue and chronic skepticism toward authentic health communication—what Chesney and Citron [11] termed the liar's dividend.

XSL•FO
RenderX

The structural logic of the MEDF posits a sequential cascade: clinical language shielding (layer 1) is hypothesized to lower baseline cognitive resistance, rendering the user more susceptible to biometric manipulation (layer 2), which is reinforced by fabricated consensus (layer 3), ultimately contributing to systemic institutional distrust (layer 4). This cascade, if empirically confirmed, would produce harm that is nonadditive. Each layer can cause harm independently; the cascade logic constitutes the framework's most contestable and most important empirical claim, elaborated in the Interlayer Relationships section below. Figure 1 illustrates the framework's structure, the sequential cascade between layers, and the cumulative reinforcement dynamic. Each layer is hypothesized to lower the user's resistance to the next (solid arrows), while their combined operation produces nonlinear, cumulative disruption (dashed reinforcement). Table 2 differentiates the four constructs by their primary trust target, core mechanism, theoretical basis, and intervention point.

Figure 1. The Multilayered Epistemic Disruption Framework (MEDF): a proposed cascade across four interdependent layers of health-information trust. The cascade is a theoretical proposition advanced for empirical testing and is not yet experimentally validated.

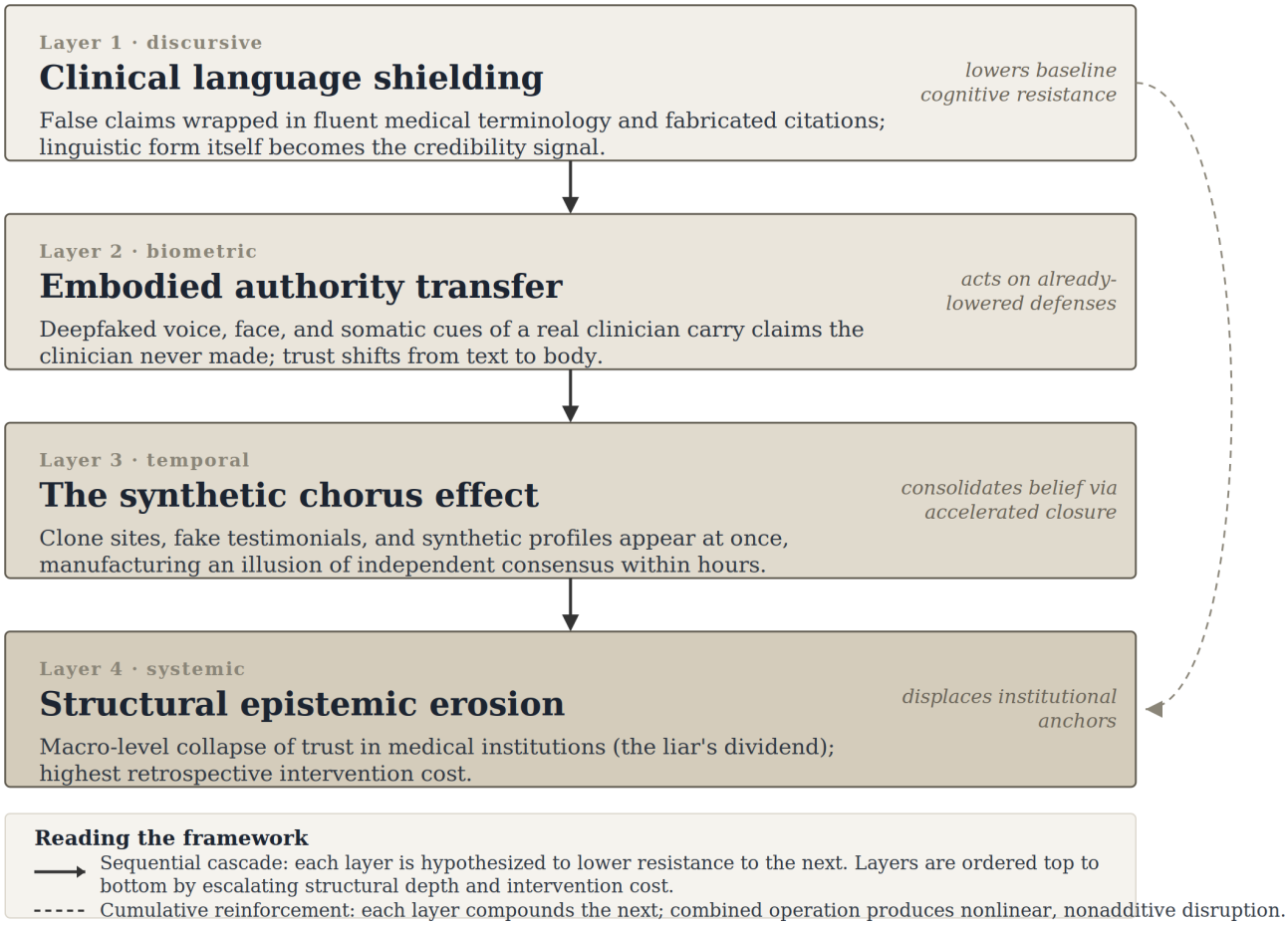


Table . Differentiation of the four MEDF[a] constructs by trust target, mechanism, theoretical basis, and intervention point.

Layer and construct	Primary trust target	Core mechanism	Principal theoretical basis	Primary intervention point
Layer 1: discursive: clinical language shielding	Linguistic markers of clinical competence	False claims embedded in fluent medical terminology and fabricated citations	Machine heuristic [9]; truth-default bias [10]	Source identity verification
Layer 2: biometric: embodied authority transfer	Embodied identity of recognized clinicians	Deepfaked face and voice carry claims the clinician never made	Deepfake perception research [17]	Clinician identity protection
Layer 3: temporal: synthetic chorus effect	Perceived independence and consensus of sources	Near-instantaneous multichannel fabrication of corroborating sources	Source multiplicity [16,18]; need for cognitive closure [14]	Ecosystem-level governance
Layer 4: systemic: structural epistemic erosion	Institutional credibility of medical authority	Pervasive synthetic content induces chronic skepticism toward authentic communication	Liar's dividend [11]; post-truth distrust dynamics [19]	Proactive trust infrastructure

[a]MEDF: Multilayered Epistemic Disruption Framework.

XSL•FO
RenderX

Theoretical Framework and Conceptual Positioning

Gap in the Literature and Conceptual Positioning

The MEDF is distinguished from prior work by 3 features: it treats threat and defense as two faces of one mechanism rather than separate lists; it defines the layers as independent but interconnected, such that simultaneous operation yields nonlinear rather than additive harm; and it is domain-specific. Layered frameworks exist in general disinformation research, but none systematically theorizes the vulnerability dynamics particular to health—the clinical language shield, the lowered epistemic threshold that accompanies patient anxiety, and the reflex of deference to clinical authority.

The closest conceptual neighbors in the literature are as follows:

1. Sundar's machine heuristic [9] describes the automatic trust users extend to machine-generated content and supplies the psychological foundation for clinical language shielding. The fourth layer builds on the liar's dividend [11]. The epistemic trust literature by Fonagy et al [12] provides grounding for understanding trust mechanisms in health contexts. Park and Nan's [1] epistemic ambivalence offers the first systematic theorization of AI's paradoxical capacity to both construct and erode public knowledge. These works are the MEDF's predecessors; none of them, however, integrates these mechanisms into a unified framework specific to health misinformation.
2. A word on conceptual boundaries: clinical language shielding differs from the "generative illusion" [20] and "authority cue" concepts in that those capture general large language model (LLM) competence illusions, while clinical language shielding addresses the health-specific version of that illusion and its interaction with patient vulnerability. Likewise, the synthetic chorus effect differs from astroturfing and coordinated inauthentic behavior in that those concepts rest on manual coordination and political motivation; the synthetic chorus effect describes a low-cost, near-instantaneous multichannel production capacity now accessible to individual actors, and its specific consequences in health contexts.
3. Positioning within existing frameworks: a foundational distinction in recent behavioral public policy and health communication research is the contrast between the i-frame and the s-frame [13]. The i-frame locates solutions at the level of individual cognition and behavior; the s-frame addresses the structural, systemic conditions that shape an information ecosystem. Traditional interventions against health misinformation—individual media literacy campaigns, localized fact-checking, and content labeling—operate within the i-frame. The MEDF is explicitly an s-frame model: its central argument is that the primary harm of GenAI does not stem from individual users being deceived case by case, but from the systematic restructuring of the digital environment in ways that lower epistemic thresholds for all users simultaneously.
4. Established public health frameworks for infodemic management provide valuable structural precedents. The 4-I Fact Framework [21] offers a structured approach to information assessment and institutional trust rebuilding. Socioecological and environmental health models [22] have mapped how misinformation circulates at micro, meso, and macro levels. These frameworks were, however, largely developed for pre-GenAI disinformation dynamics or manual, human-coordinated campaigns. The MEDF builds on their structural orientation while addressing three specific dynamics outside their scope: ecosystem-level automation at near-zero marginal cost; the weaponization of clinical epistemology; and a nonlinear cascade logic in which each layer actively degrades resistance to subsequent layers.
5. From structural approaches to epistemic injustice [23], the MEDF draws the insight that epistemic harm operates not only through individual deception but through the systematic undermining of the institutional structures through which credibility is assessed and granted. Applied to AI-mediated health communication, what is eroded is not merely individual belief, but the structural conditions under which clinical authority can function as a credibility signal at all.
6. Construct originality: each of the MEDF's four constructs requires explicit positioning relative to adjacent concepts in prior literature. The locus of the framework's novelty claim should first be stated precisely: although the constructs vary in their distance from existing concepts, the MEDF's primary contribution resides not in any single component but in the interaction architecture—the integration of four mechanisms into a unified, health-specific structure with a hypothesized compounding logic and a layer-specific intervention mapping. Within that architecture, the individual constructs are positioned as follows. Clinical language shielding is the framework's most original analytical contribution. While analogous dynamics—medical framing as a persuasion strategy, professional language as a credibility heuristic—have been documented in health communication and contested illness research, prior literature addresses these as human-authored, limited-scale phenomena. Clinical language shielding formalizes the AI-specific version, characterized by LLM fluency, scalability, and the automated generation of fabricated clinical citations. Embodied authority transfer draws on established deepfake research; its specific contribution is formalizing biometric identity theft as a proposed public health mechanism, distinct from privacy violation or political disinformation contexts, and treating the empirical confirmation of this mechanism as an explicit research agenda item. The synthetic chorus effect is theoretically novel in combining source multiplicity, simultaneous cross-channel presence, and GenAI-enabled instantaneous deployment; existing astroturfing and coordinated inauthentic behavior frameworks, designed for manual, slower-moving operations, do not capture this dynamic. Structural epistemic erosion draws directly on the liar's dividend [11]; the MEDF's contribution is applying this concept to health communication and articulating its differential geographic and institutional impact.

XSL•FO
RenderX

Interlayer Relationships: The Cumulative Disruption Dynamic

The question of how the four layers reinforce one another is the framework's most critical, and most contestable dimension. A fair critique will note that layers overlap in practice: clinical language shielding operates inside deepfake videos; the synthetic chorus effect directly feeds structural erosion. This overlap is not a methodological problem; it is the framework's central claim. The layers intersect. Their analytical value lies precisely in the fact that they require different points of intervention. Countering clinical language shielding calls for source authentication; countering embodied authority transfer calls for identity protection; countering the synthetic chorus effect requires ecosystem-level governance; and countering structural epistemic erosion requires proactive trust infrastructure. No single intervention can address all four simultaneously, and that is why treating AI-driven health misinformation as a single problem with a single fix has so far produced limited results.

The theoretical basis for the nonlinear, cumulative character of this harm rests on two bodies of evidence already present in the literature. First, the work by Fonagy et al [12] on epistemic trust in psychotherapeutic contexts established that trust mechanisms are hierarchically organized, with credibility assessments at one level modulating openness to information at the next. It must be stated transparently that this framework was developed to account for mentalizing processes in clinical relationships; applying it to AI-mediated health information ecosystems is a deliberate theoretical extension, and the cascade mechanism proposed here should be read as a theoretically motivated hypothesis rather than an established finding, pending direct experimental validation. The proposed dynamic is as follows: when clinical language shielding successfully mimics technical competence, it is hypothesized to lower the user's baseline cognitive resistance; deepfake imagery then encounters a subject whose epistemic defenses have already been partially lowered; the synthetic chorus effect consolidates a belief formed under conditions of reduced resistance; and structural erosion sets in as those beliefs begin to displace institutional anchors. This proposed cascade, if empirically confirmed, would produce harm that is nonadditive.

Second, Kruglanski and Webster's [14] need for cognitive closure establishes that once a satisfying explanation is reached, the tendency is to defend rather than reopen it. Critically, the speed at which closure is reached increases with the number of consistent signals encountered. Multichannel exposure, the defining feature of the third layer, does not merely add to the probability of belief; it accelerates the closure process in a nonlinear fashion, because each additional consistent signal reduces the cognitive resources the subject is willing to invest in continued scrutiny. The synthetic chorus effect exploits this dynamic directly. Taken together, these two frameworks supply the psychological mechanism that bridges individual layer effects to the cumulative harm the framework predicts: prior trust erosion lowers the threshold for closure, and closure accelerates with signal density in a way that no linear model can capture.

One clarification prevents misreading: the cascade is not advanced as a claim of necessary sequence. Any layer can be activated independently: a user may encounter a clinician deepfake without prior exposure to LLM-generated text, and structural erosion can be fed by the mere public awareness that synthetic content exists [11]. The framework's claim is that when layers cooccur, their joint operation is hypothesized to produce more-than-additive harm; it is the compounding, not the ordering, that constitutes the central empirical proposition.

Layer 1—Discursive: Clinical Language Shielding

Mechanism and Positioning in the Literature

LLMs can command virtually any domain of expertise at high fluency, including medical terminology. That fluency, combined with the models' inability to verify the accuracy of what they produce, generates a health-specific threat: clinical language shielding. The mechanism works as follows: a false health claim is wrapped in medical terminology, plausible-sounding clinical reasoning, and fabricated or distorted references. When a user reads the result, it is not the content they evaluate, it is the language itself that functions as the credibility signal. This explains both Sundar's machine heuristic [9]—the automatic trust attribution users extend to machine-generated output—and the paradox identified by Markowitz and Hancock [10]: LLMs are more truth-biased than humans, which is precisely what makes their production of false claims so dangerous.

Empirical Evidence

A 2026 Lancet Digital Health study by Omar and colleagues provided systematic documentation of this mechanism, analyzing 3.4 million prompts across 20 LLMs [24]. Testing with real hospital discharge notes, social media health myths, and 300 clinician-approved scenarios, the researchers found that existing safety filters could not reliably distinguish true from fabricated claims when those claims were wrapped in familiar clinical language. A related randomized controlled experiment involving medical students [25] found that misleading AI explanations significantly degraded diagnostic accuracy, while accurate AI explanations produced no statistically significant improvement. This asymmetry, the harm of false information outweighing the benefit of true information, is precisely what makes clinical language shielding so dangerous in health contexts.

The scope of this layer extends well beyond supplement marketing. Research on vaccine hesitancy shows that LLMs can present inaccurate claims about vaccine safety within a scientific-looking frame when safety guardrails are bypassed [26]. In oncology contexts, herbal treatments and off-protocol regimens can be legitimized through clinical terminology. The same system-instruction vulnerabilities documented by Modi et al [26] make it technically feasible to frame unproven treatments in scientifically credible language once guardrails are circumvented.

XSL•FO
RenderX

Testable Hypothesis

Hypothesis 1 predicts that identical false claims framed in clinical terminology produce higher belief and compliance than in everyday language.

Layer 2—Biometric: Embodied Authority Transfer

Mechanism

Where clinical language shielding operates at the discursive level, deepfake technology carries the threat into a biometric one: it is no longer language that provides reassurance, but the body itself. A recognized physician's face, voice, posture, and eye contact activate social recognition and parasocial trust mechanisms that operate below conscious scrutiny. This framework terms this dynamic embodied authority transfer: the weaponization of a real medical authority's biometric identity to carry claims that authority never made. Barrington et al [17] found that people cannot reliably distinguish AI-generated voice clones from real speech: participants perceived a cloned voice to be the same as its real counterpart approximately 80% of the time, and correctly identified a voice as AI-generated only about 60% of the time. Liu et al [27] demonstrated experimentally that AI doctor avatars appearing older elicit higher trust ratings, suggesting that deepfake designers are deliberately modeling preexisting cognitive biases in their audiences.

Documented Patterns and Topic Distribution

A Bitdefender Labs investigation covering March–May 2024 on Meta platforms identified over 1000 deepfake health videos promoting more than 40 supplement products, targeting diabetes, joint pain, vision problems, and cardiac conditions [28]. A Full Fact investigation documented networks of accounts using deepfaked academics and clinicians across TikTok (ByteDance) and Instagram (Meta Platforms, Inc) [29], and reporting in the medical press has described the same impersonation pattern across other major platforms [30]. A 2026 statement by the American Medical Association chief executive officer in STAT News [30] documented the same pattern concentrated around glucagon-like peptide-1 alternatives, weight loss pills, and diabetes treatments.

The topics targeted are not random. Diabetes, obesity, and weight management, menopause, joint pain, and antiaging products cluster around conditions characterized by chronicity, stigma, waiting-list pressure, and a perceived inadequacy of conventional care. This mirrors broader health misinformation patterns: vaccine misinformation research shows an identical vulnerability mapping, with high-anxiety and low-institutional-trust contexts as primary targets [31]. The Wellness Nest case illustrates the dynamic concretely: Professor David Taylor-Robinson's expertise in child health was weaponized to promote menopause supplements — a deliberate mismatch between domain authority and topic that reveals how the mechanism works. Affected clinicians have had to spend hours trying to have deepfake content removed, defending both their professional reputation and their patients' safety.

Testable Hypothesis

Hypothesis 2 predicts belief and compliance increase from text to anonymous avatar to recognized-clinician deepfake, with the largest gap at the deepfake step.

Layer 3—Temporal: The Synthetic Chorus Effect

Mechanism

The synthetic chorus effect names a specific kind of skepticism erosion: the dissolution of doubt that occurs when AI constructs a comprehensive credibility infrastructure around a single claim—websites, academic-looking articles, patient testimonials, clinician blogs, social media profiles, and videos—simultaneously and within hours. This mechanism is best understood as the intersection of three well-established psychological processes, now operating in an environment their original theorists could not have anticipated. First, source multiplicity research consistently shows that information attributed to multiple independent sources is perceived as substantially more credible than the same information from a single source [16,18]. Second, repeated exposure, independent of source, increases the perceived truth of a statement: the illusory truth effect [16]. The synthetic chorus effect is analytically distinct from, and complementary to this mechanism: the illusory truth effect operates through temporal repetition of the same claim across successive encounters, whereas the chorus operates through the structural density of apparently independent corroborating sources encountered at a single moment. Repetition builds familiarity over time; the chorus manufactures consensus all at once. Third, Paivio's [15] dual coding theory and Mayer's [32] cognitive theory of multimedia learning demonstrate that information encoded simultaneously through verbal and audiovisual channels leaves a stronger, more durable memory trace than single-channel exposure, reinforcing both retention and subjective certainty. Add to this Kruglanski and Webster's [14] need for cognitive closure: once a person reaches a satisfying explanation, they tend to defend that judgment rather than reopen it. Multichannel exposure accelerates that closure. The synthetic chorus effect exploits all of this at once. Every apparent source—the website, the blog, the patient story, and the video—was produced by a single actor within hours. The user perceives independent corroboration; what they are experiencing is a coordinated illusion of consensus.

The Temporal Factor and Collapse of Defensive Capacity

The most critical feature of this layer is its speed. Traditional disinformation operations took time to build credibility: the age of a website, the history of a social media account, accumulated engagement patterns—these were all verification signals. AI eliminates that cost. A coherent, mutually reinforcing information ecosystem spanning multiple channels can now be built in hours. Even where health information consumers approach isolated videos or single promotional pages with skepticism, the framework proposes that encountering a whole system—website, blog, social media profiles, patient

XSL•FO
RenderX

testimonials, and user reviews—can erode that skepticism considerably; this proposition is formalized in hypothesis 3 below.

The synthetic chorus effect's emergence is best understood against the backdrop of how social media had already transformed information environments before GenAI. Research on online news spread established that false stories diffuse faster, more broadly, and more deeply than true ones across social networks, driven primarily by human sharing behavior [33]. Earlier analyses documented how the structural incentive architecture of social media, algorithmic amplification of engagement-maximizing content, had already converted passive information consumers into active amplifiers, a dynamic Phillips [34] described as the oxygen of amplification in the context of political disinformation. GenAI did not create this vulnerability; it automated its exploitation at a previously inaccessible scale. The concepts of astroturfing and coordinated inauthentic behavior describe the political predecessors of the synthetic chorus effect, but were developed around the manual coordination capacity of the pre-GenAI era. The DC Weekly operation's documented increase in content production through AI use [35] illustrates the scale shift: what previously required organized human networks can now be achieved by a single actor within hours.

Empirical Gap and Testable Hypothesis

No empirical study has directly tested the synthetic chorus effect—this is one of the paper's most important research gap findings. Existing disinformation research predominantly addresses isolated content units. Controlled experiments measuring the effect of simultaneous multichannel presence on user trust and health behavior represent the most pressing research agenda this framework generates.

Hypothesis 3 predicts simultaneous multichannel presence raises trust nonlinearly, the gain from two to four channels exceeding that from one to two.

Layer 4—Systemic: Structural Epistemic Erosion

Mechanism: The Liar's Dividend and Institutional Trust

The cumulative effect of the first three layers feeds the fourth and deepest: structural epistemic erosion. This dynamic, what Chesney and Citron [11] named the liar's dividend, the way that the mere existence of fake content erodes trust in real content, is particularly destructive in health communication. A patient who has watched a deepfake of a doctor may approach their next real clinical consultation, or the next message from a public health authority, with a residual skepticism that was not there before. This erosion does not target individual health decisions; it undermines institutional trust in medical authority as such. A nationally representative US study of 5041 adults found that those who reported encountering substantial health misinformation on social media were 1.66 times more likely to report low trust in the health care system (odds ratio 1.66, 95% CI 1.11 - 2.48) [36].

Documented cases, including patients who abandoned prescribed treatments after watching deepfake physician endorsements of alternative remedies [29,30], illustrate how medical authority can be structurally displaced. Health communication research shows that once institutional distrust is established, the effectiveness of corrective messages drops substantially [19]. On this proposed logic, the fourth layer may represent the most consequential: structural epistemic erosion carries the highest projected retrospective intervention cost.

Epistemic Fatigue and Beyond the Global North

A second mechanism within structural erosion is epistemic fatigue: the cognitive exhaustion of constantly asking which information is real and which was generated by AI. Research shows that exposure to highly realistic disinformation weakens confidence in distinguishing truth from fabrication—a kind of cynicism some researchers have begun calling truth fatigue [37]. The authority being eroded is, as medical sociology has long documented, a historically constructed achievement rather than a natural fact: professional dominance over health knowledge and the lay–expert boundary were institutionalized over more than a century of professionalization [38,39]. Structural epistemic erosion can therefore be read as the synthetic acceleration of that construction's reversal—an authority assembled over a century, destabilized at the pace of automated content generation. Experimental evidence is consistent with this proposed dynamic: exposure to deepfakes has been shown to generate uncertainty more readily than outright deception, and that uncertainty in turn reduces trust in news encountered on social media [40] — a cynicism pathway that structural epistemic erosion extends to the health domain.

The geographic dimension of this layer deserves attention. Countries with developed digital infrastructure but relatively weak regulatory frameworks and uneven health literacy, for example, Brazil, India, and Nigeria, represent the contexts where this framework's predictions are likely sharpest. In lower- and middle-income countries (LMICs), deepfake detection infrastructure is limited, platform oversight is thin, and clinicians have limited capacity to monitor their own digital footprint. The MEDF's health-domain specificity is most pronounced in these settings: where baseline institutional health distrust is already high, structural epistemic erosion can advance considerably faster.

Testable Hypothesis

Hypothesis 4 predicts cumulative deepfake exposure reduces trust in clinicians and institutions and depresses health-seeking behavior, most sharply where baseline distrust is high.

Evaluation of Existing Defense Mechanisms

Content Labeling and Provenance Standards

China's Cyberspace Administration regulations, effective September 2025, require all AI-generated content to carry labeling at both the visual and metadata level. The EU AI Act and US state-level legislation impose comparable transparency obligations. Technical standards such as C2PA (Coalition for

XSL•FO
RenderX

Content Provenance and Authenticity) aim for cryptographic verification of content origin. These measures offer a partial response to layers 1 and 2, but carry two fundamental limitations.

The first is an enforcement gap: synthetic health content frequently circulates without any transparency label, as metadata can be stripped during peer-to-peer reposting and visual watermarks can be deprioritized in platform interface designs. Initiatives such as Google's SynthID and TikTok's C2PA integration are promising, but their effectiveness depends on platform compliance that is not yet mandatory or consistently monitored. The second limitation is that labeling does not directly target the synthetic chorus effect (Layer 3) or structural epistemic erosion (Layer 4): flagging a single content unit within a coherent ecosystem does not undo the system's persuasive force.

Detection Systems and Their Limitations

The World Health Organization's EARS (early AI-supported response with social listening) platform and similar systems hold real potential for early detection of health misinformation. However, the Deepfake-Eval-2024 benchmark found that state-of-the-art deepfake detection models show area under the curve drops of up to 50% for video, 48% for audio, and 45% for images when evaluated on real-world content versus controlled datasets [41]. The synthetic chorus effect compounds the challenge further: when campaign components are distributed across platforms and appear independent of one another, algorithmic identification of their coordination is technically difficult.

Prebunking: The Individual-Level Limitation

The demonstrated inadequacy of post hoc correction has driven interest in prebunking. Experimental evidence supports the effectiveness of gamified prebunking interventions in certain contexts [42]. Viewed through the MEDF lens, however, prebunking carries a structural limitation that the i-frame and s-frame distinction [13] makes precise: it is an i-frame intervention targeting individual cognition, whereas the synthetic chorus effect operates at the s-frame level of the information environment itself. Equipping individuals to critically evaluate isolated content units may not be enough to hold back an ecosystem-level manipulation. A person who remains skeptical of a single piece of content can still find their defenses eroded by a coherent ecosystem of dozens of apparently independent sources. Individual media literacy is necessary but cannot substitute for structural, s-frame interventions at the ecosystem level.

Policy Recommendations

The MEDF makes a straightforward argument: because each of its four layers requires a different point of intervention, policy frameworks that address only one or two of them will leave significant damage pathways open. Consistent with the framework's s-frame orientation [13], the recommendations below target the structural conditions of the information ecosystem rather than individual user behavior. They are derived directly from the framework's structure and are calibrated against what existing initiatives address and what they miss.

Source Identity Verification

An effective defense against clinical language shielding must target the source, not the language. This means developing mandatory source authentication standards for AI systems that generate health information, and expanding the availability of health communication AI models trained on peer-reviewed medical literature with strengthened safety barriers. The practical obstacles are real: platform-level compliance requires global coordination, and companies such as Meta and Google may resist such constraints as encroachments on commercial autonomy. Joint American Medical Association–World Health Organization platform standardization efforts are promising in this space but have not yet produced binding frameworks.

Clinician Identity Protection Protocols

Recognizing clinician identity as a legally protected professional asset, as the American Medical Association has proposed, and requiring platforms to verify accounts using physician identities for health content are critical steps. The US TAKE IT DOWN Act (May 2025) mandates removal of synthetic personal content within 48 hours, but targets privacy violations rather than health misinformation. A health-specific equivalent does not yet exist. Article 50 transparency requirements in the European Union AI Act could partially close this gap, but enforcement mechanisms need strengthening. A further structural constraint follows from the framework's own temporal logic: false content diffuses significantly faster than true content across social networks [33], and the synthetic chorus effect compresses ecosystem construction into hours, so a 48-hour removal window can intervene only after the bulk of organic dissemination has already occurred. Removal mandates are necessary, but they are temporally outpaced by the threat they target.

Ecosystem-Level Governance

The synthetic chorus effect requires interventions that go beyond individual content moderation. Risk-scoring systems that evaluate account age, engagement patterns, and content consistency together, and shared threat intelligence mechanisms that enable detection of coordinated campaigns across platforms, are the natural policy response. Researcher–platform collaboration models such as Stanford Internet Observatory's Spamouflage tracking offer good precedents; health-specific equivalents have not yet been institutionalized. Cross-platform threat intelligence also presupposes moderation capacity that is unevenly distributed: in the LMIC contexts where this framework predicts the sharpest effects, platform moderation resources and local-language coverage are thinnest, and operations hosted outside national jurisdictions remain largely beyond the reach of instruments such as the European Union AI Act.

Proactive Trust Infrastructure

Structural epistemic erosion cannot be meaningfully addressed through reactive correction alone. User-friendly tools that allow clinicians to cryptographically sign authentic content (C2PA-compatible), institutional mechanisms for health organizations to mark their digital presence as verifiable, and

XSL•FO
RenderX

health literacy programs redesigned to include ecosystem awareness, teaching people to recognize coordinated information systems, not just individual false claims, all belong in this layer. In LMIC contexts, building this infrastructure requires additional resources; it also offers the greatest potential impact in precisely those settings. Given the temporal limits of removal noted above, platform architectures could also explore correction mechanisms that preserve rather than delete. Because removal can feed a censorship narrative consistent with the liar's dividend [11], and because corrective messages lose effectiveness once institutional distrust is established [19], replacing an identified synthetic impersonation with a clinician-authorized disclaimer, while leaving its distribution links intact, would convert the falsehood's accumulated reach into a delivery vector for authentic correction. Such interception protocols remain untested and would face substantial regulatory, privacy, and platform-liability constraints; they are advanced here as a direction for design research rather than as deployable solutions.

A final consideration concerns sequencing and evaluability. These four intervention classes differ in their deployment horizons. Provenance labeling and ecosystem-level risk governance can be pursued immediately through existing instruments, C2PA standards and, in the European Union, the systemic-risk assessment obligations that the Digital Services Act already imposes on very large online platforms, whereas clinician biometric protection requires new institutional mechanisms, and proactive trust infrastructure is a long-horizon investment whose effects compound slowly. The framework's hypotheses double as evaluation criteria: the relationship that hypothesis 4 predicts between cumulative deepfake exposure and institutional trust, for example, provides a measurable baseline against which the effectiveness of any of these interventions can be assessed over time. The recommendations are thus offered not as a finished regulatory program but as a layer-indexed agenda whose components can be prioritized by deployment horizon and monitored through the same empirical apparatus the framework proposes.

Testable Hypotheses, Conclusions, and Limitations

Research Agenda

Four testable hypotheses, formally stated in their respective layer sections above, are proposed to guide empirical validation of the MEDF. In summary: hypothesis 1 predicts a clinical-terminology advantage in belief and behavioral compliance for identical false claims (layer 1); hypothesis 2 predicts an ordered increase in trust from text to anonymous synthetic avatar to recognized-clinician deepfake, with the largest gap at the deepfake step (layer 2); hypothesis 3 predicts nonlinear gains in trust under simultaneous multi-channel presence (layer 3); and hypothesis 4 predicts that cumulative deepfake exposure erodes trust in real clinicians, health institutions, and health-seeking behavior, most sharply in populations with high baseline institutional distrust (layer 4).

A and B experiments are the natural methodology for testing hypothesis 1, hypothesis 2, and hypothesis 3; longitudinal survey designs for hypothesis 4; and cross-cultural comparative studies for the LMIC dimensions of hypothesis 3 and hypothesis 4. Full operational specifications for all four hypotheses are provided in Multimedia Appendix 1.

Conclusions

This viewpoint has proposed an analytical model, the MEDF, that addresses AI's role in health misinformation beyond the dual-role threat-opportunity framing that currently dominates the literature. The argument is that the danger does not lie in the falsehood of individual content units but in AI's capacity to embed false information simultaneously across four distinct layers of human trust systems.

The paper makes three contributions. First, it systematically theorizes the vulnerability dynamics specific to health misinformation, clinical language shielding, embodied authority transfer, the synthetic chorus effect, and structural epistemic erosion, and explicitly differentiates each from similar concepts in adjacent literature. Second, it maps which existing defense mechanisms address which layers and which layers they leave untouched, deriving policy recommendations with an honest account of their practical constraints. Third, it acknowledges the research gap openly, advances testable hypotheses for each layer, and contributes to shaping an empirical research agenda.

Limitations

Three limitations deserve acknowledgment. First, the MEDF is an untested theoretical framework. The synthetic chorus effect in particular, and the cumulative harm claim more broadly, are propositions awaiting controlled experimental validation. Second, the paper draws predominantly on English-language and Western-context literature. Different authority perceptions, local language dynamics, and digital literacy patterns in LMIC contexts are underrepresented—a limitation that has been flagged throughout the analysis but that comparative research will need to systematically address. Third, the documented cases come predominantly from the supplement and commercial health fraud context; systematic mapping of similar patterns in vaccine misinformation, pandemic management, and oncology would substantially extend the framework's scope.

Acknowledgments

In accordance with JMIR Publications policy on the use of artificial intelligence (AI) and large language models (LLMs) in scholarly work, the author discloses the following. The intellectual core of this manuscript—the conception of the Multilayered Epistemic Disruption Framework (MEDF), the definition and differentiation of its four constructs, the cascade logic linking them, the policy analysis, and the testable hypotheses—was developed by the human author, who originally drafted the work in Turkish. LLM-based tools (OpenAI ChatGPT, Google Gemini, and Anthropic Claude) were used as assistive instruments under the author's

XSL•FO
RenderX

direction for the following tasks: translation of the author's Turkish-language drafts into academic English; restructuring of the manuscript to conform to JMIR formatting requirements (eg, the numbered reference style); language editing for concision and readability; and assistance in locating, formatting, and verifying bibliographic references during revision. All AI-assisted output was reviewed, edited, and verified by the author, who takes full responsibility for the accuracy, integrity, and originality of the entire manuscript, including all claims, citations, and conclusions. The author confirms that no AI tool is listed or qualifies as an author, consistent with International Committee of Medical Journal Editors and JMIR authorship criteria.

Funding

No specific funding was received for the research. The article processing charge was partially supported by İstanbul Medipol University.

Data Availability

Not applicable. This is a viewpoint paper presenting a theoretical framework; no datasets were generated or analyzed.

Authors' Contributions

MM, as sole author, conceived the framework and all of its constructs, conducted the literature review, drafted the manuscript in Turkish, directed and verified the AI-assisted translation and formatting process described in the disclosure above, and reviewed and approved the final English text.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full operationalization of the four MEDF hypotheses, including proposed conditions, predicted direction of effect, and indicative empirical designs for each layer.
[DOCX File, 11 KB - infodemiology_v6i1e96664_app1.docx]

References

1. Park S, Nan X. Generative AI and misinformation: a scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation. AI & Soc 2026 Feb;41(2):1501-1515. [doi: 10.1007/s00146-025-02620-3]
2. National Academies of Sciences, Engineering, and Medicine. Understanding and addressing misinformation about science. : The National Academies Press; 2024. [doi: 10.17226/27894]
3. Lai YK, Lai Z, Zhao X. Counteracting cyberchondria in Chinese chronic disease patients: The divergent roles of health-related social media use and online patient-centered communication. Patient Educ Couns 2025 Dec;141:109337. [doi: 10.1016/j.pec.2025.109337] [Medline: 40945192]
4. Lyons B, King AJ, Barter RL, Kaphingst KA. Exposure to low-credibility online health content is limited and is concentrated among older adults. Nat Aging 2026 Feb;6(2):454-462. [doi: 10.1038/s43587-025-01059-x] [Medline: 41639217]
5. Delmastro M, Paciello M. Depression, reduced education, and bias perceptions as risk factors of beliefs in misinformation. Sci Rep 2022 Sep 30;12(1):16408. [doi: 10.1038/s41598-022-20640-7] [Medline: 36180772]
6. Vidgen B, Taylor H, Pantazi M, Anastasiou Z, Inkster B, Margetts H. Understanding Vulnerability to Online Misinformation: The Alan Turing Institute; 2021. URL: https://www.turing.ac.uk/news/publications/understanding-vulnerability-online-misinformation [accessed 2026-07-27]
7. Borg MA, Waisfisz B, Frank U. Quantitative assessment of organizational culture within hospitals and its relevance to infection prevention and control strategies. J Hosp Infect 2015 May;90(1):75-77. [doi: 10.1016/j.jhin.2014.12.015] [Medline: 25676113]
8. Touboul-Lundgren P, Jensen S, Drai J, Lindbæk M. Identification of cultural determinants of antibiotic use cited in primary care in Europe: a mixed research synthesis study of integrated design "Culture is all around us". BMC Public Health 2015 Sep 17;15:908. [doi: 10.1186/s12889-015-2254-8] [Medline: 26381376]
9. Sundar SS. The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ, Flanagin AJ, editors. Digital Media, Youth, and Credibility: MIT Press; 2008:73-100 URL: https://search.issuelab.org/resources/875/875.pdf [accessed 2026-07-27]
10. Markowitz DM, Hancock JT. Generative AI are more truth-biased than humans: a replication and extension of core truth-default theory principles. J Lang Soc Psychol 2024 Mar;43(2):261-267. [doi: 10.1177/0261927X231220404]
11. Chesney R, Citron DK. Deep fakes: a looming challenge for privacy, democracy, and national security. Calif Law Rev 2019;107:1753-1819. [doi: 10.15779/Z38RV0D15J]
12. Fonagy P, Luyten P, Allison E, Campbell C. Mentalizing, epistemic trust and the phenomenology of psychotherapy. Psychopathology 2019;52(2):94-103. [doi: 10.1159/000501526] [Medline: 31362289]

XSL•FO
RenderX

13. Chater N, Loewenstein G. The i-frame and the s-frame: how focusing on individual-level solutions has led behavioral public policy astray. Behav Brain Sci 2022 Sep 5;46:e147. [doi: 10.1017/S0140525X22002023] [Medline: 36059098]
14. Kruglanski AW, Webster DM. Motivated closing of the mind: “seizing” and “freezing”. Psychol Rev 1996 Apr;103(2):263-283. [doi: 10.1037/0033-295x.103.2.263] [Medline: 8637961]
15. Paivio A. Mental Representations: A Dual Coding Approach: Oxford University Press; 1986. [doi: 10.1093/acprof:oso/9780195066661.001.0001]
16. Nadarevic L, Reber R, Helmecke AJ, Köse D. Perceived truth of statements and simulated social media postings: an experimental investigation of source credibility, repeated exposure, and presentation format. Cogn Res Princ Implic 2020 Nov 11;5(1):56. [doi: 10.1186/s41235-020-00251-4] [Medline: 33175284]
17. Barrington S, Cooper EA, Farid H. People are poorly equipped to detect AI-powered voice clones. Sci Rep 2025 Mar 31;15(1):11004. [doi: 10.1038/s41598-025-94170-3] [Medline: 40164656]
18. Harkins SG, Petty RE. Information utility and the multiple source effect. J Pers Soc Psychol 1987;52(2):260-268. [doi: 10.1037/0022-3514.52.2.260]
19. Lewandowsky S, Ecker UKH, Cook J. Beyond misinformation: understanding and coping with the “post-truth” era. J Appl Res Mem Cogn 2017;6(4):353-369. [doi: 10.1016/j.jarmac.2017.07.008]
20. Joseph J, Jose B, Jose J. The generative illusion: how ChatGPT-like AI tools could reinforce misinformation and mistrust in public health communication. Front Public Health 2025;13:1683498. [doi: 10.3389/fpubh.2025.1683498] [Medline: 41080885]
21. Sundelson AE, Jamison AM, Huhn N, Pasquino SL, Sell TK. Fighting the infodemic: the 4 i Framework for Advancing Communication and Trust. BMC Public Health 2023 Aug 30;23(1):1662. [doi: 10.1186/s12889-023-16612-9] [Medline: 37644563]
22. John JN, Gorman S, Scales D. Understanding interventions to address infodemics through epidemiological, socioecological, and environmental health models: framework analysis. JMIR Infodemiology 2025 Mar 24;5:e67119. [doi: 10.2196/67119] [Medline: 40127460]
23. Fricker M. Epistemic Injustice: Power and the Ethics of Knowing: Oxford University Press; 2007. [doi: 10.1093/acprof:oso/9780198237907.001.0001]
24. Omar M, Sorin V, Wieler LH, et al. Mapping the susceptibility of large language models to medical misinformation across clinical notes and social media: a cross-sectional benchmarking analysis. Lancet Digit Health 2026 Jan;8(1):100949. [doi: 10.1016/j.landig.2025.100949] [Medline: 41672646]
25. Teng D, Tan L, Cao Q, et al. Impact of AI misinformation on diagnostic accuracy and confidence calibration in novice medical students. NPJ Digit Med 2026 Mar 17;9(1):356. [doi: 10.1038/s41746-026-02547-z] [Medline: 41844809]
26. Modi ND, Menz BD, Awaty AA, et al. Assessing the system-instruction vulnerabilities of large language models to malicious conversion into health disinformation chatbots. Ann Intern Med 2025 Aug;178(8):1172-1180. [doi: 10.7326/ANNALS-24-03933] [Medline: 40550134]
27. Liu T, Wang P, Pan D, Liu R. Credibility of AI generated and human video doctors and the relationship to social media use. Front Public Health 2025;13:1559378. [doi: 10.3389/fpubh.2025.1559378]
28. Deep dive on supplement scams: how AI drives ‘Miracle Cures’ and sponsored health-related scams on social media. Bitdefender. URL: https://www.bitdefender.com/en-us/blog/labs/deep-dive-on-supplement-scams-how-ai-drives-miracle-cures-and-sponsored-health-related-scams-on-social-media [accessed 2026-07-27]
29. Revealed: how academics are being deepfaked on TikTok and Instagram to promote supplements. Full Fact. 2025. URL: https://fullfact.org/health/academics-deepfaked-tiktok-wellness-nest/ [accessed 2026-06-27]
30. AMA CEO: deepfake doctors are a threat to public health. STAT News. 2026. URL: https://www.statnews.com/2026/02/17/deepfake-doctors-scam-ama/ [accessed 2026-02-17]
31. Jafar Z, Quick JD, Larson HJ, et al. Social media for public health: reaping the benefits, mitigating the harms. Health Promot Perspect 2023;13(2):105-112. [doi: 10.34172/hpp.2023.13] [Medline: 37600540]
32. Mayer RE. Multimedia Learning, 2nd edition: Cambridge University Press; 2009.
33. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 9;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
34. Phillips W. The oxygen of amplification. Data & Society. 2018. URL: https://datasociety.net/wp-content/uploads/2018/05/2-PART-2_Oxygen_of_Amplification_DS.pdf [accessed 2026-07-21]
35. Wack M, Ehrett C, Linvill D, Warren P. Generative propaganda: evidence of AI’s impact from a state-backed disinformation campaign. PNAS Nexus 2025 Apr;4(4):pgaf083. [doi: 10.1093/pnasnexus/pgaf083] [Medline: 40171239]
36. Chandrasekaran R, Moustakas E, Sadiq MT. Racial and demographic disparities in susceptibility to health misinformation on social media: national survey-based analysis. J Med Internet Res 2024 Nov 6;26:e55086. [doi: 10.2196/55086] [Medline: 39504121]
37. AI in the age of fake (imagined) content. Stimson. 2026 Feb. URL: https://www.stimson.org/2026/ai-in-the-age-of-fake-imagined-content/ [accessed 2026-07-21]

XSL•FO
RenderX

38. Freidson E. Profession of Medicine: A Study of the Sociology of Applied Knowledge: Dodd, Mead; 1970. URL: https://archive.org/details/professionofmedi0000frei [accessed 2026-07-27]
39. Starr P. The Social Transformation of American Medicine: Basic Books; 1982. URL: https://archive.org/details/socialtransforma0000star [accessed 2026-07-27]
40. Vaccari C, Chadwick A. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Soc Media Soc 2020 Jan;6(1):2056305120903408. [doi: 10.1177/2056305120903408]
41. Chandra NA, Murtfeldt R, Qiu L, et al. Deepfake-eval-2024: a multi-modal in-the-wild benchmark of deepfakes circulated in 2024. arXiv. Preprint posted online on Mar 4, 2025. [doi: 10.48550/arXiv.2503.02857]
42. Kozyreva A, Lorenz-Spreen P, Herzog SM, et al. Toolbox of individual-level interventions against online misinformation. Nat Hum Behav 2024 Jun;8(6):1044-1052. [doi: 10.1038/s41562-024-01881-0] [Medline: 38740990]

Abbreviations

C2PA: Coalition for Content Provenance and Authenticity
GenAI: generative AI
LLM: large language model
LMIC: lower- and middle-income countries

Edited by T Purnat; submitted 30.Mar.2026; peer-reviewed by D Scales, T Purnat; revised version received 09.Jul.2026; accepted 09.Jul.2026; published 20.Aug.2026.

Please cite as:
Malkoç M
Multilayered Epistemic Disruption in AI-Driven Health Misinformation: Conceptual Framework and Viewpoint
JMIR Infodemiology 2026;6:e96664
URL: https://infodemiology.jmir.org/2026/1/e96664
doi:10.2196/96664

© Muzaffer Malkoç. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 20.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Information Void in Asymptomatic Chronic Disease: A Digital Health Framework for Understanding Social Media Health Information Seeking in Young Adults

Victoria Sze Min Ekstrom[1,2], MBBS, MRCP, MSc

[1]Department of Gastroenterology and Hepatology, Singapore General Hospital, 20 College Rd, Singapore, Singapore

[2]SingHealth Duke-NUS Medical Humanities Institute, Singapore

Corresponding Author:
Victoria Sze Min Ekstrom, MBBS, MRCP, MSc
Department of Gastroenterology and Hepatology, Singapore General Hospital, 20 College Rd, Singapore, Singapore

Abstract

Nearly 1 in 4 young adults has a chronic condition, yet many feel well despite their diagnosis. Asymptomatic conditions such as prediabetes and hypertension create a unique vulnerability to digital health misinformation, particularly on platforms where inaccurate content is prevalent. Conventional clinical responses, which often just warn patients about online misinformation, fail to address the underlying drivers of this behavior. This viewpoint proposes a novel disease characteristic–based vulnerability framework to understand this challenge, grounded in established behavioral science theories such as the capability, opportunity, and motivation–behavior model; temporal discounting; and the concept of information voids in infodemiology. We identify a critical "information void" for asymptomatic conditions managed primarily through lifestyle modification. This void, created by the absence of symptomatic feedback combined with delayed clinical biomarker feedback, compels patients to seek information online. Instead of viewing this information seeking as a problematic deviation, we reframe it as a "digital phenotype" indicating a patient's readiness for behavior change. Through case studies illustrating how this framework applies to specific conditions (prediabetes, nonalcoholic fatty liver disease, and untreated hypertension), we demonstrate its practical utility for clinicians, health systems, and policymakers. Evidence supports a multipronged approach: integrating digital health literacy into clinical encounters, providing curated evidence-based resources, and pursuing strategic institutional engagement in digital spaces. While acknowledging the framework's deliberate simplification and the need for culturally sensitive adaptation across diverse health care settings, this viewpoint offers a generalizable strategy for engaging with patients' information needs, helping transform a public health challenge into an opportunity for empowerment.

(JMIR Infodemiology 2026;6:e86489) doi:10.2196/86489

KEYWORDS

digital health; social media; health information seeking; asymptomatic chronic disease; prediabetes; dyslipidemia; hypertension; fatty liver disease; nonalcoholic fatty liver disease; NAFLD; young adults; digital health literacy; health literacy; behavior change; patient engagement; misinformation; TikTok; infodemiology

Introduction

Asymptomatic metabolic and cardiovascular conditions are increasingly prevalent among young adults, creating a growing public health concern. Prediabetes now affects approximately 24% of this demographic [1], the prevalence of nonalcoholic fatty liver disease has nearly doubled over 2 decades [2], and 22.4% of adults aged 18 to 39 years have hypertension [3]. Initial management of these conditions typically emphasizes lifestyle modification and relies on biomarker reassessment months later. This approach creates a prolonged "information void" where patients feel healthy but lack actionable guidance. To fill this void, young adults increasingly turn to digital platforms, with scoping reviews confirming their primary use of search engines and social media for health information [4]. For example, a recent survey found that 65.5% of young women intentionally seek health advice on TikTok [5]. However, these platforms contain widespread misinformation. One systematic review found misinformation rates of around 40% for noncommunicable diseases [6], while topic-specific analyses show that less than half of educational videos from influencers may be factual [7] and more than half of popular attention-deficit/hyperactivity disorder content is misleading [8].

This convergence of rising disease prevalence and digital information seeking creates a significant challenge for evidence-based care. Clinical guidelines often formalize the long intervals between encounters that define the information void. For instance, recommendations include lipid monitoring every 3 to 12 months [9] or repeat diabetes screening every 3 years for normal results [10]. While the individual components of this problem are recognized, a comprehensive explanatory

XSL•FO
RenderX

framework is missing. There is no model that links specific disease characteristics, such as asymptomatic presentation and delayed, abstract feedback, to a differential vulnerability to health misinformation in young adults. Consequently, conventional clinical advice to simply "be careful online" fails to address the underlying drivers of this vulnerability or offer practical, evidence-based alternatives.

The need to address this gap is urgent. The rising epidemiologic burden is substantial, with 27.1% of young adults now living with multiple chronic conditions [11]. This trend intersects with their established digital health information–seeking behaviors [4,5]. Failing to intervene has serious, long-term consequences. Asymptomatic conditions diagnosed in young adulthood are not benign. Early hypertension is associated with a significantly higher risk of later cardiovascular events [12] and adverse late-life neuroimaging biomarkers [13]. Furthermore, prediabetes in this population progresses to diabetes in 9.5% of cases within 5 years [14]. To mitigate these risks, it is critical to develop targeted strategies that fill the information void with credible guidance.

This article has 3 objectives: First, we synthesize evidence from infodemiology, behavioral science, and digital health to construct a disease characteristic–based framework that explains why patients with certain conditions are disproportionately vulnerable to online misinformation. Second, we apply this framework through illustrative case studies to demonstrate its practical utility. Third, we propose evidence-based digital health interventions designed to address this vulnerability. Our intended audience encompasses frontline clinicians who encounter these patients daily, health systems and policymakers seeking scalable strategies, and researchers working at the intersection of digital health and chronic disease prevention.

Evidence and Theoretical Foundations

Digital Health Information Seeking in Chronic Disease Management

A growing body of evidence documents the role of digital platforms in shaping health behaviors among individuals with chronic conditions. A systematic review by Zhao and Zhang [15] found that social media platforms serve as important sources of both informational and emotional support for individuals managing chronic diseases, with users actively seeking peer experiences and practical self-management strategies. Importantly, the type of platform used and the quality of the information found vary considerably by condition. For highly symptomatic conditions with visible communities (such as diabetes or inflammatory bowel disease), patients often find structured peer support forums and professionally curated content [16]. In contrast, for asymptomatic or preclinical conditions such as prediabetes and early-stage fatty liver disease, the digital landscape is far less organized, with information seekers encountering a fragmented mix of commercial wellness content, personal testimonials, and algorithmically amplified claims [17].

This disparity in digital information quality has measurable consequences. A cross-sectional analysis of TikTok content related to nonalcoholic fatty liver disease found that most popular videos were produced by nonmedical content creators and contained significant inaccuracies regarding dietary recommendations and disease reversibility [18]. Similarly, studies evaluating diabetes and metabolic health information on TikTok have found low reliability and quality scores, with a predominance of commercially motivated messaging from noncredentialed health coaches and content that lacks actionable, evidence-based guidance [19]. These findings are not limited to Western platforms; analyses of Weibo and WeChat health content in China have identified comparable patterns of misinformation in metabolic disease discussions, suggesting that the phenomenon transcends specific cultural or platform contexts [20].

Theoretical Underpinnings

Our proposed framework draws on several established theoretical models. First, it is grounded in the concept of "data voids" and "information voids" from infodemiology, which describes how gaps in authoritative information create vacuums that are filled by low-quality or misleading content [21,22]. The work by Golebiewski and Boyd [21] on data voids in search engines demonstrates that, when reliable information is scarce for a given query, algorithmically ranked results disproportionately surface fringe or commercially motivated content. We extend this concept to the clinical domain, arguing that the temporal gap between diagnosis and biomarker feedback creates a functionally analogous information void for patients.

Second, the framework builds on the Behavior Change Wheel (capability, opportunity, and motivation–behavior; COM-B) model, which posits that behavior arises from the interaction of capability, opportunity, and motivation [23]. From this lens, the information void represents a deficit in both psychological capability (knowledge of how to implement lifestyle changes) and environmental opportunity (access to credible, actionable resources). The patient's subsequent turn to social media is a motivated attempt to address these deficits, which the COM-B model would predict.

Third, we draw on temporal discounting theory from behavioral economics [24,25]. This well-established phenomenon whereby individuals systematically undervalue delayed rewards relative to immediate ones is particularly relevant to asymptomatic conditions. When the clinical message is essentially "change your behavior now to prevent a disease outcome in 10 to 30 years," temporal discounting predicts low motivational salience in the absence of proximate feedback. Social media content creators exploit this cognitive bias by offering promises of rapid, tangible results (eg, "reverse your prediabetes in 30 days"), providing the short-term validation that is absent from long-horizon prevention messages.

Finally, uses and gratifications theory [26] provides a complementary perspective, explaining why young adults actively choose social media over clinical sources. This theory posits that individuals are active agents who select media to satisfy specific needs, including informational needs, social identity needs, and emotional validation needs. For a young adult newly diagnosed with an asymptomatic condition, social media uniquely satisfies all 3: it provides practical "how-to"

XSL•FO
RenderX

information, connects the individual with a peer community that shares their experience, and offers emotional reassurance through success narratives. Traditional clinical encounters, constrained by time and format, often address only the informational need and may not even do so adequately for lifestyle-focused conditions.

The Information Void Framework: Disease Characteristics as Vulnerability Determinants

Building on the theoretical foundations outlined above, our proposed disease characteristic–based framework provides a structured explanation for why certain patient populations are disproportionately vulnerable to digital health misinformation as referenced in Table 1. The framework classifies chronic conditions along two axes: (1) symptom presence (whether the condition produces noticeable symptoms that provide ongoing sensory feedback to the patient) and (2) primary management strategy (whether initial management relies primarily on pharmaceutical intervention, which introduces its own structured feedback loop, or on lifestyle modification, which does not). The intersection of these axes generates 4 distinct quadrants, each characterized by a unique feedback environment and a corresponding level of vulnerability to online misinformation.

Table . Disease characteristic–based framework for vulnerability to digital health misinformation.

	Symptomatic		Asymptomatic	
	Pharmaceutical (quadrant 1)	Lifestyle (quadrant 2)	Pharmaceutical (quadrant 3)	Lifestyle (quadrant 4)
Examples	Type 1 diabetes, rheumatoid arthritis, and epilepsy	Symptomatic IBS[a] (dietary) and chronic pain (exercise based)	Treated hypertension and statin-managed dyslipidemia	Prediabetes, untreated hypertension, NAFLD[b], and early dyslipidemia
Symptom feedback	Present: symptoms validate management (eg, glucose swings)	Present: symptoms provide direct feedback on lifestyle changes	Absent: no somatic cues	Absent: no somatic cues; the patient feels well
Clinical feedback	Rapid and structured: daily glucose monitoring and dose titration	Moderate: symptom diaries and functional improvement over weeks	Structured but abstract: periodic biomarkers (eg, BP[c] readings and lipid panels) linked to medication adjustments	Delayed and abstract: biomarker reassessment months later with no medication anchor; feedback disconnected from daily effort
Information void	Minimal: multiple feedback sources reduce the need for external validation	Low-moderate: symptom changes provide some validation of effort	Moderate: medication provides a structured anchor, but the lack of symptoms reduces urgency for additional information seeking	Maximal: no symptoms, no medication anchor, and no proximate feedback; the patient is left with broad lifestyle advice and months of uncertainty
Misinformation vulnerability	Low	Low-moderate	Moderate	High

[a]IBS: irritable bowel syndrome.

[b]NAFLD: nonalcoholic fatty liver disease.

[c]BP: blood pressure.

We define "vulnerability to digital health misinformation" in this context as the degree to which a patient's clinical situation creates conditions that predispose them to seek, encounter, and potentially act upon inaccurate health information online. This vulnerability is not a fixed individual trait but rather an emergent property of the interaction among disease characteristics, the clinical feedback environment, and the digital information landscape. It is highest when three conditions converge: (1) the patient lacks somatic feedback (no symptoms to confirm or disconfirm advice), (2) clinical feedback is infrequent or abstract (delayed biomarker results rather than immediate, tangible changes), and (3) the information void created by these first 2 factors intersects with a digital ecosystem where low-quality content is algorithmically amplified.

The critical distinction between quadrants 3 and 4 warrants further elaboration. In quadrant 3 (asymptomatic and pharmaceutically managed), conditions such as statin-treated dyslipidemia involve a structured feedback loop: the patient takes a medication daily, attends periodic follow-up appointments, and receives biomarker results (eg, low-density lipoprotein cholesterol levels) that are directly linked to medication dose adjustments. Although the patient feels no symptoms, the medication itself serves as a tangible, daily anchor to the management plan. The act of taking a pill is a concrete behavior with a clear link to the clinical goal, and adjustments to that regimen provide relatively prompt clinical feedback. In quadrant 4, in contrast, no such anchor exists. The patient with prediabetes managed through lifestyle modification alone is told to "eat better and exercise more" and then returns months later for a repeat hemoglobin A_{1c} test. During those intervening months, there is no structured touchpoint, no daily behavioral anchor linked to the clinical plan, and no proximate feedback mechanism to confirm whether their efforts are succeeding. It is precisely this absence of any feedback loop—somatic, pharmacological, or clinical—that creates the

XSL•FO
RenderX

maximal information void and drives patients to seek validation and guidance from digital sources.

We acknowledge that this 2×2 framework is a deliberate simplification of a complex reality. Chronic conditions do not always fit neatly into 1 quadrant; many evolve over time (prediabetes progressing to treated diabetes moves from quadrant 4 to quadrant 3 or 1), and individual patients may occupy different positions based on their specific treatment regimen, comorbidities, and health care context. Moreover, the level of vulnerability is modulated by factors not captured in the framework itself, including health literacy, cultural context, socioeconomic status, health care system structure, and the availability of community-based support [27,28]. A patient with prediabetes in a resource-rich setting with regular access to a multidisciplinary care team faces a very different information void from that of a patient in a resource-limited setting without such support. We present the framework not as a comprehensive model of all factors influencing misinformation vulnerability but as a parsimonious heuristic that isolates 2 modifiable, disease-level characteristics with clear implications for clinical practice and health system design.

Applying the Framework: Illustrative Case Studies

To demonstrate the practical utility of the framework, we present 3 illustrative case studies (Textboxes 1-3) showing how the information void manifests differently across quadrant 4 conditions and how it maps onto specific types of misinformation encountered on social media platforms.

Textbox 1. Case study 1: prediabetes and "reverse your diabetes" content on TikTok.

- **Clinical scenario:** an individual aged 28 y is diagnosed with prediabetes (hemoglobin A_{1c} [HbA_{1c}]=6.1%) at a routine checkup. She feels completely well. Her clinician advises dietary modification and increased physical activity, with a repeat HbA_{1c} measurement in 6 months.
- **Information void:** for the next 6 months, this patient has no clinical feedback whatsoever. She has no symptoms to monitor, no medication regimen to anchor her daily behavior, and no way to know whether her dietary changes are "working."
- **Digital misinformation pathway:** she searches TikTok for "how to reverse prediabetes." The algorithm surfaces high-engagement videos promising "reverse your prediabetes in 30 days with this one supplement" and "the food your doctor won't tell you about." These videos exploit temporal discounting by promising rapid results and fill the information void with concrete (if unfounded) action plans. The patient purchases an unregulated supplement and follows an elimination diet with no evidence base.
- **Framework application:** the framework predicts this trajectory. Quadrant 4 conditions (asymptomatic + lifestyle managed) create maximal vulnerability because the combination of no somatic feedback and delayed clinical feedback leaves the patient entirely reliant on external sources for validation that their efforts are worthwhile.

Textbox 2. Case study 2: nonalcoholic fatty liver disease (NAFLD) and "liver detox" misinformation.

- **Clinical scenario:** a man aged 32 y is incidentally found to have hepatic steatosis on abdominal ultrasound during workup for another issue. He has mildly elevated alanine aminotransferase. He is advised to lose weight through diet and exercise, with repeat liver function tests and imaging in 12 months.
- **Information void:** the patient has no symptoms and no medication and faces a 12-month wait for clinical feedback. He may not even fully understand what "fatty liver" means or how serious it could become.
- **Digital misinformation pathway:** searching "fatty liver cure" on social media yields a flood of "liver detox" products, "liver cleanse" juice protocols, and testimonials claiming dramatic improvement from unproven supplements. The NAFLD information landscape is particularly prone to commercial exploitation because the term "liver detox" is already embedded in wellness culture, creating a preexisting market for products targeting health-anxious consumers [29].
- **Framework application:** NAFLD exemplifies quadrant 4 vulnerability in its most extreme form—an exceptionally long feedback interval (often 6-12 months), a condition name that is poorly understood by the public, and a commercial misinformation ecosystem that is already well established around "liver health."

Textbox 3. Case study 3: untreated hypertension and supplement-based "natural" blood pressure (BP) cures.

- **Clinical scenario:** an individual aged 25 y has stage 1 hypertension (BP of 135/85 mm Hg) identified at a routine visit. Given his age, low cardiovascular risk score, and absence of target organ damage, his physician recommends a trial of lifestyle modification (sodium restriction, Dietary Approaches to Stop Hypertension diet, and exercise) before initiating medication, with follow-up in 3 months.
- **Information void:** unlike patients initiated on antihypertensive medication (quadrant 3), this patient has no pharmacological anchor. He may or may not own a home BP monitor. If he does, the natural variability in readings may itself become a source of anxiety and confusion.
- **Digital misinformation pathway:** searching "lower blood pressure naturally" yields content promoting magnesium supplements, beet juice "cures," and breathing techniques with exaggerated efficacy claims. Some content explicitly frames medication as harmful, reinforcing the patient's desire to avoid pharmaceuticals and potentially delaying necessary treatment if lifestyle modification proves insufficient.
- **Framework application:** this case illustrates how quadrant 4 vulnerability can have particularly serious consequences when it delays transition to quadrant 3 (pharmaceutical management). The information void does not just expose patients to misinformation—it can actively impede appropriate escalation of care.

XSL•FO
RenderX

Reframing Information Seeking as a Clinical Signal: The Digital Phenotype Concept

In response to the challenge described by the framework, we advocate for a fundamental shift in clinical perspective: reframing active online information seeking from a problematic behavior to a valuable clinical signal, which we term the "digital phenotype." This reconceptualization is firmly grounded in the COM-B model described above [23,30]. From this perspective, a patient who actively searches for information about their new diagnosis is demonstrating psychological capability (understanding the diagnosis), motivation (a desire to act), and an attempt to create opportunity (finding resources for self-management). This proactive behavior maps directly onto theoretical domains framework domains such as "social influences" and "environmental context and resources," which are critical determinants of health behavior but are often underaddressed in time-constrained clinical encounters. Instead of viewing this as noncompliance or a challenge to medical authority, clinicians can interpret it as a positive prognostic indicator.

This concept finds support in a growing body of literature that links online health information seeking with positive health indicators, such as higher patient activation levels, more appropriate use of health services, and greater vaccine uptake [29,31,32]. However, the digital phenotype is a nuanced signal that requires careful interpretation. The same behavior can also stem from confusion, anxiety, or low digital health literacy, which may predispose individuals to seek out and trust unreliable sources [33,34]. Furthermore, some studies have shown that general online searching does not always correlate with adherence to specific clinical recommendations such as cancer screening guidelines [35]. Therefore, the clinical utility of the digital phenotype lies not in a simple binary assessment but in its function as a triage tool. A brief, nonjudgmental inquiry at diagnosis—such as "Many people look for information online after a new diagnosis. Have you found anything helpful or confusing?"—can open a dialogue. This allows clinicians to stratify support, offering foundational digital literacy skills to those who seem overwhelmed or misinformed while providing curated, high-quality resources to those who are already motivated and capable, thereby channeling their proactive energy toward evidence-based pathways.

Evidence-Based Digital Health Interventions

The evidence base supports a multipronged intervention strategy to constructively engage this digital phenotype and fill the information void. The first pillar of this strategy is the direct provision of digital health literacy education. This approach is backed by high-level evidence; a 2023 systematic review and meta-analysis of digital interventions for chronic diseases found a large and statistically significant improvement in eHealth literacy (standardized mean difference=1.22) [33]. Other meta-analyses have corroborated these findings, demonstrating that health literacy interventions can also lead to improvements in health status and self-efficacy [36]. Importantly, these reviews highlight that the most effective interventions are not one-off educational sessions but, rather, those that involve sustained engagement, use appropriate tools, and are integrated with primary care. The feasibility of this approach in the target population has been demonstrated in a recent pilot study of the Get Health'e' intervention, which successfully increased digital health knowledge and confidence among young adults [37]. Real-world implementation of such approaches is already emerging; for example, the Sun Life-KKH LITE Programme in Singapore provides a multidisciplinary, digitally delivered lifestyle intervention for children and families managing obesity integrating virtual coaching, peer WhatsApp communities, and structured physical activity sessions [38].

The second pillar involves the creation and dissemination of curated, developmentally aligned digital resources. It is not enough to teach patients how to evaluate information; health systems must also actively contribute high-quality, engaging content to the digital ecosystem. Currently, a significant gap exists. While medical professionals produce more accurate content, they struggle to compete with the reach and visibility of nonmedical influencers, who often have commercial incentives to promote unproven products or services [6]. This challenge is compounded by platform-specific algorithmic dynamics. On TikTok and similar short-form video platforms, recommendation algorithms prioritize engagement metrics (watch time, shares, and comments) over content accuracy [39]. This creates a structural disadvantage for evidence-based content, which tends to be more nuanced and less sensational than misinformation. Content from high–follower count influencers receives preferential algorithmic amplification regardless of its accuracy, creating a feedback loop in which misleading health claims gain visibility precisely because they are designed to maximize engagement [40]. To close this gap, health care organizations must create content that is not only evidence-based but also platform-native, leveraging formats that resonate with young adults, such as short-form video, and addressing the needs that drive them to these platforms: peer-style success narratives, practical tips, and emotional validation [5].

The third pillar is strategic institutional engagement in digital spaces. The current approach, which largely relies on the ad hoc efforts of individual physicians, is insufficient to counter the tide of coordinated, commercially driven misinformation. Many health care professionals already use social media for professional purposes, including patient education, and many report that it influences their own clinical perceptions and habits [41,42]. A coordinated strategy led by health care systems or professional societies could provide the resources, editorial oversight, and consistent branding necessary to build a trusted presence online. Critically, such strategies must also engage with platform governance. This could involve advocating for algorithmic transparency in health content curation, collaborating with platforms on content-labeling systems that distinguish evidence-based information from user-generated testimonials, and partnering with platform-native creators who have established credibility with young adult audiences to

XSL•FO
RenderX

coproduce content that is both engaging and accurate [43]. Without addressing the structural and algorithmic factors that disadvantage evidence-based content, even the highest-quality clinical material will struggle to reach its intended audience.

Research Priorities and Implementation Challenges

The implications of this framework are far-reaching, demanding action from clinicians, health systems, and digital platforms alike. For frontline clinicians, the immediate takeaway is to shift from a paternalistic stance of warning patients away from the internet to a collaborative one. This involves routinely and nonjudgmentally assessing online information-seeking behaviors; using the digital phenotype concept to gauge readiness for change; and explicitly addressing the information void by setting realistic expectations for biomarker improvement and providing proximate, behavior-linked goals. For health systems and policymakers, our analysis provides a strong rationale for investing in scalable digital health literacy programs and exploring new reimbursement models that compensate for digital education. Furthermore, it highlights the need for clear institutional guidelines on professional boundaries and for policy discussions on the responsibilities of social media platforms in curating a healthier information environment.

Our analysis is characterized by several strengths, including its novel synthesis of infodemiology, behavioral science, and digital health research into a cohesive and actionable framework supported by established theoretical models. By triangulating evidence from diverse sources and grounding the framework in recognized theories (COM-B, temporal discounting, and uses and gratifications theory), we provide a comprehensive perspective on a complex problem. The inclusion of illustrative case studies demonstrates the framework's practical application and clinical relevance. Nevertheless, we acknowledge important limitations. As a viewpoint, this work is conceptual and does not present new primary data. The 2×2 framework deliberately simplifies a complex reality; individual patients' vulnerability is modulated by numerous factors beyond disease characteristics, including cultural context, health system structure, socioeconomic status, digital access, and individual health literacy [27,28]. The framework has been developed primarily with reference to Western health care contexts and social media platforms; its applicability to different cultural settings, health care systems, and digital ecosystems (eg, WeChat-dominated health information environments in China or settings with different patient-health care provider dynamics) requires empirical validation [20]. The digital landscape is in constant flux, meaning that platform-specific findings may quickly become outdated. Moreover, the digital divide remains a critical challenge; interventions that rely on digital access and literacy risk exacerbating existing health disparities if not designed with equity at their core. The evidence base for interventions, while strong for intermediate outcomes such as literacy, requires more research linking these interventions to hard clinical end points.

Conclusions

The intersection of rising asymptomatic chronic diseases in young adults and their universal engagement with social media has created a critical public health challenge. The information void that characterizes these conditions, grounded in the convergence of absent somatic feedback and delayed clinical feedback, makes this population uniquely vulnerable to digital misinformation. However, this challenge also presents an opportunity. By reframing online information seeking as a digital phenotype, we can transform a perceived problem into a powerful tool for patient engagement. The evidence clearly points toward a constructive path forward: a sustained, multipronged strategy that builds digital health literacy, delivers high-quality curated resources, addresses the algorithmic structures that amplify misinformation, and fosters strategic institutional engagement. Future research must now rise to the challenge of validating this framework through prospective studies across diverse health care settings and populations; designing and testing interventions that directly address the psychological drivers of this behavior; and developing novel methods to evaluate their effectiveness in the complex, dynamic digital ecosystem.

Acknowledgments

The authors would like to thank Dr Pearlly Chang (KK Women's and Children's Hospital) and Dr Poon Shi Hui (Neuropsychiatry Associates) for their valuable guidance and input during the preparation of this manuscript.

Funding

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

Authors' Contributions

The manuscript was conceptualized and written by VSME.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Andes LJ, Cheng YJ, Rolka DB, Gregg EW, Imperatore G. Prevalence of prediabetes among adolescents and young adults in the United States, 2005-2016. JAMA Pediatr 2020 Feb 1;174(2):e194498. [doi: 10.1001/jamapediatrics.2019.4498] [Medline: 31790544]
2. Li W, Ng CH, Quek J, et al. The growing prevalence of nonalcoholic fatty liver disease (NAFLD), determined by fatty liver index, amongst young adults in the United States. A 20-year experience. Metab Target Organ Damage 2022;2(4):19. [doi: 10.20517/mtod.2022.24]
3. Ostchega Y, Fryar CD, Nwankwo T, Nguyen DT. Hypertension prevalence among adults aged 18 and over: United States, 2017-2018. NCHS Data Brief 2020 Apr;364(364):1-8. [Medline: 32487290]
4. Stifjell K, Sandanger TM, Wien C. Exploring online health information-seeking behavior among young adults: scoping review. J Med Internet Res 2025 Sep 9;27:e70379. [doi: 10.2196/70379] [Medline: 40925001]
5. Kirkpatrick CE, Lawrie LL. TikTok as a source of health information and misinformation for young women in the United States: survey study. JMIR Infodemiology 2024 May 21;4(1):e54663. [doi: 10.2196/54663] [Medline: 38772020]
6. 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]
7. Dimitroyannis R, Fenton D, Cho S, Nordgren R, Pinto JM, Roxbury CR. A social media quality review of popular sinusitis videos on TikTok. Otolaryngol Head Neck Surg 2024 May;170(5):1456-1466. [doi: 10.1002/ohn.688] [Medline: 38431902]
8. 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]
9. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol. J Am Coll Cardiol 2019 Jun;73(24):e285-e350. [doi: 10.1016/j.jacc.2018.11.003]
10. US Preventive Services Task Force, Davidson KW, Barry MJ, et al. Screening for prediabetes and type 2 diabetes: US Preventive Services Task Force recommendation statement. JAMA 2021 Aug 24;326(8):736-743. [doi: 10.1001/jama.2021.12531] [Medline: 34427594]
11. Watson KB, Wiltz JL, Nhim K, Kaufmann RB, Thomas CW, Greenlund KJ. Trends in multiple chronic conditions among US adults, by life stage, behavioral risk factor surveillance system, 2013-2023. Prev Chronic Dis 2025 Apr 17;22:E15. [doi: 10.5888/pcd22.240539] [Medline: 40245168]
12. Yano Y, Reis JP, Colangelo LA, et al. Association of blood pressure classification in young adults using the 2017 American College of Cardiology/American Heart Association Blood Pressure guideline with cardiovascular events later in life. JAMA 2018 Nov 6;320(17):1774-1782. [doi: 10.1001/jama.2018.13551] [Medline: 30398601]
13. George KM, Maillard P, Gilsanz P, et al. Association of early adulthood hypertension and blood pressure change with late-life neuroimaging biomarkers. JAMA Netw Open 2023 Apr 3;6(4):e236431. [doi: 10.1001/jamanetworkopen.2023.6431] [Medline: 37010868]
14. Suzuki Y, Kaneko H, Okada A, et al. Prediabetes in young adults and its association with cardiovascular health metrics in the progression to diabetes. J Clin Endocrinol Metab 2022 Jun 16;107(7):1843-1853. [doi: 10.1210/clinem/dgac247] [Medline: 35446413]
15. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
16. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85. [doi: 10.2196/jmir.1933] [Medline: 23615206]
17. Pagoto S, Waring ME, Xu R. A call for a public health agenda for social media research. J Med Internet Res 2019 Dec 19;21(12):e16661. [doi: 10.2196/16661] [Medline: 31855185]
18. Cheng D, Ren K, Gao X, et al. Video quality of nonalcoholic fatty liver disease on TikTok: a cross-sectional study. Medicine (Baltimore) 2024 Aug 23;103(34):e39330. [doi: 10.1097/MD.0000000000039330] [Medline: 39183399]
19. Vasistha S, Kanchibhatla A, Blanchette JE, Rieke J, Hughes AS. The sugar-coated truth: the quality of diabetes health information on TikTok. Clin Diabetes 2025;43(1):53-58. [doi: 10.2337/cd24-0042] [Medline: 39829690]
20. Zhang X, Wen D, Liang J, Lei J. How the public uses social media WeChat to obtain health information in China: a survey study. BMC Med Inform Decis Mak 2017 Jul 5;17(Suppl 2):66. [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
21. Golebiewski M, Boyd D. Data voids: where missing data can easily be exploited. Data & Society. 2019. URL: https://datasociety.net/library/data-voids [accessed 2026-03-13]
22. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11. [doi: 10.2196/jmir.1157] [Medline: 19329408]
23. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
24. Hunter RF, Tang J, Hutchinson G, Chilton S, Holmes D, Kee F. Association between time preference, present-bias and physical activity: implications for designing behavior change interventions. BMC Public Health 2018 Dec;18(1):1388. [doi: 10.1186/s12889-018-6305-9]

XSL•FO
RenderX

25. DePalma MT, Heichel KT, O'Leary A. Assessing the relationship between present bias and diabetes self-care behavior. J Health Psychol 2026 Mar;31(3):1001-1014. [doi: 10.1177/13591053251350057] [Medline: 40684384]
26. Katz E, Blumler JG, Gurevitch M. Uses and gratifications research. Public Opin Q 1973;37(4):509. [doi: 10.1086/268109]
27. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 4;18(10):e264. [doi: 10.2196/jmir.6349] [Medline: 27702738]
28. Nutbeam D, Lloyd JE. Understanding and responding to health literacy as a social determinant of health. Annu Rev Public Health 2021 Apr 1;42:159-173. [doi: 10.1146/annurev-publhealth-090419-102529] [Medline: 33035427]
29. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. J Med Internet Res 2021 Jun 11;23(6):e29802. [doi: 10.2196/29802] [Medline: 34043526]
30. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37. [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
31. Smith SG, Pandit A, Rush SR, Wolf MS, Simon C. The association between patient activation and accessing online health information: results from a national survey of US adults. Health Expect 2015 Dec;18(6):3262-3273. [doi: 10.1111/hex.12316]
32. Li H, Li D, Zhai M, Lin L, Cao Z. Associations among online health information seeking behavior, online health information perception, and health service utilization: cross-sectional study. J Med Internet Res 2025;27:e66683. [doi: 10.2196/66683]
33. Verweel L, Newman A, Michaelchuk W, Packham T, Goldstein R, Brooks D. The effect of digital interventions on related health literacy and skills for individuals living with chronic diseases: a systematic review and meta-analysis. Int J Med Inform 2023 Sep;177:105114. [doi: 10.1016/j.ijmedinf.2023.105114] [Medline: 37329765]
34. Yuen E, Winter N, Savira F, et al. Digital health literacy and its association with sociodemographic characteristics, health resource use, and health outcomes: rapid review. Interact J Med Res 2024 Jul 26;13(1):e46888. [doi: 10.2196/46888] [Medline: 39059006]
35. Chou WYS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
36. Galmarini E, Marciano L, Schulz PJ. The effectiveness of visual-based interventions on health literacy in health care: a systematic review and meta-analysis. BMC Health Serv Res 2024 Jun 11;24(1):718. [doi: 10.1186/s12913-024-11138-1] [Medline: 38862966]
37. Manganello JA, Colvin KF, Hadley M, O'Brien K. Get Health'e': a pilot test of a digital health literacy intervention for young adults. Health Lit Res Pract 2024 Oct;8(4):e224-e235. [doi: 10.3928/24748307-20240723-01] [Medline: 39642045]
38. Sun Life-KKH LITE Programme. KK Women's and Children's Hospital. 2026. URL: https://www.kkh.com.sg/content/dam/singhealth-web/kkh/imported-assets/assets/patient-care/areas-of-care/childrens-services/documents/Sun-Life-KKH-LITE-Programme.pdf [accessed 2026-03-13]
39. Karizat N, Delmonaco D, Eslami M, Andalibi N. Algorithmic folk theories and identity: how TikTok users co-produce knowledge of identity and engage in algorithmic resistance. Proc ACM Hum-Comput Interact 2021 Oct 13;5(CSCW2):1-44. [doi: 10.1145/3476046]
40. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552. [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
41. Klee D, Covey C, Zhong L. Social media beliefs and usage among family medicine residents and practicing family physicians. Fam Med 2015 Mar;47(3):222-226. [Medline: 25853534]
42. Hasham MA, Parshuram V, Bandre GR. Social media in healthcare: transforming patient engagement, education, and research dissemination. Int J Acad Med 2025;11(1):4-9 [FREE Full text] [doi: 10.4103/ijam.ijam_48_24]
43. Bautista JR, Zhang Y, Gwizdka J. Healthcare professionals' acts of correcting health misinformation on social media. Int J Med Inform 2021 Apr;148:104375. [doi: 10.1016/j.ijmedinf.2021.104375] [Medline: 33461008]

Abbreviations

COM-B: capability, opportunity, and motivation–behavior

Edited by T Mackey; submitted 25.Oct.2025; peer-reviewed by A Ahmadvand, R Nordgren; revised version received 21.Feb.2026; accepted 02.Mar.2026; published 07.Apr.2026.

Please cite as:
Ekstrom VSM
The Information Void in Asymptomatic Chronic Disease: A Digital Health Framework for Understanding Social Media Health Information Seeking in Young Adults
JMIR Infodemiology 2026;6:e86489
URL: https://infodemiology.jmir.org/2026/1/e86489
doi:10.2196/86489

XSL•FO
RenderX

© Victoria Sze Min Ekstrom. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 7.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Online COVID-19 Misinformation and Infodemic Governance During the Peak Pandemic Period: Scoping Review of Social Media Research From 2020 to 2022

Yang Liu[1], PhD; Zhiying Yue[2], PhD

[1]Department of Computer Information Systems, School of Business, North Carolina Central University, 1801 Fayetteville St, Durham, NC, United States

[2]Digital Wellness Lab, Boston Children's Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Yang Liu, PhD
Department of Computer Information Systems, School of Business, North Carolina Central University, 1801 Fayetteville St, Durham, NC, United States

Abstract

Background: During the COVID-19 pandemic, social media platforms became major channels for public health communication and public sensemaking, but they also enabled the rapid circulation of rumors, false claims, conspiracy narratives, and misleading health information. Research on online COVID-19 misinformation expanded rapidly during the peak pandemic period, yet the evidence base was fragmented across public health, computing, communication, and information science.

Objective: This scoping review synthesizes empirical research on online COVID-19 misinformation on social media during the peak pandemic period from February 2020 to October 2022. It examines how misinformation was conceptualized, which platforms and public health contexts were emphasized, and how computational methods shaped the evidence base for infodemic management. Its distinct contribution is an integrated comparison of conceptual definitions, thematic and platform concentration, the computational research pipeline, and the gap between detection and actionable infodemic governance.

Methods: We conducted a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-informed scoping review of empirical studies indexed in PubMed and IEEE Xplore in November 2022 using COVID-19, social media platforms, and misinformation-related terms. Eligible reports were English-language, peer-reviewed empirical studies published between February 2020 and October 2022 that analyzed social media data using computational, quantitative, or social media analytic methods; the language restriction applied to the publication, not to the underlying dataset. The final synthesis included 100 studies selected from 1688 records.

Results: Within the included corpus, the literature was concentrated around vaccine-related discourse, public health and social measures-related misinformation, and text-based data from X (X Corp), formerly known as Twitter. Studies commonly used misinformation detection, classification, sentiment analysis, topic modeling, and network analysis. Fewer studies examined diffusion pathways, source tracing, cross-platform movement, intervention effectiveness, or platform governance. The evidence base remains more developed for detecting and describing misinformation than for evaluating actionable responses to it.

Conclusions: Peak-pandemic COVID-19 misinformation research produced important computational and empirical knowledge, but it remains detection-centered and unevenly connected to public health action. Future infodemiology research should link computational outputs to intervention design, risk communication, fact-checking workflows, platform moderation, multilingual and multimodal analysis, and the evaluation of infodemic governance strategies.

(JMIR Infodemiology 2026;6:e79434) doi:10.2196/79434

KEYWORDS

COVID-19; misinformation; infodemiology; social media; infodemic management; public health communication; natural language processing; machine learning; scoping review

Introduction

Social media platforms have become central infrastructures for digital information production, circulation, and consumption. Their participatory affordances, algorithmic curation mechanisms, and networked diffusion structures enable rapid information exchange across large and diverse publics [1-3]. At the same time, these same features can accelerate the spread of misinformation, amplify unverified claims, and make misleading information appear credible through repetition, engagement, and social endorsement [4-6]. Online misinformation is not only a problem of false content but also

XSL•FO
RenderX

a broader digital information problem shaped by the interaction of users, platforms, algorithms, institutions, and communication environments [7,8].

The COVID-19 pandemic provides a critical empirical setting for examining this problem. During the pandemic, social media platforms became essential channels for public health communication, institutional messaging, community discussion, and real-time public response [9-12]. However, they also facilitated the rapid circulation of rumors, conspiracy theories, misleading claims, and false information concerning the virus, its origins, prevention measures, treatments, vaccines, and public health policies [13,14]. Because COVID-19 combined high uncertainty, high information demand, rapidly changing scientific evidence, and widespread platform-mediated communication, it offers a particularly important context for understanding how online misinformation is produced, circulated, measured, and governed.

Researchers have responded to this challenge by using social media data and computational methods to study COVID-19 misinformation at scale [15]. Natural language processing (NLP), machine learning (ML), deep learning (DL), topic modeling, sentiment analysis, and network analysis have been used to detect misinformation, classify content, identify themes, examine affective responses, and monitor public discourse [16-20]. These methods have generated valuable knowledge about online misinformation during a major public health crisis. However, the literature remains difficult to synthesize because studies differ substantially in how they define misinformation, which platforms and populations they examine, what forms of social media data they collect, and which computational techniques they apply.

This gap motivates the present peak-pandemic scoping review. This peak-pandemic scoping review integrates technical and public health research by examining COVID-19 misinformation as an infodemiology and online health information domain during the intensive period from February 2020 to October 2022. The review does not claim to cover the entire postpandemic misinformation literature; instead, it focuses on the period in which COVID-19 misinformation research rapidly developed around urgent public health uncertainty, high-volume social media data, and computational monitoring. Specifically, we review 100 empirical studies selected from 1688 records identified through the targeted database searches. We analyze how this literature conceptualizes misinformation, what social media contexts it emphasizes, and how computational methods shape what researchers are able to observe, measure, and act upon. In this review, infodemic governance refers to the sociotechnical arrangements through which platforms, institutions, technologies, and users shape the visibility, credibility, correction, moderation, and public health use of online information. This definition allows us to examine misinformation not only as content to be detected but also as an information problem to be managed across platform, institutional, technological, and public communication settings. The distinctive contribution is the joint analysis of conceptual definitions, thematic and platform coverage, the end-to-end computational research pipeline, and the distance between analytic outputs and governance action within one focused corpus. The review is guided by 3 research questions (RQs):

- RQ1: How was COVID-19 misinformation on social media conceptualized and operationalized in empirical infodemiology and online health information research during the peak pandemic period?
- RQ2: What themes, platforms, populations, and online public health information contexts were emphasized or underexamined in this literature?
- RQ3: How were NLP, ML, and related computational methods used to study online misinformation, and what limitations did these methods create for intervention and infodemic governance?

By addressing these questions, this study makes 3 main contributions:

- It clarifies the peak-pandemic conceptual landscape of online COVID-19 misinformation research by synthesizing how misinformation, disinformation, fake news, rumors, and infodemic-related content were defined and operationalized.
- It maps the empirical structure of the literature by identifying dominant themes, platform concentrations, population coverage, and underexamined online contexts during the 2020 - 2022 review window.
- It reframes detection-centered computational misinformation research as part of a broader sociotechnical infodemic governance agenda, showing why detection approaches need to be complemented by research on diffusion, cross-platform movement, source tracing, intervention effectiveness, platform governance, and public health communication.

A central challenge in online misinformation research is conceptual inconsistency. Across the literature, terms such as misinformation, disinformation, fake news, rumor, conspiracy theory, and infodemic are often used in overlapping ways, even though they refer to related but analytically distinct phenomena. The COVID-19 pandemic brought particular attention to the concept of an infodemic, understood as the rapid spread of misinformation, rumors, fake news, and stigma during disease outbreaks [21]. In this study, we defined misinformation as false or misleading information regardless of whether there is intent to deceive [22]. This definition distinguishes misinformation from disinformation, which refers more specifically to false or misleading information deliberately created or shared to deceive, manipulate, or cause harm.

The distinction between misinformation and disinformation is important because intent, harm, and information accuracy have different implications for measurement, labeling, moderation, and intervention. During the COVID-19 crisis, some individuals unintentionally circulated inaccurate claims or anecdotes, whereas others intentionally promoted deceptive narratives. To establish conceptual clarity, we adopt these definitions throughout the study.

This distinction is further supported by prior efforts to classify misleading content more precisely. For example, the information disorder framework of Wardle [7] identifies multiple forms of

XSL•FO
RenderX

mis- and disinformation, ranging from satire or parody, which may mislead despite lacking harmful intent, to false context and fully fabricated content designed to deceive. Such typologies clarify that problematic information online is not a single phenomenon. Rather, it spans multiple forms that vary in intent, construction, context, and potential harm. This broader conceptualization is important for understanding how COVID-19 misinformation was studied across social media platforms and how different operational definitions shaped the empirical literature.

When describing individual studies, we retained the source authors' labels, including fake news, rumor, misinformation, and disinformation, rather than retroactively relabeling their datasets. These labels are not treated as interchangeable. Misinformation serves as the umbrella term for synthesis, whereas disinformation is reserved for cases with evidence of deceptive intent. Opinions, personal experiences, fears, and hesitancy are not themselves misinformation unless they contain a verifiably false or misleading factual claim.

Social media platforms are central infrastructures for information dissemination. Their role is inherently double-edged: they enable rapid access to diverse information, but they also facilitate the circulation and amplification of unverified rumors and misleading narratives [23]. This challenge is intensified by algorithmic curation, which prioritizes content partly based on user preferences and prior behavior [4]. As a result, users may be disproportionately exposed to information consistent with their existing views, while conflicting evidence receives less visibility. Prior research has linked this dynamic to echo chambers and greater polarization [5,6]. In such environments, misinformation can spread more easily, gain legitimacy through repetition, and become more resistant to correction.

The public health consequences of this dynamic are significant. Misinformation can undermine coordinated societal responses during health crises, thereby worsening epidemic conditions [8]. In a review of health-related misinformation published between 2012 and 2018, Wang et al [23] found that vaccination was the most frequently studied topic. Their review further suggested that social media may intensify vaccine hesitancy by granting misinformation disproportionate visibility and perceived credibility, thereby increasing public skepticism and resistance to vaccination. During COVID-19, this broader problem became especially visible. False and misleading claims concerning the virus's origins, transmission, prevention, treatment, and consequences spread rapidly across platforms. These narratives obscured accurate health guidance, encouraged harmful practices, and aggravated both physical and mental health burdens worldwide [24].

At the same time, the scale and structure of social media data create both opportunity and difficulty for researchers. These datasets are large, heterogeneous, and often multimodal, comprising text, images, and video. While such characteristics make social media valuable for studying public discourse during crises, they also increase analytic complexity and measurement risk. Noise, ambiguity, and misinformation make it difficult to extract valid signals for public health and platform decision-making. In response, researchers have increasingly relied on computational social science approaches [15]. NLP methods, such as sentiment analysis, named-entity recognition, and topic modeling, are used to characterize discourse and identify misinformation-related narratives at scale. ML and DL approaches support detection and classification tasks, while network analysis helps examine diffusion patterns and influence structures. These approaches offer analytical and operational value, including support for warning systems and content moderation. However, their usefulness depends not only on technical performance but also on careful evaluation of their broader sociotechnical implications.

Taken together, these conditions justify a structured and auditable synthesis of research at the intersection of social media, misinformation, and COVID-19 during the peak pandemic period from 2020 to 2022. Such a review can clarify what has been studied, identify which computational approaches have dominated the literature, and reveal where important evidence gaps remain. By doing so, it can provide a stronger foundation for the design, evaluation, and governance of human-centered misinformation management during public health emergencies.

Methods

Study Design and Reporting Framework

This study used a scoping review approach to synthesize a focused corpus of empirical social media studies at the intersection of COVID-19, misinformation, and computational analysis. The objective was to identify conceptual, empirical, and methodological patterns. The review does not estimate aggregate effect sizes. We reported the available search criteria, eligibility rules, and aggregate counts. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) informed the reporting of the search and selection process, to the extent applicable to this non–meta-analytic scoping review (Checklist 1). The available documentation includes database-level identification totals, database-specific eligibility totals before cross-database deduplication, the cross-database duplicate count, and the final included corpus.

Search Strategy Development

Our exploration of the landscape of social media "misinformation" began with an initial exploratory search in November 2022, using digital repositories such as Google Scholar and our digital library databases. Using "misinformation" and "COVID-19" as our starting keywords, we used a dynamic method to adjust our search query. We initially used various query approaches with these terms to capture the relevant literature (n=20). As we progressed, we identified additional keywords that frequently co-occurred in our preliminary findings. These new keywords were then incorporated into our primary search. This evolving approach of keyword discovery and inclusion not only broadened our literature base but also sharpened our research focus, aiding in the precise formulation of our RQs, refinement of terminology, and establishment of specific inclusion and exclusion criteria for our study.

XSL•FO
RenderX

Information Sources and Search Strategy

Building upon the exploratory phase, we conducted a targeted search in PubMed and IEEE Xplore. These 2 databases were selected because they jointly cover 2 major streams of COVID-19 misinformation research relevant to this review. PubMed captures biomedical, public health, and health communication studies that examine misinformation as a public health and infodemic problem [25], whereas IEEE Xplore captures computing, engineering, AI, NLP, and ML studies that examine misinformation through computational detection, classification, and analysis. By combining these 2 repositories, our goal was to construct a focused corpus of empirical studies that reflects the intersection of public health misinformation and computational social media analysis. We acknowledge that this search strategy does not cover all possible information science, communication, or social science databases; therefore, the review should be interpreted as a structured synthesis of a focused multidisciplinary corpus rather than a comprehensive systematic review of all COVID-19 misinformation research.

Eligibility Criteria

During the targeted repository search, we used a tripartite context-platform-topic keyword strategy. The sample was restricted to peer-reviewed empirical studies based on original data and analysis; reviews, commentaries, survey papers, and meta-analyses were excluded. Eligible reports were published in English between February 2020 and October 2022 and used computational, quantitative, or social media analytic methods. The English-language criterion applied to the publication, not to the language of the underlying social media data; studies analyzing non-English datasets remained eligible. Studies were included only when the searchable title, abstract, or keyword/index-term fields contained at least one term from each of the three categories below, subject to the field structure available in each database.

1. Context: the study had to concern itself with the COVID-19 pandemic. The COVID-19 context was identified from the searchable title, abstract, or keyword/index-term fields using the following terms: “covid-19,” “covid19,” “covid,” “coronavirus,” and “acute respiratory syndrome.”
2. Platform: we identified the platform as social media by keywords such as “social media,” “social network,” “Twitter,” “tweet,” “Reddit,” “Facebook,” or “Weibo.” Our search also included keywords for “Instagram,” “Snapchat,” and “YouTube.” Because the primary goal of the current study addressed text-centered social media misinformation research, studies focused solely on nontextual image or video content outside this scope were excluded. The final sample nevertheless included studies addressing misinformation on YouTube (Google LLC; n=2) and Instagram (Meta Platforms, Inc; n=2) when they met the review criteria.
3. Topic: the exact misinformation-related search terms were “misinformation,” “disinformation,” “fake news,” “rumor,” and “infodemic.” No additional adjacent labels, such as “inaccurate information,” “unreliable information,” or “low-quality information,” were included in the targeted search. The search identifies studies that self-described their focus using the specified misinformation vocabulary; studies using only adjacent terminology may have been missed. The reported terms describe the recoverable component terms of the targeted search; a complete database-exported record of the original field-specific query syntax was not retained. Additional details on the search logic, database-specific record counts, supplementary title/abstract search audit, and PRISMA-ScR–informed reporting are provided in Multimedia Appendix 1.

The 2 researchers divided the initial screening tasks by database; one researcher screened records retrieved from IEEE Xplore, and the other screened records retrieved from PubMed. Following the initial screening, every inclusion and exclusion decision was verified by the other researcher against the prespecified eligibility criteria. This verification was conducted sequentially and was not independent of the initial screening decision. Disagreements were resolved through discussion and consensus. No disagreement remained unresolved. Therefore, third-party adjudication was not used. No formal interrater reliability statistic was calculated.

Data Extraction and Synthesis

Each researcher extracted data from the studies retained from their assigned database using the same structured extraction framework. The framework recorded publication year, title, source link, methodological approach, platform, topic, inclusion or exclusion decision, major RQ, specific misinformation topic, sample selection strategy, sample size, statistical analyses, model strengths and weaknesses, and conclusion. Data extraction was divided by database and was not performed independently in duplicate. Accordingly, no formal interrater reliability coefficient, such as Cohen kappa, was calculated.

The thematic synthesis used the common extraction framework as its evidentiary basis. Information recorded under the topic, major RQ, specific misinformation topic, methodological approach, platform, sample selection, statistical analysis, model strengths and weaknesses, and conclusion fields was compared across the included studies and organized around the 3 review questions. The resulting themes were used to identify recurring conceptual definitions, public health topics, platform contexts, data acquisition strategies, computational methods, and governance-relevant gaps. Because the included studies used heterogeneous data sources, analytical methods, and outcome measures, we used a conceptual synthesis and did not estimate aggregate effect sizes. The themes and analytical lenses were not treated as mutually exclusive study categories. The retained records do not document independent duplicate thematic coding or a formal reliability assessment for the thematic synthesis; accordingly, the resulting structure should be interpreted as an author-developed conceptual synthesis rather than a validated study-level multilabel coding scheme. The 2 researchers jointly reviewed the proposed thematic organization and resolved differences in interpretation through discussion and consensus.

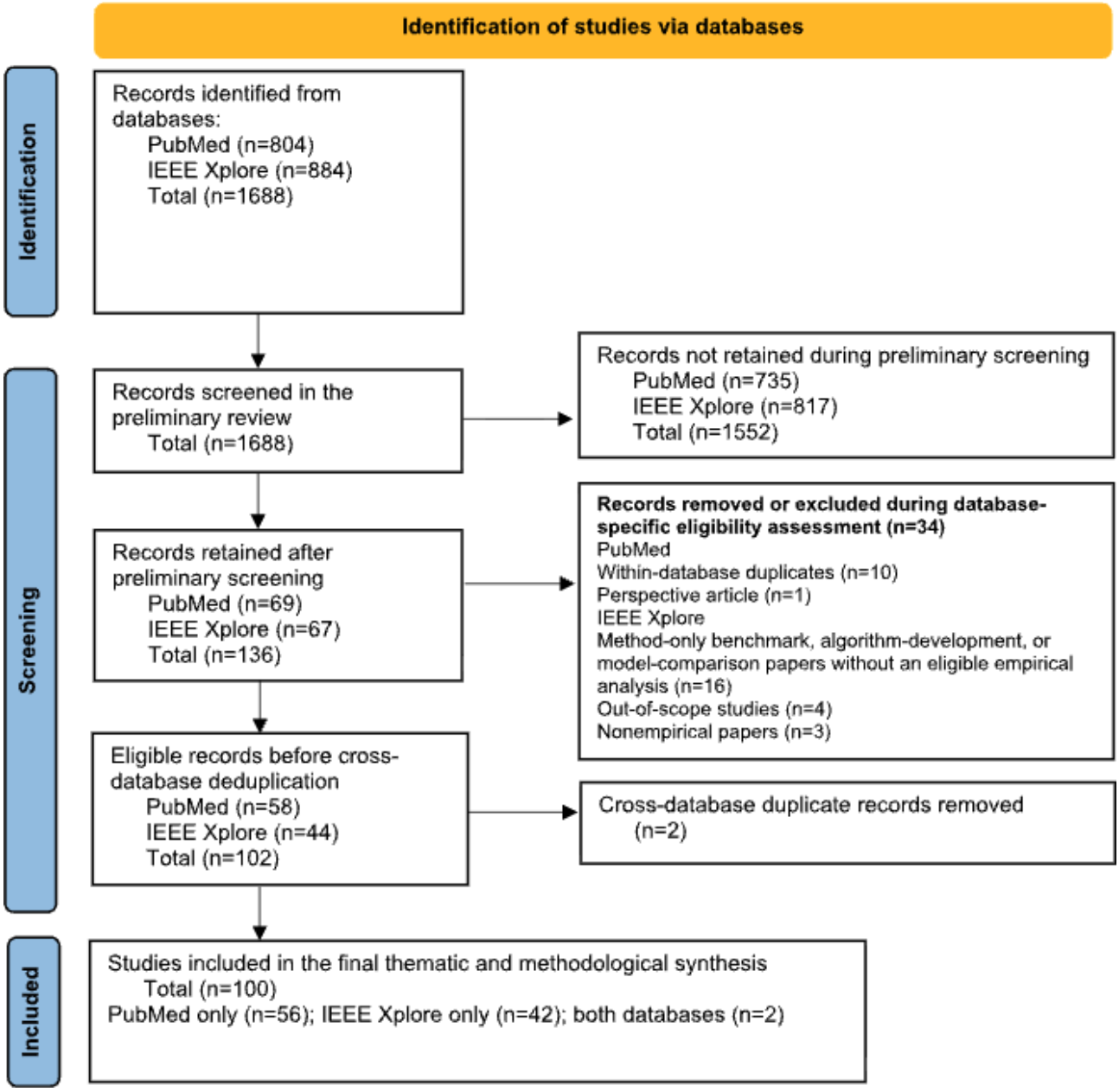
XSL•FO
RenderX

Results

Study Selection and Sample Characteristics

The process of literature search and identification is shown in Figure 1.

Figure 1. Overview of database identification, aggregate screening and eligibility assessment, cross-database deduplication, and final study selection.



The targeted database searches identified 1688 records, comprising 804 records from PubMed and 884 records from IEEE Xplore. Preliminary screening retained 69 PubMed records and 67 IEEE Xplore records for detailed assessment; 735 PubMed records and 817 IEEE Xplore records were not retained at this stage. Within the PubMed set, 10 within-database duplicate records were removed, leaving 59 unique records. One perspective article was subsequently excluded because it did not report eligible original empirical research, leaving 58 eligible PubMed studies. Among the 67 IEEE Xplore records retained after preliminary screening, 23 were excluded during eligibility assessment: 16 were benchmark, algorithm-development, or model-comparison papers lacking an eligible empirical analysis of COVID-19 misinformation on social media; 4 concerned non-COVID-19 topics, general fake news, or platforms or topics outside the prespecified scope; and 3 were theoretical, policy-oriented, research-agenda, or other nonempirical papers. This process left 44 eligible IEEE Xplore studies. After cross-database deduplication removed 2 duplicate records, the final synthesis comprised 100 unique empirical studies: 56 identified only through PubMed, 42 identified only through IEEE Xplore, and 2 identified through both databases. The retained screening documentation supports the reporting of database-level identification totals, aggregate preliminary-screening counts, within-database duplicate removal, principal eligibility exclusion categories, the cross-database duplicate count, and the final included corpus. However, a complete contemporaneous item-level screening

XSL•FO
RenderX

log and record-level exclusion reasons for every record not retained during preliminary screening were not preserved. Accordingly, Figure 1 presents an aggregate summary of study selection and should not be interpreted as a complete PRISMA flow diagram.

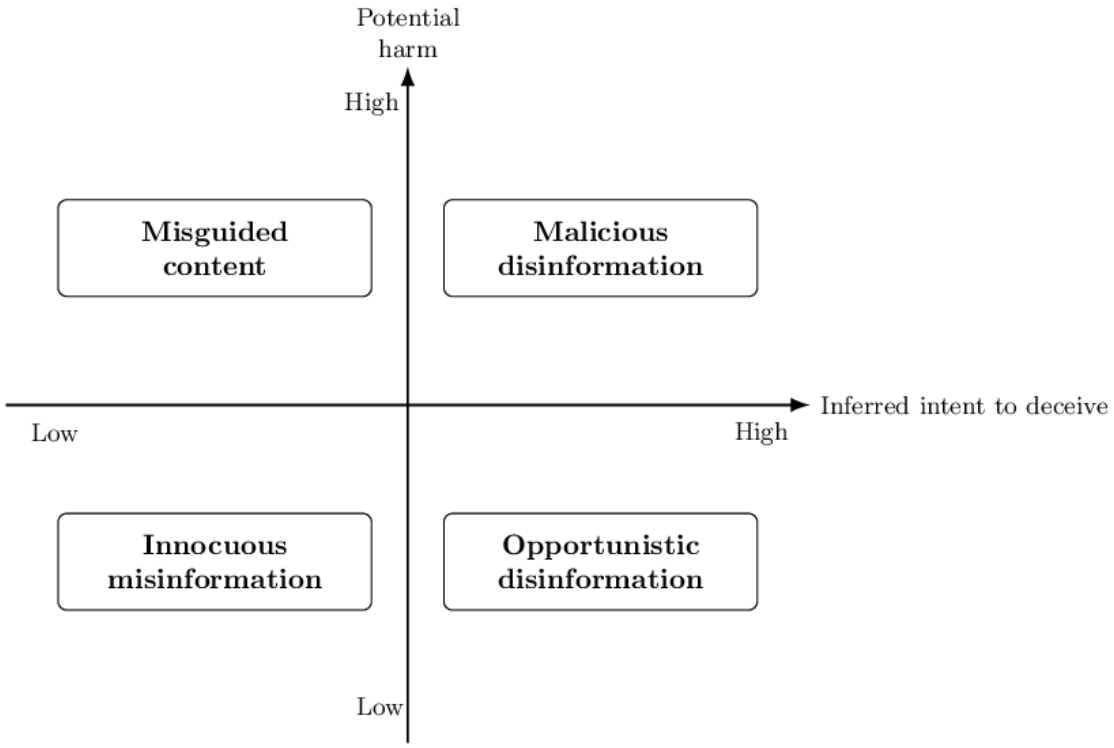
The largest number of articles were published in 2021 (n=44), followed by a comparable number in 2022 (n=42), and fewer articles in 2020 (n=14). Most studies analyzed data in X (X Corp), formerly known as Twitter (n=58, 58%), followed by multiplatform/cross-platform (n=25, 25%) and other individual platforms (Facebook; Meta: n=5; Weibo; Sina.com Technology (China) Co, Ltd: n=4; Reddit; Reddit, Inc: n=3; Instagram; Meta: n=2; YouTube; Google: n=2; LINE; LY Corporation: n=1).

Conceptualization and Typology of Misinformation

Overview

To frame our analysis, it is important to clarify how misinformation is conceptualized across the reviewed studies. We found that authors do not always use consistent definitions or typologies of misinformation and disinformation. Through our review, we identified several categories of misinformation-related content that differ in both the nature of the falsehood and the intent underlying it. Drawing on prior frameworks, particularly the information disorder framework proposed by Wardle [7], we developed a working typology that provides a conceptual framework for interpreting the forms of misinformation and disinformation described in the reviewed literature. It was not used to assign all 100 studies to mutually exclusive categories or to generate category-specific frequencies. The typology classifies content and acts of dissemination; it does not classify unverifiable subjective experiences. Sender intent differentiates misguided, opportunistic, and malicious dissemination when intent can be supported; receiver interpretation or behavior is a downstream outcome and does not itself determine whether content is misinformation. Because intent is often difficult to infer from content alone, these categories should be treated as conceptual distinctions rather than definitive judgments about individual users. Figure 2 presents this typology along 2 conceptual dimensions: intent to deceive and potential harm.

Figure 2. An author-developed conceptual typology of misleading information by inferred sender intent and potential harm.



Innocuous Misinformation

This category encompasses factually false or misleading content shared without the intent to deceive or cause harm [7]. It may arise from satire, parody, reporting errors, humor, or sincere misunderstanding. Personal fears, concerns, opinions, and experiences are not misinformation unless they contain a fact-checkable false or misleading claim. This distinction is particularly important when interpreting vaccine-related social media discourse. For example, Stevens et al [26] examined anxiety, anger, sadness, and incivility in tweets about COVID-19 vaccine mandates but did not assess the factual accuracy of those tweets; therefore, the emotional expressions analyzed in that study should not themselves be classified as misinformation. By contrast, a claim such as “ice cream prevents COVID-19 infection,” if repeated as a joke or in a sincere mistake without deceptive intent, would constitute innocuous misinformation.

Misguided Content

This category involves inaccurate or distorted information shared by individuals who sincerely believe it to be true. Rumors and unverified claims, especially common during crises when

XSL•FO
RenderX

reliable information may be scarce, are central to this category [27]. For example, claims such as "mRNA alters a person's DNA" or "the COVID-19 vaccine contains tissue from aborted fetuses" fall into this category when they are spread by individuals who mistakenly believe them to be accurate.

Opportunistic Disinformation

This category refers to false or misleading information deliberately disseminated to obtain personal, political, professional, or commercial advantage [28-30]. Unlike malicious disinformation, its primary objective is benefit seeking rather than the deliberate infliction of substantial harm, although harmful consequences may still occur [28]. Such content may exploit timely events, public attention, or existing concerns to attract engagement, influence perceptions, generate revenue, or promote interests. Examples include clickbait, exaggerated claims, and commercially motivated false or misleading health or product claims [29,30]. Its placement in the lower-harm quadrant indicates comparatively lower or more indirect potential harm than malicious disinformation; it does not imply that the consequences are necessarily harmless or minor.

Malicious Disinformation

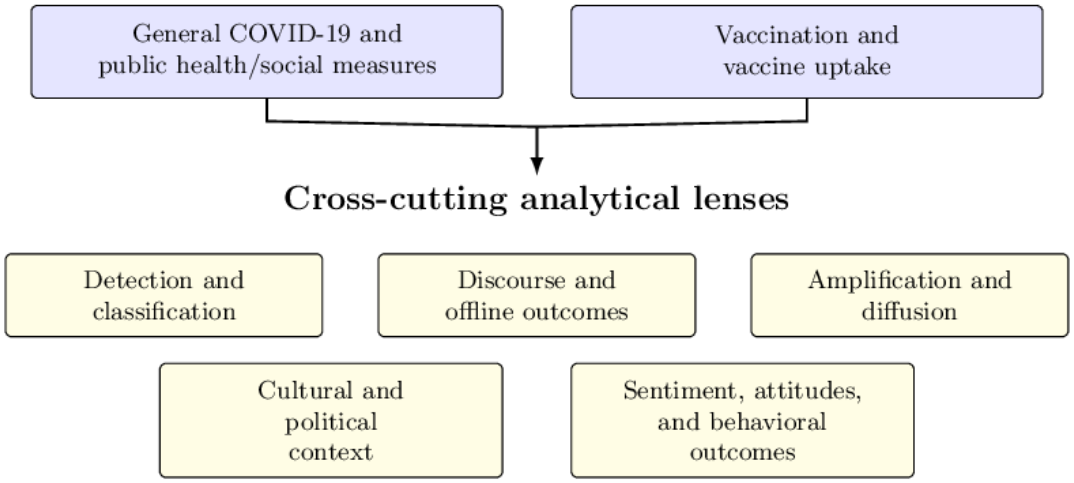
This category refers to content that is knowingly and deliberately disseminated to mislead, distort public perceptions, or cause harm. It may include fabricated content, manipulated content, imposter content, and other misleading messages when there is evidence of deliberate deceptive intent [31]. The false claim linking fifth-generation (5G) mobile networks to the spread of COVID-19 is an example of a fabricated COVID-19 misinformation narrative with the potential to mislead the public and provoke fear [32]. It should be classified as malicious disinformation only when the available evidence supports knowing and deliberate dissemination; the falsity or potential harmfulness of the claim alone does not establish sender intent.

Thematic and Platform Concentration

Overview

To address RQ2, we examined the online public health information contexts and thematic emphases represented in the reviewed studies. Instead of treating the categories as mutually exclusive groups of studies, we organize recurring findings into 2 parallel topic domains (general COVID-19 and public health/social measures and vaccination/vaccine uptake) and a set of cross-cutting analytical lenses. The domains and lenses can overlap conceptually; for example, amplification and cultural context may apply to both topic domains. The available review records do not support reliable per-theme study counts or an overlap matrix, so Figure 3 is presented as a conceptual map rather than a quantitative classification. This distinction is important for infodemiology because the observed concentration of topics and platforms shapes what kinds of misinformation become visible, which communities are studied, and what governance implications can be drawn.

Figure 3. Conceptual map of 2 parallel topic domains and cross-cutting analytical lenses in the reviewed literature; domains and lenses are not mutually exclusive.



General COVID-19 and Public Health/Social Measure–Related Misinformation

Overview

This topic domain captures studies of general COVID-19 misinformation and misinformation concerning public health and social measures, including nonpharmaceutical measures such as mask-wearing, distancing, isolation, and related prevention guidance. It includes research on how misleading information shapes, reflects, or interacts with public attitudes, discourse, and behavior. Studies in this domain commonly use content analysis, sentiment analysis, ML, and network-based approaches to analyze large-scale social media data.

Detection and Classification

This cross-cutting analytical lens focuses on the identification, classification, and interpretation of misinformation at scale. The studies grouped here primarily aim to develop or apply computational approaches that can detect misleading content and characterize its patterns across social media environments. In this sense, their central contribution lies not only in identifying false or misleading content but also in improving

XSL•FO
RenderX

the methodological capacity to analyze misinformation systematically. For example, Bangyal et al [33] proposed a high-accuracy approach for sentiment analysis of COVID-19 fake news on Twitter. Their study compared multiple ML and DL techniques, including naive Bayes (NB), AdaBoost, k-nearest neighbors (KNN), convolutional neural networks (CNN), long short-term memory (LSTM), recurrent neural networks (RNN), and gated recurrent units (GRU). By evaluating a range of models on COVID-19 fake news data, the study illustrates how advanced computational techniques can be used to strengthen large-scale misinformation detection. Rastogi and Bansal [31] developed an integrated approach combining social science theory with computational analysis to distinguish among real information, disinformation, and satire. Their framework is notable because it extends beyond binary authenticity classification to consider communicative intent, using features such as frequency, complexity, uncertainty, sentiment, and subjectivity. Studies examined through this analytical lens deepen understanding of misinformation patterns, improve classification strategies, and clarify the analytical dimensions through which misinformation can be examined during global crises.

Discourse and Offline Outcomes

This cross-cutting analytical lens concerns relationships between online discourse and offline social or public health outcomes, including whether online signals mirror, anticipate, or coincide with behavior, events, or policy-relevant developments. The core premise is that social media data may function as an indicator of broader public response during a health emergency. For instance, Al-Ramahi et al [34] analyzed a large volume of tweets related to antimask discourse during the COVID-19 pandemic and found that increases in negative mask-related tweets were associated with rises in new COVID-19 cases. Their findings suggest that online sentiment may reflect or coincide with consequential behavioral patterns offline. Jang et al [35] analyzed COVID-19-related tweets and found that identified topics and their temporal trajectories aligned closely with public health promotions and interventions. These studies show that social media can serve as a useful lens through which to observe public concerns, reactions, and behavioral tendencies in near real time. This analytical lens also includes work on rumor-specific discourse that gains traction and produces tangible social consequences. Bahja and Safdar [32] examined the rumor linking 5G networks to the spread of COVID-19. Using social network analysis, they identified key actors involved in spreading the claim and the persuasive strategies used to sustain it. This line of work is important because it demonstrates how specific misinformation narratives can move beyond online circulation and contribute to harmful offline outcomes.

Cultural and Political Context

This cross-cutting analytical lens examines how misinformation is shaped by national, cultural, political, and social context. It recognizes that dissemination, uptake, and persistence vary across informational and societal conditions. As a result, comparative analysis becomes essential for understanding the global yet uneven nature of the COVID-19 infodemic. For example, Park et al [36] investigated Twitter discourse related to COVID-19 in 4 Asian countries and found variation in public attention, misinformation patterns, and topic diversity across national contexts. de Oliveira and Albuquerque [37] examined the spread of false information on Twitter in Brazil and highlighted the possible role of cultural context and social isolation in sustaining misinformation within the information ecosystem. These studies suggest that misinformation should not be analyzed as a context-free phenomenon. Instead, its forms and trajectories must be understood in relation to the cultural and social environments in which it is produced, circulated, and interpreted.

Vaccination and Vaccine Uptake–Related Misinformation

Overview

This second topic domain concerns misinformation and related discourse about COVID-19 vaccination, vaccine confidence and hesitancy, and vaccine uptake. It examines not only false or misleading vaccine claims but also the discursive, behavioral, and political environments in which those claims circulate. The following paragraphs emphasize vaccine-related examples while illustrating analytical lenses that can also operate across the general COVID-19 and public health/social measures domain.

Detection and Classification in Vaccine-Related Studies

When applied to vaccination-related research, this analytical lens includes studies that identify, classify, and interpret vaccine-related misinformation and closely related forms of public discourse. A central concern in this line of work is to distinguish misinformation from other types of vaccine-related content, such as opinion, news, or general expressions of hesitancy, to better understand the structure of online vaccine discourse. For example, Weinzierl and Harabagiu [38] developed a graph link-prediction approach for the automated detection of COVID-19 vaccine misinformation. Ke et al [39] analyzed more than 646 million COVID-19-related tweets to investigate public health beliefs about mask-wearing and vaccination. Their study emphasized the role of "high-profile cues to action," such as guidelines and statements from major institutions including the WHO and CDC, in shaping public beliefs and narratives. These studies show that vaccine discourse research is concerned not only with detecting false claims but also with mapping the broader informational and attitudinal landscape within which such claims circulate.

A more specific strand within this analytical lens focuses directly on the detection of vaccine-related misinformation and rumors. Hayawi et al [40] developed ML-based approaches to detect misinformation related to COVID-19 vaccines. Shang et al [41] similarly identified vaccine-related misleading posts on social media using community-contributed COVID-19 fact data (CCFD), while also addressing user privacy concerns. These studies demonstrate the importance of scalable detection methods for identifying misleading vaccine content and supporting timely public health responses.

Amplification and Diffusion in Vaccine-Related Studies

A further analytical lens concerns the mechanisms through which vaccine-related misinformation is amplified on social

XSL•FO
RenderX

media. These studies extend the analysis beyond the identification of misleading content to examine the relational, discursive, and political processes that enable misinformation to spread, gain attention, and reinforce hesitancy. For example, Guntuku et al [42] examined interaction patterns among users and showed how these dynamics can amplify both misinformation and negative sentiment, thereby intensifying vaccine hesitancy. Sauvayre et al [43] applied ML and NLP techniques to French-language tweets about vaccination and found that political considerations played a significant role in shaping vaccine-related discourse on X, formerly known as Twitter. Their findings suggest that vaccine misinformation does not circulate in a neutral informational environment; rather, it is often embedded in politically charged discussions that influence public interpretation and vulnerability to misleading claims. This line of research is important because it shifts attention from misinformation content alone to the social and political conditions that sustain its visibility and influence.

Sentiment, Attitudes, and Behavioral Outcomes in Vaccine-Related Studies

Another analytical lens examines the relationship between online vaccine discourse and real-world vaccination behavior. These studies are especially important because they connect social media analysis to measurable public health outcomes, thereby extending the relevance of online discourse beyond communication patterns alone. Bari et al [44] developed a real-time big data analytics framework using NLP-based sentiment analysis to track vaccine sentiment on Twitter. Their findings showed that negative sentiment increased after vaccine rollout, indicating that public discourse became more polarized over time. At the same time, they found that increases in positive Twitter sentiment were followed by increases in vaccination rates. Cross-correlation analyses across US regions further suggested that Twitter sentiment scores were associated with vaccination rates with an approximate one-week lag. These findings indicate that social media sentiment may function as an early signal of vaccine uptake and highlight the potential value of social media analytics for informing targeted communication and vaccination strategies.

The findings are organized around 2 major topic domains and multiple cross-cutting analytical lenses. This structure is intended to organize the findings; the topic domains and analytical lenses are not mutually exclusive. Cultural context, amplification, and discourse-behavior relationships may apply across both topic domains. Concentration around vaccines and public health/social measures is understandable given the urgency of the pandemic and the significance of vaccine uptake, but it narrows the empirical basis of the corpus. Importantly, the apparent scarcity of multimodal and behavioral evidence partly reflects the eligibility criteria: noncomputational behavioral studies and studies focused solely on nontextual image or video content were excluded. The identified gaps characterize the included corpus, not the entire COVID-19 misinformation literature. Future work should move beyond topic-specific and platform-specific snapshots toward comparative, cross-platform, multilingual, and multimodal analysis.

Computational Approaches as Online Information Processing Tools

Data overview

Having examined the thematic and contextual structure of the literature, we now turn to the computational approaches used to study online COVID-19 misinformation. This section addresses RQ3 by synthesizing how researchers collect, prepare, model, and analyze social media data. We interpret these methods not only as technical tools, but also as online information processing mechanisms that shape what researchers can observe, classify, explain, and potentially govern. Across the reviewed studies, the analytical pipeline generally involves data collection, data preparation, and modeling and analysis.

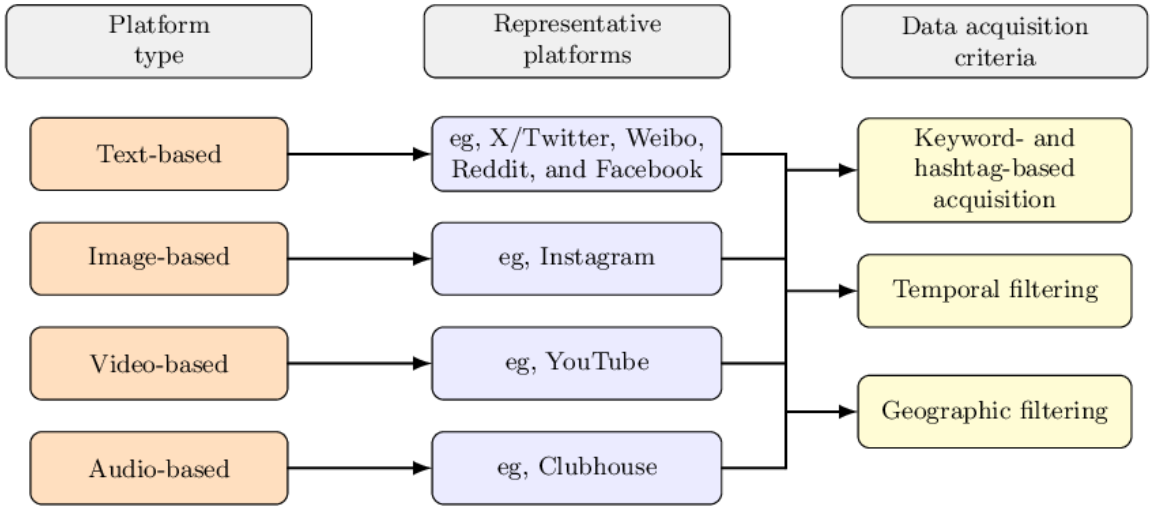
Data Collection

Overview

Data collection is a foundational stage in COVID-19 misinformation research because the choice of platform and acquisition strategy shapes both the scope of analysis and the kinds of claims that can be made. In the reviewed literature, most studies collected their own datasets from social media platforms, while a smaller number relied on publicly available datasets. Across these studies, data collection is structured by 2 core considerations: the selection of platform sources and the criteria used to retrieve relevant data. Together, these decisions determine what forms of discourse enter the dataset and what dimensions of misinformation can subsequently be observed. Figure 4 distinguishes data sources and communication modalities represented in the included studies from modalities that were absent or underrepresented in the corpus. The latter are shown to identify future research opportunities and should not be interpreted as evidence that corresponding studies were included in the reviewed sample.

XSL•FO
RenderX

Figure 4. Social media data sources and communication modalities represented in the reviewed literature, together with underrepresented modalities identified as priorities for future research.



Data Sources: Social Media Platforms

The reviewed studies drew primarily on major text-based social media platforms. X, formerly known as Twitter, is the most frequently used source [45-49], followed by Weibo [50-53], Reddit [54-56], and Facebook [57-60]. These platforms are particularly prominent because they provide large volumes of user-generated textual data that are well-suited to misinformation analysis, discourse modeling, and computational classification.

X, formerly known as Twitter, is the most extensively used platform in the reviewed literature, largely because it offers high-volume, real-time, and relatively accessible discourse for analyzing rumor propagation, coordinated amplification, sentiment, and credibility. For example, Kirn and Hinders [61] analyzed more than 60 million tweets and used ridge count thresholding to identify coordinated tweetstorms and the communities most responsible for amplifying narratives across the platform. Abdelminaam et al [62] applied ML, DL, and NLP techniques to detect COVID-19 fake news across 4 benchmark datasets. Abe et al [63] used Twitter data to study rumors about toilet paper shortages during the pandemic. Similarly, Al-Rakhami and Al-Amri [64] developed a credibility assessment framework for COVID-19 tweets using 6 ML algorithms. These studies show that X/Twitter is especially useful for examining both large-scale diffusion dynamics and automated misinformation detection.

The dominance of X/Twitter also reflects differential platform data availability. During much of the review window, X/Twitter functioned as an easy-data environment because researchers could obtain large quantities of structured public data through its APIs, whereas platforms such as Facebook offered fewer public data and imposed stricter API limits [65]. Therefore, the number of X/Twitter studies should not be interpreted as a direct measure of that platform's public health importance or misinformation prevalence.

Weibo is a major source of data for studies focusing on Chinese-language social media environments. Its importance lies in its ability to capture platform-specific discourse within the Chinese context, where rumor dynamics, language use, and interaction patterns may differ from those found on Western platforms. Lu et al [52] constructed a Weibo COVID-19 rumor dataset and proposed a few-shot learning-based fusion model integrating Bidirectional Encoder Representations from Transformers (BERT) embeddings, bidirectional gated recurrent units (Bi-GRUs), and meta-learning. Shi et al [53] used text, user, interaction, and emotion features to build an extreme gradient boosting (XGBoost) rumor detection model that achieved 91% accuracy. These studies suggest that Weibo is both an important regional data source and a useful setting for methodological experimentation in rumor detection.

Reddit occupies a distinct position because its subreddit structure allows researchers to study misinformation within topic-specific communities. Unlike more open broadcasting platforms, Reddit organizes discourse into semi-bounded spaces, making it well-suited to the analysis of community-level narratives, ideological reinforcement, and conspiracy-oriented discussion. Shahsavari et al [66] used Reddit and 4chan posts to identify emerging COVID-19 conspiracy theories through an ML-based graphical narrative discovery pipeline. Bhattacharyya [67] examined r/liberal and r/conservative to analyze the propagation and acceptance of COVID-related misinformation over time. Together, these studies show that Reddit is particularly valuable for understanding how misinformation becomes embedded in community dynamics and polarized discourse.

Facebook is less frequently represented than X/Twitter in the reviewed literature, but it remains an important platform because of its scale and influence on public communication. Research using Facebook data is particularly useful for examining interactions between misinformation, fact-checking, and user engagement. For instance, Xue et al [68] collected more than 12,000 fact-checking Facebook posts related to COVID-19 vaccines and over 122,000 associated comments. Using NLP and statistical analysis, they examined public attitudes, emotional patterns, and linguistic features in the vaccine fact-checking ecosystem. Their findings highlight the importance

XSL•FO
RenderX

of third-party fact-checkers and suggest declining uncertainty about vaccines over time.

A smaller subset of studies also incorporates multimedia-sharing environments such as Instagram, LINE, and YouTube. These platforms expand the analytical scope of COVID-19 misinformation research by capturing visual, video-based, or message-chain communication that may not be fully represented in text-centered platforms. Zarei et al [69] published a large Instagram dataset and examined engagement patterns, hashtag use, and user behavior. Wang et al [70] studied rumor groups on LINE using a classification-based clustering approach, while Ginossar et al [71] analyzed YouTube videos shared in vaccine-related tweets and found a strong presence of conspiracy-related content. Although the reviewed literature remains predominantly focused on text-based data, these studies indicate that misinformation research is beginning to extend across multiple communication modalities.

Data Acquisition Criteria

Across the reviewed studies, social media data are typically acquired using 3 main criteria: keyword- or hashtag-based queries [72-75], temporal boundaries [76-79], and geographic filters [42,80-82]. These criteria are often used in combination to define the scope of a dataset, align collection procedures with a specific RQ, and facilitate the analysis of large-scale social media data. More importantly, they shape not only what data are collected but also what forms of misinformation become visible to analysis.

Keyword- and hashtag-based acquisition is common in studies focused on topical discourse, rumor detection, and vaccine-related conversation. Hagen et al [77] collected more than 1.3 million tweets using vaccine-related keywords and used network analysis and unsupervised ML to show that vaccine conversations on Twitter were highly polarized. Sharma et al [79] collected 40,000 tweets using COVID-related hashtags and then applied BERT and topic modeling techniques to identify false-information topics. These studies illustrate how keyword selection defines the topical boundaries of the dataset and directly affects the narratives that can be analyzed.

Temporal filtering is important in studies concerned with change over time. Benis et al [76] collected more than 2.7 million tweets across a broad time period to examine how influenza- and COVID-19-vaccination discussions evolved. By delimiting a specific temporal window, such studies can trace changes in attention, sentiment, and misinformation narratives across different stages of the pandemic.

Geographic filtering is especially useful when the goal is to analyze localized misinformation ecosystems or region-specific public discourse. Clamor et al [80] collected tweets from Metro Manila and nearby areas to examine COVID-19 misinformation in the Philippines. Using topic modeling and classification techniques, they identified multiple forms of user activity, including joking, awareness raising, criticism of government responses, and the sharing of preventive advice. This study illustrates how geographic constraints can help isolate context-specific discourse and reveal how misinformation is embedded in local communicative practices.

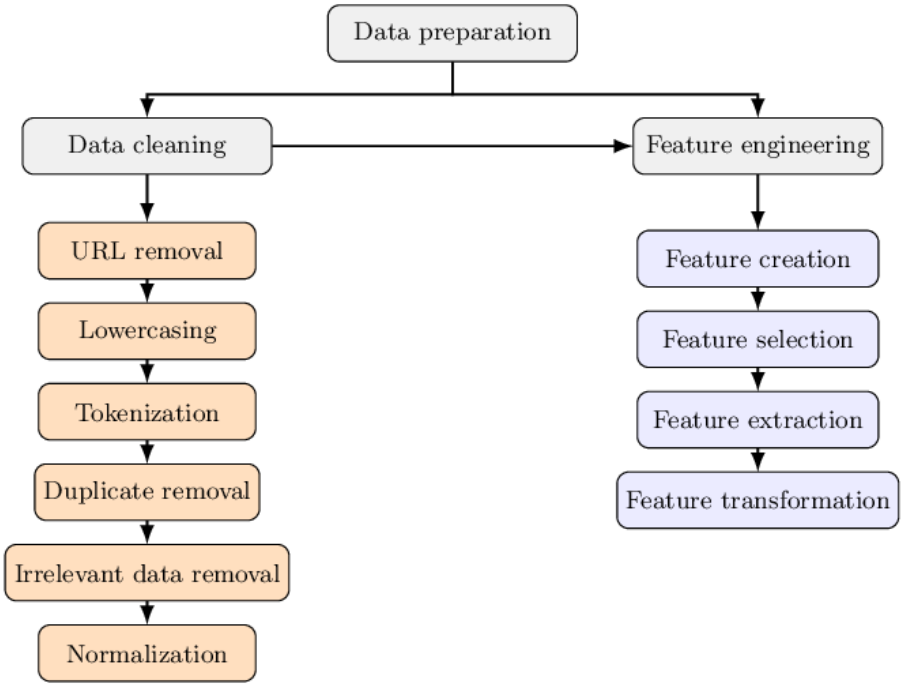
Overall, data acquisition criteria are not neutral procedural choices. They define the scope of the dataset, shape the patterns that become visible to analysis, and constrain the conclusions that can be drawn. Keyword selection determines topical inclusion, temporal boundaries capture discursive change, and geographic filters enable contextual specificity. These criteria form a key part of the methodological and interpretive boundaries of COVID-19 misinformation research.

Data Preparation

Because raw social media data are noisy, heterogeneous, and platform-dependent, data preparation is essential for ensuring that subsequent analysis is both reliable and interpretable. User-generated content frequently contains spelling errors, grammatical inconsistencies, abbreviations, vernacular expressions, and platform-specific conventions, all of which complicate systematic analysis. Figure 5 summarizes the data preparation pipeline identified in the reviewed literature. This pipeline generally involves 2 closely related stages: data cleaning and feature engineering. Although these stages are methodologically distinct, they are tightly interconnected. Data cleaning reduces noise, feature engineering shapes content representation, and subsequent modeling depends on the quality of both.

XSL•FO
RenderX

Figure 5. Key stages of data preparation in the reviewed studies, including data cleaning and feature engineering.



Data cleaning is a foundational step in improving dataset quality and analytical usability and is widely used across the reviewed studies [83]. Common techniques include URL removal, lowercasing, tokenization, duplicate removal, elimination of irrelevant elements such as punctuation, stop words, and digits, as well as normalization procedures such as stemming and lemmatization. These steps are intended to reduce noise, standardize textual inputs, and improve downstream model performance. For example, Sharma et al [84] used a COVID-19 dataset collected from Twitter, Facebook, Reddit, and other platforms to classify real and fake news. Their preprocessing pipeline included removing stop words and special characters, applying lemmatization and stemming, tokenizing text, and converting words into vectors. After these steps, they compared several ML models, with bidirectional LSTM (Bi-LSTM) achieving the highest accuracy. Other studies apply cleaning procedures in more targeted ways depending on the linguistic or structural characteristics of the dataset. Shushkevich and Cardiff [85] removed terms that appeared either too rarely or too frequently, thereby reducing sparsity and overrepresentation. Alsudias and Rayson [86], who worked with Arabic tweets, included the removal of non-Arabic words as part of their cleaning process, while Bahja and Safdar [32] replaced abbreviations with full expressions during preprocessing. These studies show that data cleaning is not a uniform procedure but a context-sensitive stage shaped by language, platform, and research objective.

Feature engineering refers to the process of generating, selecting, extracting, and transforming raw data into representations suitable for ML models [87]. In the reviewed literature, this stage plays a central role in determining what aspects of misinformation become analytically visible. Because misinformation is multidimensional, researchers often construct features that capture not only lexical content but also contextual, structural, behavioral, and psycho-linguistic characteristics of posts and users. Some studies emphasize contextualized semantic representations. Chowdhury et al [88] used fine-tuned BERT and RoBERTa models to extract embeddings from COVID-19-related tweets, combined them through stacked embeddings, and passed them through a BiLSTM module before classification. Other studies extend feature engineering beyond message text. Mourad et al [89] analyzed not only tweet content but also user profiles, metadata, and tweet context, showing how the pandemic was exploited to redirect attention to irrelevant topics and false medical information. Michel et al [90] operationalized tweet utility by classifying messages into communicative categories such as caution and advice, doubt and criticism, and rumors and counter-rumors, thereby linking message type to diffusion dynamics.

A further strand of the literature examines the performance implications of different feature extraction and selection techniques [91]. The study compared principal component analysis (PCA) and chi-square feature selection with bag of words (BoW) and term frequency-inverse document frequency (TF-IDF) representations across multiple ML and DL models, finding that TF-IDF and BoW performed better than the selected features and that full preprocessing yielded the strongest results. Mahbub et al [92] proposed a hybrid feature space combining content-based and psycho-linguistic features to analyze rumor propagation on Twitter, showing that psycho-linguistic attributes improved rumor detection accuracy. These studies indicate that feature engineering is not merely a technical intermediary step. It is a substantive analytical stage in which researchers determine how misinformation will be represented, measured, and ultimately interpreted.

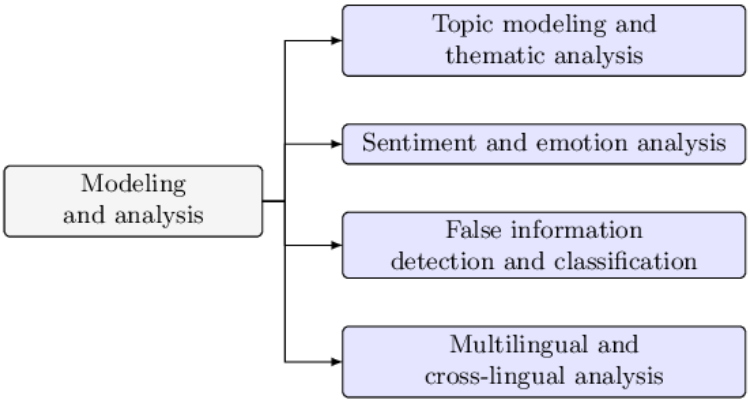
XSL•FO
RenderX

Modeling and Analysis

Modeling Overview

Once data have been collected and transformed into analyzable forms, researchers use NLP and ML methods to address a range of analytical objectives. Across the reviewed literature, these objectives cluster around 4 broad areas: topic modeling and thematic analysis [93,94], sentiment and emotion analysis [95-97], false information detection and classification, including misinformation, disinformation, and fake news detection [98-102], and multilingual or cross-lingual analysis [103-106]. These approaches are not mutually exclusive; rather, they often overlap within the same study. Together, they show that COVID-19 misinformation research spans exploratory, interpretive, predictive, and comparative tasks. Figure 6 presents a high-level taxonomy of the major modeling and analysis approaches identified in the reviewed studies.

Figure 6. A high-level taxonomy of the major modeling and analysis approaches identified in the reviewed studies.



Topic Modeling and Thematic Analysis

Topic modeling and thematic analysis are commonly used to identify recurring narratives, discursive patterns, and user communities associated with COVID-19 misinformation. These approaches are especially useful when the goal is to uncover the substantive structure of discourse rather than merely classify individual posts. Kalantari et al [107] combined topic identification, network analysis, and sentiment analysis to study misinformation-related Twitter discourse and user behavior. Velásquez et al [108] used Latent Dirichlet Allocation (LDA)-based topic modeling across multiple platforms to examine how online hate communities incorporated COVID-19 into malicious narratives and showed how platform moderation policies could influence the spread of such content. These studies demonstrate the value of topic modeling and thematic analysis for mapping discursive structure, tracing narrative evolution, and linking misinformation to broader community dynamics.

Sentiment and Emotion Analysis

A second major line of work focuses on affective expression in social media discourse. Sentiment and emotion analysis are used to examine how fear, distrust, support, skepticism, and related emotional responses emerge in discussions surrounding vaccines, health policies, and misinformation. Prabagar et al [95] applied a multimodule AI framework to analyze emotional shifts in approximately 400,000 vaccine-related posts from Twitter and Reddit. Thakur [109] used the Valence Aware Dictionary and Sentiment Reasoner (VADER) to classify positive and negative sentiment in a large Twitter dataset, while Satu et al [83] used an ML classification model and other models to identify the most frequent topics and their associated sentiment categories across 6 datasets. These studies show that misinformation is not only informational in nature but also affective, and that sentiment-based methods can help capture the emotional conditions under which misinformation gains visibility and influence.

False Information Detection and Classification

A substantial portion of the reviewed literature is devoted to the automated detection and classification of false information, including misinformation, disinformation, rumors, and fake news. Although the terminology varies across studies, the underlying methodological objective is similar: to distinguish misleading from reliable content at scale using textual, structural, network-based, or hybrid features. Ajao et al [110] tested multiple ML and DL models for detecting pandemic-related misinformation and found that support vector machine (SVM) outperformed RNN in their setting. Fedoruk et al [111] proposed the Plebeian algorithm, which combines sentiment analysis and postpopularity with a democratic moderation mechanism, thereby linking detection to platform governance. Gupta and Potika [112] examined the dissemination of fake versus real news through social network analysis and community detection, while Qureshi et al [113] found that hybrid features outperformed individual network or user features in source-based fake news detection. Other studies focus more directly on performance optimization. A total of 3 studies [114-116] compared a wide range of ML, DL, and transformer-based approaches to improve classification accuracy. Iwendi et al [117] demonstrated the value of richer feature spaces and domain-specific embeddings, while Nassif et al [118] showed the importance of language-specific contextual embedding models in Arabic fake news detection. Taken together, these studies reflect a strong technical orientation toward automating authenticity assessment. At the same time, they also suggest that classification performance

XSL•FO
RenderX

alone is insufficient to address the broader sociotechnical challenges of misinformation, particularly when interpretability, governance, and real-world deployment are taken into account.

Multilingual and Cross-Lingual Analysis

The multilingual nature of social media substantially increases the complexity of COVID-19 misinformation research. Because misleading narratives circulate across linguistic communities, approaches developed for English alone are insufficient for capturing the global scope of the infodemic. The reviewed literature includes work in Arabic [86,118,119], Chinese [51,52], Japanese [73], Sinhala [103], Persian [120], and other languages, highlighting the need for methods that are robust beyond English-centric settings. For example, Caldera et al [103] used Sinhala tweets to identify hate-speech trends and applied time-series analysis to predict commenting behavior. Ashraf et al [121] proposed an ML-based rumor detection model for English and Arabic COVID-19 tweets using TF-IDF features and multiple classifiers. In a related study, Almars et al [122] introduced a hybrid attention neural network combining CNN and Bi-LSTM to detect and interpret rumors in Arabic tweets. These studies show that multilingual and cross-lingual analysis requires more than simple language transfer; it often demands tailored feature design, annotation strategies, and model architectures. More broadly, this strand of the literature underscores that methodological robustness in misinformation research depends on sensitivity to linguistic and cultural variation. The English-language eligibility criterion applied to article publication, not to dataset language. The languages named here are illustrative examples from the subsection, not a complete census of the primary dataset languages across all 100 studies; the retained records do not support a reliable study-level language distribution.

Overall, the reviewed literature demonstrates the value of computational methods for processing large-scale online misinformation data. However, it also reveals a detection-centered orientation. Most studies use computational models to identify, classify, cluster, or describe misinformation-related content, while fewer studies connect these outputs to downstream governance tasks such as moderation, warning design, fact-checking workflows, public communication strategies, or intervention evaluation. This creates a gap between information processing and information governance. For online information research, the key challenge is not only whether misinformation can be detected accurately but also whether computational outputs can support explainable, actionable, and context-sensitive responses in real-world information environments. A related limitation is the insufficiently visible role of human-in-the-loop interpretation. Still, several included studies demonstrate ways to integrate human judgment with computational analysis. Dhungana Sainju et al [45] combined supervised ML with independent human coding and qualitative content analysis; Roe et al [74] paired automated sentiment analysis with questionnaire data; and Hagen et al [77] integrated network analysis and ML with manual inductive coding of community discourse. Other studies manually labeled Arabic tweets using categories derived from the Arabic Infectious Disease Ontology [86], evaluated explainable model outputs in a human-participant experiment [116], or developed detailed annotation guidelines for fine-grained, multilabel misinformation coding and assessed interannotator reliability using a second annotator on a 1500-tweet subsample [119]. These examples show that human expertise can contribute during annotation, validation, interpretation, and triangulation rather than only at the downstream governance stage. However, because human involvement was not retained as a separate study-level extraction variable, the review does not estimate its prevalence across the full corpus.

Discussion

Principal Findings

This review shows that COVID-19 misinformation research on social media has produced a substantial body of empirical knowledge, but that this knowledge remains unevenly developed from the perspective of infodemiology and online health information research. The literature has made important progress in detecting misinformation, classifying misleading content, identifying themes, and measuring sentiment. However, it remains less developed in explaining how misinformation circulates across platforms, gains credibility among users, interacts with platform affordances, and can be addressed through public health communication and governance-oriented interventions. These findings suggest that the field needs to move from a detection-centered orientation toward a broader infodemic governance agenda [123-125].

Conceptual Fragmentation Limits Cumulative Knowledge

A first implication concerns conceptual clarity. The reviewed studies use terms such as misinformation, disinformation, fake news, rumor, and infodemic in overlapping ways. Although these terms are related, they differ in intent, accuracy, source, circulation pattern, and potential harm. When studies define and operationalize these concepts differently, it becomes difficult to compare findings, combine evidence, or evaluate the effectiveness of interventions across contexts. For example, a model trained to detect “fake news” may not capture rumor, satire, misleading framing, or sincere but inaccurate health beliefs. Similarly, a study of vaccine hesitancy may capture attitudes that are not necessarily misinformation. Future infodemiology research should adopt clearer conceptual distinctions and make explicit how definitions are translated into data collection, annotation, modeling, and evaluation [7,123]. Accordingly, this review preserves study-specific labels when describing individual papers while using misinformation as an explicitly defined umbrella term in the synthesis; source labels are not assumed to be conceptually equivalent.

Platform Concentration Narrows Understanding of Online Information Disorder

A second implication concerns platform concentration. The reviewed literature is heavily weighted toward text-based and relatively accessible platforms, especially X/Twitter [39,47,77]. This concentration has practical advantages because such platforms provide large-scale public data suitable for computational analysis. However, it also narrows the field’s

XSL•FO
RenderX

understanding of online misinformation. Misinformation also circulates through Facebook groups, Reddit communities, YouTube videos, Instagram posts, TikTok videos, messaging platforms, podcasts, and cross-platform information flows [69-71]. These environments differ in visibility, moderation, affordances, audience structure, and data accessibility. As a result, findings from X/Twitter-based studies should not be assumed to generalize to all online information environments. Future research should develop more comparative and cross-platform designs that examine how misinformation changes as it moves across platforms, media formats, and communities.

Detection-Centered Computation Is Necessary but Insufficient

A third implication concerns the dominant role of detection-centered computation. NLP, ML, DL, topic modeling, and sentiment analysis have made it possible to study online misinformation at scale. These methods are valuable because they can identify misleading content, detect emerging narratives, measure affective patterns, and support large-scale monitoring. However, detection alone does not explain why misinformation becomes credible, how it spreads through social networks, how users interpret it, or which interventions reduce its influence. High classification accuracy also does not necessarily translate into real-world usefulness if model outputs are not interpretable, timely, culturally sensitive, or integrated into governance workflows. Future research should evaluate computational methods not only by predictive performance but also by their actionability for public communication, content moderation, fact-checking, and institutional response [126-128].

Comparison With Prior Work

Prior reviews provide important foundations for understanding COVID-19 misinformation on social media. Gabarron et al [129] conducted a systematic review of COVID-19-related misinformation on social media, Joseph et al [130] produced a scoping review of COVID-19 misinformation on social media, Sanaullah et al [131] systematically reviewed ML applications for COVID-19 misinformation, and Adebesin et al [132] used bibliometric analysis to examine social media's role in health misinformation and disinformation during the pandemic. The present study does not claim to be the first or an exhaustive review. Its distinct contribution is to integrate 4 dimensions within a focused corpus of 100 empirical studies from the peak-pandemic period: conceptual definitions and operationalizations; topic, platform, population, and language coverage; the end-to-end computational pipeline from data acquisition through modeling; and the gap between detection-oriented outputs and actionable infodemic governance. This joint synthesis connects public health infodemiology with computational social media research and clarifies how review scope and methodological choices shape the gaps that become visible.

Implications for Infodemic Governance

Table 1 summarizes the proposed shift from detection-centered misinformation research toward infodemic governance. The findings point toward an infodemic governance agenda for online misinformation research. Such an agenda treats misinformation not only as false content to be detected but also as a sociotechnical phenomenon shaped by information quality, platform design, algorithmic amplification, user interpretation, institutional credibility, and intervention capacity [123,125]. This perspective highlights several priorities. First, research should connect misinformation detection to intervention design, including warning labels, correction strategies, fact-checking systems, moderation policies, and public communication campaigns [126,127]. Second, studies should examine diffusion and source tracing to understand where misinformation originates, how it gains visibility, and how it moves across platforms. Third, future work should address multilingual, multimodal, and community-specific misinformation, especially in underrepresented languages and media formats. Fourth, research should develop evaluation criteria that go beyond accuracy to include explainability, timeliness, fairness, robustness, and usefulness for decision-making.

XSL•FO
RenderX

Table . From detection-centered misinformation research to infodemic governance.

Current limitation	Needed shift	Research opportunities
Inconsistent definitions	Conceptual standardization	Distinguish misinformation, disinformation, rumor, fake news, and infodemic-related content; report how concepts are operationalized in data collection, labeling, and modeling.
Platform concentration	Cross-platform analysis	Compare misinformation across X, formerly known as Twitter, Facebook, Reddit, YouTube, TikTok, Instagram, and messaging platforms; examine cross-platform narrative migration.
Text-centered data	Multimodal information analysis	Study images, videos, memes, audio, hyperlinks, and mixed media misinformation in addition to textual posts.
Detection-centered models	Governance-oriented systems	Link detection outputs to moderation, fact-checking, warning labels, public communication, and intervention workflows.
Static snapshots	Longitudinal and diffusion analysis	Track how misinformation emerges, spreads, mutates, and declines over time; examine source tracing and amplification pathways.
Accuracy-focused evaluation	Actionability-focused evaluation	Evaluate models by explainability, timeliness, fairness, robustness, cultural sensitivity, and usefulness for decision-making.

Limitations

This review has several limitations. First, the review scope was restricted to English-language publications published between February 2020 and October 2022. Studies published after October 2022 were excluded. The publication-language restriction did not exclude studies solely because their underlying social media datasets were non-English. However, dataset language was not retained as a complete study-level variable, so the review cannot quantify the full distribution of languages represented in the corpus. The selected time window captures the peak pandemic period and allows us to examine how online misinformation research developed during an intensive phase of public health uncertainty and platform-mediated communication. However, it does not capture more recent developments in misinformation research, including postpandemic platform changes and the rise of generative AI-enabled misinformation. Second, the literature search was limited to PubMed and IEEE Xplore. These databases were selected to capture the intersection of public health misinformation research and computational social media analysis. However, relevant studies indexed in information science, communication, management, and social science databases may not have been included. For this reason, the review should be interpreted as a structured synthesis of a focused multidisciplinary corpus, not a comprehensive systematic review of all COVID-19 misinformation research. A related methodological limitation concerns the completeness of the retained search and study-selection documentation. Although Figure 1 reports author-verified aggregate identification, preliminary-screening, eligibility, duplicate-removal, and final-inclusion counts, together with the principal eligibility exclusion categories, a complete contemporaneous item-level screening log and record-level exclusion reasons for every record not retained during preliminary screening were not preserved. The complete original database-specific query syntax was also not preserved. Consequently, exact replication of the searches and independent auditing of all study-selection decisions at the individual-record level are limited. Figure 1 should be interpreted as an aggregate study-selection summary, not as a complete PRISMA flow diagram.

Third, the targeted search used the exact topic terms “misinformation,” “disinformation,” “fake news,” “rumor,” and “infodemic.” Adjacent concepts such as inaccurate information, unreliable information, low-quality information, conspiracy theories, propaganda, extremism, information disorder, or platform manipulation were not searched unless an included title or abstract also used one of the specified terms. Relevant studies using only adjacent terminology may have been omitted. Fourth, the review intentionally privileged computational, quantitative, and social media analytic studies and excluded noncomputational behavioral research, as well as studies focused solely on nontextual image or video content. The corpus’s detection-centered and text-centered profile partly reflects the eligibility criteria and should not be interpreted as a prevalence estimate for the entire COVID-19 misinformation field. Fifth, the corpus is weighted toward conversational, text-based platforms, particularly X/Twitter, Reddit, Weibo, and Facebook, whereas image-centric, video-based, audio-based, and closed messaging environments are less represented. Differential platform access and data openness likely contributed to this imbalance, further limiting generalization across online information environments.

Conclusions

This peak-pandemic scoping review synthesized 100 empirical studies on COVID-19 misinformation on social media published between February 2020 and October 2022. We examined this literature as an infodemiology and online health information research domain concerned with how misinformation was

XSL•FO
RenderX

conceptualized, measured, analyzed, and governed in social media environments during an intensive phase of public health uncertainty and platform-mediated communication.

The review identified 3 main patterns. First, the literature applies inconsistent definitions and operationalizations of misinformation, disinformation, fake news, rumor, and infodemic-related content. This conceptual fragmentation limits comparability across studies and complicates cumulative knowledge development. Second, the literature is concentrated around a limited set of topics and platforms, especially vaccine-related discourse and X/Twitter. Although this concentration reflects the urgency, platform data availability, and eligibility criteria of the review period, it narrows understanding of how misinformation operates across different platforms, communities, languages, and media modalities. Third, computational research has focused predominantly on detection, classification, topic modeling, and sentiment analysis. These approaches are valuable, but they provide only a partial basis for understanding and governing online misinformation.

The central contribution of this review is to reframe COVID-19 misinformation research as a sociotechnical infodemic governance problem. Online misinformation is not only false or misleading content to be detected; it is also shaped by platform affordances, algorithmic amplification, user interpretation, institutional trust, information quality, and intervention capacity. Future research should move beyond detection-centered models to examine diffusion, source tracing, cross-platform movement, multilingual and multimodal misinformation, and the effectiveness of governance-oriented interventions. By mapping the conceptual, contextual, and computational patterns of the literature, this study provides a foundation for more robust, inclusive, and actionable research on misinformation in online health information environments.

Generative AI Use

The authors used ChatGPT (OpenAI; GPT-5.6) to assist with proofreading, language editing, formatting, and manuscript consistency checks. All AI-assisted suggestions were reviewed, verified, and incorporated at the authors' discretion. The authors take full responsibility for the final manuscript, including its content, accuracy, and conclusions.

Funding

This work was supported by the IBM-HBCU Quantum Center Faculty Fellowship and the Duke Power Endowed Professorship at the School of Business, North Carolina Central University.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

Conceptualization: YL, ZY
Data curation: YL, ZY
Formal analysis: YL, ZY
Methodology: YL, ZY
Visualization: YL
Writing – original draft: YL, ZY
Writing – review & editing: YL
Funding acquisition: YL
All authors reviewed and approved the final version of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) informed reporting notes.
[DOCX File, 18 KB - infodemiology_v6i1e79434_app1.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 99 KB - infodemiology_v6i1e79434_app2.pdf]

References

1. Boyd D. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications: Routledge; 2010. [doi: 10.4324/9780203876527-8]

XSL•FO
RenderX

2. Treem JW, Leonardi PM. Social media use in organizations: exploring the affordances of visibility, editability, persistence, and association. Ann Int Commun Assoc 2013 Jan;36(1):143-189. [doi: 10.1080/23808985.2013.11679130]
3. Gillespie T, Boczkowski PJ, Foot KA, editors. Media Technologies: Essays on Communication, Materiality, and Society: The MIT Press; 2014. [doi: 10.7551/mitpress/9780262525374.001.0001]
4. Kulshrestha J, Eslami M, Messias J, et al. Quantifying search bias: investigating sources of bias for political searches in social media. Presented at: Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing; Feb 25 to Mar 1, 2017. [doi: 10.1145/2998181.2998321]
5. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 2;118(9):e2023301118. [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
6. Törnberg P. Echo chambers and viral misinformation: modeling fake news as complex contagion. PLoS One 2018;13(9):e0203958. [doi: 10.1371/journal.pone.0203958] [Medline: 30235239]
7. Wardle C. Understanding information disorder: essential guides. : Firstdraft; 2020 URL: https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf [accessed 2026-10-20]
8. Kim A, Moravec PL, Dennis AR. Combating fake news on social media with source ratings: the effects of user and expert reputation ratings. J Manag Inf Syst 2019 Jul 3;36(3):931-968. [doi: 10.1080/07421222.2019.1628921]
9. Garrett L. COVID-19: the medium is the message. Lancet 2020 Mar 21;395(10228):942-943. [doi: 10.1016/S0140-6736(20)30600-0] [Medline: 32171075]
10. Chon MG, Kim S. Dealing with the COVID-19 crisis: theoretical application of social media analytics in government crisis management. Public Relat Rev 2022 Sep;48(3):102201. [doi: 10.1016/j.pubrev.2022.102201] [Medline: 35469268]
11. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: how to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380. [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]
12. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194. [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
13. Cinelli M, Quattrociocchi W, Galeazzi A, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 6;10(1):16598. [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
14. Borges do Nascimento IJ, Pizarro AB, Almeida JM, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022 Sep 1;100(9):544-561. [doi: 10.2471/BLT.21.287654] [Medline: 36062247]
15. Liu Y, Whitfield C, Zhang T, Hauser A, Reynolds T, Anwar M. Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning. Health Inf Sci Syst 2021 Dec;9(1):25. [doi: 10.1007/s13755-021-00158-4] [Medline: 34188896]
16. Joy SKS, Dofadar DF, Khan RH, Ahmed MS, Rahman R. A comparative study on covid-19 fake news detection using different transformer based models. Presented at: 2022 IEEE Symposium on Industrial Electronics & Applications (ISIEA); Jul 16-17, 2022. [doi: 10.1109/ISIEA54517.2022.9873797]
17. Khan S, Hakak S, Deepa N, Prabadevi B, Dev K, Trelova S. Detecting COVID-19-related fake news using feature extraction. Front Public Health 2021;9:788074. [doi: 10.3389/fpubh.2021.788074] [Medline: 35059379]
18. Priya A, Kumar A. Deep ensemble approach for COVID-19 fake news detection from social media. 2021 Presented at: 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN); Aug 26-27, 2021. [doi: 10.1109/SPIN52536.2021.9565958]
19. F IMA, Yaddarabullah Y, Permana SDH. Classification hoax news of COVID-19 on instagram using k-nearest neighbor. Presented at: 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT); Jul 17-18, 2021. [doi: 10.1109/COMNETSAT53002.2021.9530828]
20. Li X, Lu P, Hu L, Wang X, Lu L. A novel self-learning semi-supervised deep learning network to detect fake news on social media. Multimed Tools Appl 2022;81(14):19341-19349. [doi: 10.1007/s11042-021-11065-x] [Medline: 34093070]
21. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/S0140-6736(20)30461-X]
22. Global risks 2013: eighth edition. : World Economic Forum; 2013 URL: https://www3.weforum.org/docs/WEF_GlobalRisks_Report_2013.pdf [accessed 2026-10-20]
23. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552. [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
24. Tasnim S, Hossain MM, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 2020 May;53(3):171-174. [doi: 10.3961/jpmph.20.094] [Medline: 32498140]
25. Valente A, Holanda M, Mariano AM, Furuta R, Da Silva D. Analysis of academic databases for literature review in the computer science education field. Presented at: 2022 IEEE Frontiers in Education Conference (FIE); Oct 8-11, 2022. [doi: 10.1109/FIE56618.2022.9962393]
26. Stevens H, Rasul ME, Oh YJ. Emotions and incivility in vaccine mandate discourse: natural language processing insights. JMIR Infodemiology 2022;2(2):e37635. [doi: 10.2196/37635] [Medline: 36188420]
27. Lanyi K, Green R, Craig D, Marshall C. COVID-19 vaccine hesitancy: analysing Twitter to identify barriers to vaccination in a low uptake region of the UK. Front Digit Health 2021;3:804855. [doi: 10.3389/fdgth.2021.804855] [Medline: 35141699]
28. Fallis D. What is disinformation? Libr Trends 2015 Dec;63(3):401-426. [doi: 10.1353/lib.2015.0014]

XSL•FO
RenderX

29. Tandoc EC, Lim ZW, Ling R. Defining “fake news” a typology of scholarly definitions. Digital J 2018;6(2):137-153. [doi: 10.1080/21670811.2017.1360143]
30. Allcott H, Gentzkow M. Social media and fake news in the 2016 election. J Econ Perspect 2017 May 1;31(2):211-236. [doi: 10.1257/jep.31.2.211]
31. Rastogi S, Bansal D. Disinformation detection on social media: an integrated approach. Multimed Tools Appl 2022;81(28):40675-40707. [doi: 10.1007/s11042-022-13129-y] [Medline: 35582207]
32. Bahja M, Safdar GA. Unlink the link between COVID-19 and 5G networks: an NLP and SNA based approach. IEEE Access 2020;8:209127-209137. [doi: 10.1109/ACCESS.2020.3039168] [Medline: 34812369]
33. Bangyal WH, Qasim R, Rehman NU, et al. Detection of fake news text classification on COVID-19 using deep learning approaches. Comput Math Methods Med 2021;2021:5514220. [doi: 10.1155/2021/5514220] [Medline: 34819990]
34. Al-Ramahi M, Elnoshokaty A, El-Gayar O, Nasralah T, Wahbeh A. Public discourse against masks in the COVID-19 era: infodemiology study of Twitter data. JMIR Public Health Surveill 2021 Apr 5;7(4):e26780. [doi: 10.2196/26780] [Medline: 33720841]
35. Jang H, Rempel E, Roth D, Carenini G, Janjua NZ. Tracking COVID-19 discourse on Twitter in North America: infodemiology study using topic modeling and aspect-based sentiment analysis. J Med Internet Res 2021 Feb 10;23(2):e25431. [doi: 10.2196/25431] [Medline: 33497352]
36. Park S, Han S, Kim J, et al. COVID-19 discourse on Twitter in four Asian countries: case study of risk communication. J Med Internet Res 2021 Mar 16;23(3):e23272. [doi: 10.2196/23272] [Medline: 33684054]
37. de Oliveira DVB, Albuquerque UP. Cultural evolution and digital media: diffusion of fake news about COVID-19 on Twitter. SN Comput Sci 2021;2(6):430. [doi: 10.1007/s42979-021-00836-w] [Medline: 34485922]
38. Weinzierl MA, Harabagiu SM. Automatic detection of COVID-19 vaccine misinformation with graph link prediction. J Biomed Inform 2021 Dec;124:103955. [doi: 10.1016/j.jbi.2021.103955] [Medline: 34800722]
39. Ke SY, Neeley-Tass ES, Barnes M, Hanson CL, Giraud-Carrier C, Snell Q. COVID-19 health beliefs regarding mask wearing and vaccinations on Twitter: deep learning approach. JMIR Infodemiology 2022;2(2):e37861. [doi: 10.2196/37861] [Medline: 36348979]
40. Hayawi K, Shahriar S, Serhani MA, Taleb I, Mathew SS. ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. Public Health 2022 Feb;203:23-30. [doi: 10.1016/j.puhe.2021.11.022] [Medline: 35016072]
41. Shang L, Kou Z, Zhang Y, Chen J, Wang D. A privacy-aware distributed knowledge graph approach to qois-driven COVID-19 misinformation detection. Presented at: 2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS); Jun 10-12, 2022. [doi: 10.1109/IWQoS54832.2022.9812879]
42. Guntuku SC, Buttenheim AM, Sherman G, Merchant RM. Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States. Vaccine 2021 Jul 5;39(30):4034-4038. [doi: 10.1016/j.vaccine.2021.06.014] [Medline: 34140171]
43. Sauvayre R, Vernier J, Chauvière C. An analysis of French-language Tweets about COVID-19 vaccines: supervised learning approach. JMIR Med Inform 2022 May 17;10(5):e37831. [doi: 10.2196/37831] [Medline: 35512274]
44. Bari A, Heymann M, Cohen RJ, et al. Exploring Coronavirus disease 2019 vaccine hesitancy on Twitter using sentiment analysis and natural language processing algorithms. Clin Infect Dis 2022 May 15;74(Suppl_3):e4-e9. [doi: 10.1093/cid/ciac141] [Medline: 35568473]
45. Dhungana Sainju K, Zaidi H, Mishra N, Kuffour A. Xenophobic bullying and COVID-19: an exploration using big data and qualitative analysis. Int J Environ Res Public Health 2022 Apr 15;19(8):4824. [doi: 10.3390/ijerph19084824] [Medline: 35457691]
46. Farrell T, Gorrell G, Bontcheva K. Vindication, virtue, and vitriol: a study of online engagement and abuse toward British MPs during the COVID-19 pandemic. J Comput Soc Sci 2020;3(2):401-443. [doi: 10.1007/s42001-020-00090-9] [Medline: 33225098]
47. Gerts D, Shelley CD, Parikh N, et al. “Thought I’d Share First” and other conspiracy theory Tweets from the COVID-19 infodemic: exploratory study. JMIR Public Health Surveill 2021 Apr 14;7(4):e26527. [doi: 10.2196/26527] [Medline: 33764882]
48. Ng JY, Abdelkader W, Lokker C. Tracking discussions of complementary, alternative, and integrative medicine in the context of the COVID-19 pandemic: a month-by-month sentiment analysis of Twitter data. BMC Complement Med Ther 2022 Apr 13;22(1):105. [doi: 10.1186/s12906-022-03586-1] [Medline: 35418205]
49. Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A. Fine-tuned sentiment analysis of COVID-19 vaccine-related social media data: comparative study. J Med Internet Res 2022 Oct 17;24(10):e40408. [doi: 10.2196/40408] [Medline: 36174192]
50. Han Y. Identification and analysis of opinion leaders for COVID-19 related rumor spreading in social media. : VDE Presented at: CIBDA 2022; 3rd International Conference on Computer Information and Big Data Applications; Mar 25-27, 2022; Wuhan, China p. 1-5 URL: https://ieeexplore.ieee.org/abstract/document/9898926 [accessed 2026-10-20]
51. Leng Y, Zhai Y, Sun S, et al. Misinformation during the COVID-19 outbreak in China: cultural, social and political entanglements. IEEE Trans Big Data 2021 Mar 1;7(1):69-80. [doi: 10.1109/TBDATA.2021.3055758] [Medline: 37974653]

XSL•FO
RenderX

52. Lu HY, Fan C, Song X, Fang W. A novel few-shot learning based multi-modality fusion model for COVID-19 rumor detection from online social media. PeerJ Comput Sci 2021;7:e688. [doi: 10.7717/peerj-cs.688] [Medline: 34497874]
53. Shi A, Qu Z, Jia Q, Lyu C. Rumor detection of COVID-19 pandemic on online social networks. Presented at: 2020 IEEE/ACM Symposium on Edge Computing (SEC); Nov 12-14, 2020. [doi: 10.1109/SEC50012.2020.00055]
54. Liu Y. A comparative study of vector space language models for sentiment analysis using reddit data. Master's thesis.: North Carolina Agricultural and Technical State University; 2020 URL: https://www.proquest.com/openview/092a86ab8b3827cf04f345621b19217f/1?pq-origsite=gscholar&cbl=44156 [accessed 2026-10-20]
55. Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Sentiment analysis of the Covid-19 vaccines on social media. Stud Health Technol Inform 2022 Jun 6;290:1056-1057. [doi: 10.3233/SHTI220265] [Medline: 35673203]
56. Patel A, Meehan K. Fake news detection on reddit utilising countvectorizer and term frequency-inverse document frequency with logistic regression, multinominalnb and support vector machine. Presented at: 2021 32nd Irish Signals and Systems Conference (ISSC); Jun 10-11, 2021. [doi: 10.1109/ISSC52156.2021.9467842]
57. Alouffi B, Alharbi A, Sahal R, Saleh H. An optimized hybrid deep learning model to detect COVID-19 misleading information. Comput Intell Neurosci 2021;2021(1):9615034. [doi: 10.1155/2021/9615034] [Medline: 34790233]
58. Catapang JK, Cleofas JV. Topic modeling, clade-assisted sentiment analysis, and vaccine brand reputation analysis of COVID-19 vaccine-related Facebook comments in the Philippines. Presented at: 2022 IEEE 16th International Conference on Semantic Computing (ICSC); Jan 26-28, 2022. [doi: 10.1109/ICSC52841.2022.00026]
59. Maakoul O, Boucht S, El Hachimi K, Azzouzi S. Towards evaluating the COVID'19 related fake news problem: case of morocco. Presented at: 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS); Dec 2-3, 2020 URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9314356 [accessed 2026-10-20] [doi: 10.1109/ICECOCS50124.2020.9314517]
60. Sear RF, Velasquez N, Leahy R, et al. Quantifying COVID-19 content in the online health opinion war using machine learning. IEEE Access 2020;8:91886-91893. [doi: 10.1109/ACCESS.2020.2993967] [Medline: 34192099]
61. Kirn SL, Hinders MK. Ridge count thresholding to uncover coordinated networks during onset of the Covid-19 pandemic. Soc Netw Anal Min 2022;12(1):45. [doi: 10.1007/s13278-022-00873-0] [Medline: 35368620]
62. Abdelminaam DS, Ismail FH, Taha M, Taha A, Houssein EH, Nabil A. CoAID-DEEP: an optimized intelligent framework for automated detecting COVID-19 misleading information on Twitter. IEEE Access 2021;9:27840-27867. [doi: 10.1109/ACCESS.2021.3058066] [Medline: 34786308]
63. Abe T, Higashi D, Nishiyama D, Toriumi F, Galam S, Ishii A. The toilet paper issue during the COVID-19 crisis. Presented at: 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT); Dec 14-17, 2020. [doi: 10.1109/WIIAT50758.2020.00099]
64. Al-Rakhami MS, Al-Amri AM. Lies kill, facts save: detecting COVID-19 misinformation in Twitter. IEEE Access 2020;8:155961-155970. [doi: 10.1109/ACCESS.2020.3019600] [Medline: 34192115]
65. Trezza D. To scrape or not to scrape, this is dilemma. The post-API scenario and implications on digital research. Front Sociol 2023;8:1145038. [doi: 10.3389/fsoc.2023.1145038] [Medline: 37006635]
66. Shahsavari S, Holur P, Wang T, Tangherlini TR, Roychowdhury V. Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. J Comput Soc Sci 2020;3(2):279-317. [doi: 10.1007/s42001-020-00086-5] [Medline: 33134595]
67. Bhattacharyya S. Classification and analysis of general misinformation and COVID-related misinformation within subreddits of opposing political viewpoints. Presented at: SoutheastCon 2022; Mar 26 to Apr 3, 2022. [doi: 10.1109/SoutheastCon48659.2022.9764051]
68. Xue H, Gong X, Stevens H. COVID-19 vaccine fact-checking posts on Facebook: observational study. J Med Internet Res 2022 Jun 21;24(6):e38423. [doi: 10.2196/38423] [Medline: 35671409]
69. Zarei K, Farahbakhsh R, Crespi N, Tyson G. Dataset of coronavirus content from Instagram with an exploratory analysis. IEEE Access 2021;9:157192-157202. [doi: 10.1109/ACCESS.2021.3126552]
70. Wang AW, Lan JY, Wang MH, Yu C. The evolution of rumors on a closed social networking platform during COVID-19: algorithm development and content study. JMIR Med Inform 2021 Nov 23;9(11):e30467. [doi: 10.2196/30467] [Medline: 34623954]
71. Ginossar T, Cruickshank IJ, Zheleva E, Sulskis J, Berger-Wolf T. Cross-platform spread: vaccine-related content, sources, and conspiracy theories in YouTube videos shared in early Twitter COVID-19 conversations. Hum Vaccin Immunother 2022 Dec 31;18(1):1-13. [doi: 10.1080/21645515.2021.2003647] [Medline: 35061560]
72. Kwok SWH, Vadde SK, Wang G. Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: machine learning analysis. J Med Internet Res 2021 May 19;23(5):e26953. [doi: 10.2196/26953] [Medline: 33886492]
73. Niu Q, Liu J, Kato M, et al. Public opinion and sentiment before and at the beginning of COVID-19 vaccinations in Japan: Twitter analysis. JMIR Infodemiology 2022;2(1):e32335. [doi: 10.2196/32335] [Medline: 35578643]
74. Roe C, Lowe M, Williams B, Miller C. Public perception of SARS-CoV-2 vaccinations on social media: questionnaire and sentiment analysis. Int J Environ Res Public Health 2021 Dec 10;18(24):13028. [doi: 10.3390/ijerph182413028] [Medline: 34948638]

XSL•FO
RenderX

75. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining physicians' opinions on social media to obtain insights into COVID-19: mixed methods analysis. JMIR Public Health Surveill 2020 Jun 18;6(2):e19276. [doi: 10.2196/19276] [Medline: 32421686]
76. Benis A, Chatsubi A, Levner E, Ashkenazi S. Change in threads on Twitter regarding influenza, vaccines, and vaccination during the COVID-19 pandemic: artificial intelligence-based infodemiology study. JMIR Infodemiology 2021;1(1):e31983. [doi: 10.2196/31983] [Medline: 34693212]
77. Hagen L, Fox A, O'Leary H, et al. The role of influential actors in fostering the polarized COVID-19 vaccine discourse on Twitter: mixed methods of machine learning and inductive coding. JMIR Infodemiology 2022;2(1):e34231. [doi: 10.2196/34231] [Medline: 35814809]
78. Liu Y, Anwar M. Learning programming in social media: an nlp-powered reddit study. Presented at: 2022 Fourth International Conference on Transdisciplinary AI (TransAI); Sep 19-21, 2022. [doi: 10.1109/TransAI54797.2022.00015]
79. Sharma U, Pandey P, Kumar S. A transformer-based model for evaluation of information relevance in online social-media: a case study of Covid-19 media posts. New Gener Comput 2022;40(4):1029-1052. [doi: 10.1007/s00354-021-00151-1] [Medline: 35035023]
80. Clamor TDS, Solano GA, Oco N, Catapang JK, Cleofas J, Isip-Tan IT. Identification and analysis of COVID-19-related misinformation tweets via kullback-leibler divergence for informativeness and phraseness and biterm topic modeling. Presented at: 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC); Feb 21-24, 2022. [doi: 10.1109/ICAIIC54071.2022.9722623]
81. Liu Y, Yue Z, Anwar M. Monkeypox at-a-glance from google trends and reddit. Presented at: 2022 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); Nov 17-19, 2022; Arlington, VA p. 166-167 URL: https://ieeexplore.ieee.org/abstract/document/9983632 [accessed 2026-10-20]
82. Whitfield C, Liu Y, Anwar M. Surveillance of COVID-19 pandemic using social media: a reddit study in north carolina. Presented at: Proceedings of the 12th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics; Aug 1-4, 2021. [doi: 10.1145/3459930.3469550]
83. Satu MS, Khan MI, Mahmud M, et al. TClustVID: a novel machine learning classification model to investigate topics and sentiment in COVID-19 tweets. Knowl Based Syst 2021 Aug 17;226:107126. [doi: 10.1016/j.knosys.2021.107126] [Medline: 33972817]
84. Sharma MK, Kumar P, Rasool A, Dubey A, Mahto VK. Classification of actual and fake news in pandemic. Presented at: 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC); Nov 11-13, 2021. [doi: 10.1109/I-SMAC52330.2021.9640639]
85. Shushkevich E, Cardiff J. Detecting fake news about covid-19 on small datasets with machine learning algorithms. Presented at: 2021 30th Conference of Open Innovations Association (FRUCT); Oct 27-29, 2021. [doi: 10.23919/FRUCT53335.2021.9599970]
86. Alsudias L, Rayson P. Social media monitoring of the COVID-19 pandemic and influenza epidemic with adaptation for informal language in Arabic Twitter data: qualitative study. JMIR Med Inform 2021 Sep 17;9(9):e27670. [doi: 10.2196/27670] [Medline: 34346892]
87. Liu Z, Lin Y, Sun M, editors. Representation Learning for Natural Language Processing: Springer Nature; 2020. [doi: 10.1007/978-981-99-1600-9]
88. Chowdhury AN, Guha S, Amin N, Khan SI. Exploiting ensemble of transformer models for detecting informative tweets. Presented at: 2021 IEEE 18th India Council International Conference (INDICON); Dec 19-21, 2021. [doi: 10.1109/INDICON52576.2021.9691734]
89. Mourad A, Srour A, Harmanani H, Jenainati C, Arafeh M. Critical impact of social networks infodemic on defeating Coronavirus COVID-19 pandemic: Twitter-based study and research directions. IEEE Trans Netw Serv Manage 2020;17(4):2145-2155. [doi: 10.1109/TNSM.2020.3031034]
90. Michel K, Smith M, Brown B, King M, Dozier G. A study of social network messages during the COVID-19 infodemic: salient features and the propagation of information types. Presented at: SoutheastCon 2021; Mar 10-13, 2021. [doi: 10.1109/SoutheastCon45413.2021.9401830]
91. Narra M, Umer M, Sadiq S, et al. Selective feature sets based fake news detection for COVID-19 to manage infodemic. IEEE Access 2022;10:98724-98736. [doi: 10.1109/ACCESS.2022.3206963]
92. Mahbub S, Pardede E, Kayes ASM. COVID-19 rumor detection using psycho-linguistic features. IEEE Access 2022;10:117530-117543. [doi: 10.1109/ACCESS.2022.3220369]
93. Hua Y, Jiang H, Lin S, et al. Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. J Am Med Inform Assoc 2022 Sep 12;29(10):1668-1678. [doi: 10.1093/jamia/ocac114] [Medline: 35775946]
94. Micallef N, He B, Kumar S, Ahamad M, Memon N. The role of the crowd in countering misinformation: a case study of the COVID-19 infodemic. Presented at: 2020 IEEE International Conference on Big Data (Big Data); Dec 10-13, 2020. [doi: 10.1109/BigData50022.2020.9377956]

XSL•FO
RenderX

95. Prabagar K, Srikandabala K, Loganathan N, et al. Investigating COVID-19 vaccine messaging in online social networks using artificial intelligence. Presented at: 2022 15th International Conference on Human System Interaction (HSI); Jul 28-31, 2022. [doi: 10.1109/HSI55341.2022.9869484]
96. Ramakrishnan K, Balakrishnan V. SentiMage: a sentiment-image-based covid-19 health misinformation detection using machine learning. Presented at: 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME); Nov 16-18, 2022. [doi: 10.1109/ICECCME55909.2022.9987818]
97. Smith GY, Kabban CMS, Hopkinson KM, Oxley ME, Noel GE, Cheng H. Sensor fusion for context analysis in social media covid-19 data. Presented at: NAECON 2021 - IEEE National Aerospace and Electronics Conference; Aug 16-19, 2021. [doi: 10.1109/NAECON49338.2021.9696396]
98. Elhadad MK, Li KF, Gebali F. Detecting misleading information on COVID-19. IEEE Access 2020;8:165201-165215. [doi: 10.1109/ACCESS.2020.3022867] [Medline: 34786288]
99. Röchert D, Shahi GK, Neubaum G, Ross B, Stieglitz S. The networked context of COVID-19 misinformation: informational homogeneity on YouTube at the beginning of the pandemic. Online Soc Netw Media 2021 Nov;26:100164. [doi: 10.1016/j.osnem.2021.100164] [Medline: 34493994]
100. Varshney D, Vishwakarma DK. Analysing and identifying crucial evidences for the prediction of false information proliferated during covid-19 outbreak: a case study. Presented at: 2021 8th International Conference on Smart Computing and Communications (ICSCC); Jul 1-3, 2021. [doi: 10.1109/ICSCC51209.2021.9528205]
101. Hande A, Puranik K, Priyadharshini R, Thavareesan S, Chakravarthi BR. Evaluating pretrained transformer-based models for covid-19 fake news detection. Presented at: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC); Apr 8-10, 2021. [doi: 10.1109/ICCMC51019.2021.9418446]
102. Sari AM, Ariyani NF, Ahmadiyah AS. Evaluating the preliminary models to identify fake news on COVID-19 tweets. Presented at: 2021 13th International Conference on Information & Communication Technology and System (ICTS); Oct 20-21, 2021. [doi: 10.1109/ICTS52701.2021.9607996]
103. Caldera HMM, Meedin GSN, Perera I. Time series based trend analysis for hate speech in twitter during COVID 19 pandemic. Presented at: 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer); Nov 4-7, 2020. [doi: 10.1109/ICTer51097.2020.9325491]
104. Kar D, Bhardwaj M, Samanta S, Azad AP. No rumours please! a multi-indic-lingual approach for covid fake-tweet detection. Presented at: 2021 Grace Hopper Celebration India (GHCI); Feb 19 to Mar 12, 2021. [doi: 10.1109/GHCI50508.2021.9514012]
105. Saghayan MH, Ebrahimi SF, Bahrani M. Exploring the impact of machine translation on fake news detection: a case study on persian tweets about covid-19. Presented at: 2021 29th Iranian Conference on Electrical Engineering (ICEE); May 18-20, 2021. [doi: 10.1109/ICEE52715.2021.9544409]
106. Vogel I, Meghana M. Detecting fake news spreaders on twitter from a multilingual perspective. Presented at: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA); Oct 6-9, 2020. [doi: 10.1109/DSAA49011.2020.00084]
107. Kalantari N, Liao D, Motti VG. Characterizing the online discourse in twitter: users' reaction to misinformation around COVID-19 in twitter. Presented at: 2021 IEEE International Conference on Big Data (Big Data); Dec 15-18, 2021. [doi: 10.1109/BigData52589.2021.9671740]
108. Velásquez N, Leahy R, Restrepo NJ, et al. Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms. Sci Rep 2021 Jun 15;11(1):11549. [doi: 10.1038/s41598-021-89467-y] [Medline: 34131158]
109. Thakur N. MonkeyPox2022Tweets: a large-scale Twitter dataset on the 2022 Monkeypox outbreak, findings from analysis of Tweets, and open research questions. Infect Dis Rep 2022 Nov 14;14(6):855-883. [doi: 10.3390/idr14060087] [Medline: 36412745]
110. Ajao O, Garg A, Da Costa-Abreu M. Exploring content-based and meta-data analysis for detecting fake news infodemic: a case study on COVID-19. Presented at: 2022 12th International Conference on Pattern Recognition Systems (ICPRS); Jun 7-10, 2022. [doi: 10.1109/ICPRS54038.2022.9854058]
111. Fedoruk B, Nelson H, Frost R, Fucile Ladouceur K. The plebeian algorithm: a democratic approach to censorship and moderation. JMIR Form Res 2021 Dec 21;5(12):e32427. [doi: 10.2196/32427] [Medline: 34854812]
112. Gupta K, Potika K. Fake news analysis and graph classification on a covid-19 twitter dataset. Presented at: 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService); Aug 23-26, 2021. [doi: 10.1109/BigDataService52369.2021.00013]
113. Qureshi KA, Malick RAS, Sabih M, Cherifi H. Complex network and source inspired COVID-19 fake news classification on Twitter. IEEE Access 2021;9:139636-139656. [doi: 10.1109/ACCESS.2021.3119404] [Medline: 35211361]
114. Fatima R, Samad Shaikh N, Riaz A, et al. A natural language processing (NLP) evaluation on COVID-19 rumour dataset using deep learning techniques. Comput Intell Neurosci 2022;2022(1):6561622. [doi: 10.1155/2022/6561622] [Medline: 36156967]

XSL•FO
RenderX

115. Khan Y, Thakur S. Fake news detection of south african COVID-19 related tweets using machine learning. Presented at: 2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD); Aug 4-5, 2022. [doi: 10.1109/icABCD54961.2022.9856272]
116. Ayoub J, Yang XJ, Zhou F. Combat COVID-19 infodemic using explainable natural language processing models. Inf Process Manag 2021 Jul;58(4):102569. [doi: 10.1016/j.ipm.2021.102569] [Medline: 33776192]
117. Iwendi C, Mohan S, Khan S, Ibeke E, Ahmadian A, Ciano T. Covid-19 fake news sentiment analysis. Comput Electr Eng 2022 Jul;101:107967. [doi: 10.1016/j.compeleceng.2022.107967] [Medline: 35474674]
118. Nassif AB, Elnagar A, Elgendy O, Afadar Y. Arabic fake news detection based on deep contextualized embedding models. Neural Comput Appl 2022;34(18):16019-16032. [doi: 10.1007/s00521-022-07206-4] [Medline: 35529091]
119. Obeidat R, Gharaibeh M, Abdullah M, Alharahsheh Y. Multi-label multi-class COVID-19 Arabic Twitter dataset with fine-grained misinformation and situational information annotations. PeerJ Comput Sci 2022;8:e1151. [doi: 10.7717/peerj-cs.1151] [Medline: 36532803]
120. Parvizimosaed M, Esnaashari M, Damia A, Bahmanyar R. Using supervised learning models for creating a new fake news analysis and classification of a COVID-19 dataset: a case study on COVID-19 in iran. Presented at: 2022 8th International Conference on Web Research (ICWR); May 11-12, 2022. [doi: 10.1109/ICWR54782.2022.9786244]
121. Ashraf N, Nayel H, Taha M. A comparative study of machine learning approaches for rumors detection in covid-19 tweets. Presented at: 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC); May 8-9, 2022. [doi: 10.1109/MIUCC55081.2022.9781707]
122. Almars AM, Almaliki M, Noor TH, Alwateer MM, Atlam E. HANN: hybrid attention neural network for detecting Covid-19 related rumors. IEEE Access 2022;10:12334-12344. [doi: 10.1109/ACCESS.2022.3146712]
123. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11. [doi: 10.2196/jmir.1157] [Medline: 19329408]
124. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206. [doi: 10.2196/16206] [Medline: 32310818]
125. Tangcharoensathien V, Calleja N, Nguyen T, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659. [doi: 10.2196/19659] [Medline: 32558655]
126. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
127. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 2020 Jul;31(7):770-780. [doi: 10.1177/0956797620939054] [Medline: 32603243]
128. Pierri F, Perry BL, DeVerna MR, et al. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. Sci Rep 2022 Apr 26;12(1):5966. [doi: 10.1038/s41598-022-10070-w] [Medline: 35474313]
129. Gabarron E, Oyeyemi SO, Wynn R. COVID-19-related misinformation on social media: a systematic review. Bull World Health Organ 2021 Jun 1;99(6):455-463A. [doi: 10.2471/BLT.20.276782] [Medline: 34108756]
130. Joseph AM, Fernandez V, Kritzman S, et al. COVID-19 misinformation on social media: a scoping review. Cureus 2022 Apr;14(4):e24601. [doi: 10.7759/cureus.24601] [Medline: 35664409]
131. Sanaullah AR, Das A, Das A, Kabir MA, Shu K. Applications of machine learning for COVID-19 misinformation: a systematic review. Soc Netw Anal Min 2022;12(1):94. [doi: 10.1007/s13278-022-00921-9] [Medline: 35919516]
132. Adebesin F, Smuts H, Mawela T, Maramba G, Hattingh M. The role of social media in health misinformation and disinformation during the COVID-19 pandemic: bibliometric analysis. JMIR Infodemiology 2023 Sep 20;3(1):e48620. [doi: 10.2196/48620] [Medline: 37728981]

Abbreviations

5G: fifth-generation
BERT: Bidirectional Encoder Representations from Transformers
Bi-GRUs: bidirectional gated recurrent units
Bi-LSTM: bidirectional long short-term memory
BoW: bag of words
CCFD: community-contributed COVID-19 fact data
CNN: convolutional neural network
DL: deep learning
GRU: gated recurrent units
KNN: k-nearest neighbors
LDA: Latent Dirichlet Allocation

XSL•FO
RenderX

LSTM: long short-term memory
ML: machine learning
NB: naïve Bayesian
NLP: natural language processing
PCA: principal component analysis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RNN: recurrent neural network
RQ: research question
SVM: support vector machine
TF-IDF: term frequency-inverse document frequency
VADER: Valence Aware Dictionary and Sentiment Reasoner
XGBoost: extreme gradient boosting

Edited by T Purnat; submitted 17.Jun.2026; peer-reviewed by C Voegeli, E Wilhelm; revised version received 02.Sep.2026; accepted 23.Sep.2026; published 29.Oct.2026.

Please cite as:
Liu Y, Yue Z
Online COVID-19 Misinformation and Infodemic Governance During the Peak Pandemic Period: Scoping Review of Social Media Research From 2020 to 2022
JMIR Infodemiology 2026;6:e79434
URL: https://infodemiology.jmir.org/2026/1/e79434
doi:10.2196/79434

© Yang Liu, Zhiying Yue. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 29.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Impact of Social Media Videos on Quantitative Health Outcomes: Systematic Review

Fangyue Chen[1], MBBCHIR, MPH, MA; Tricia Tay[2], MBChB, MSc; Hannah Thould[3], BMBCh, BA; Chatchaya Nangsue[4]; Simon Dryden[1], MBChB, MPH, PhD; Amish Acharya[1], BSc, MBBS, PhD; Ara Darzi[1], MD; Kate Grailey[1], BMBCh, BA, PhD

[1]Fleming Initiative, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[2]Helix Centre, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[3]University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, United Kingdom

[4]Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

Corresponding Author:
Fangyue Chen, MBBCHIR, MPH, MA
Fleming Initiative, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

Abstract

Background: Social media has transformed the landscape of health communication. Video content can optimally activate our cognitive systems, enhance learning, and deliver accessible information. Evidence has suggested the positive impact of videos on health knowledge and health-related behaviors, yet the impact of social media videos on quantitative health outcomes is underresearched. Evaluating such outcomes poses unique challenges in measuring exposure and outcomes within internet-based populations.

Objective: We aimed to evaluate the impact of social media videos on quantitative health outcomes, examine methodologies used to measure these effects, and describe the characteristics of video interventions and their delivery.

Methods: In accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, MEDLINE, Embase, Web of Science, CINAHL, and Google Scholar were searched. Studies were eligible if they were original research evaluating long-form social media video interventions addressing any health-related condition, delivered via social media platforms, and reported quantitative health outcomes. The primary outcome was the effect of social media videos on quantitative health outcomes. Additional outcomes included participant characteristics, video features, delivery methods, and the use of theoretical frameworks. A narrative synthesis was conducted. A subgroup meta-analysis was performed to synthesize health outcomes mentioned in 2 or more studies with sufficient homogeneity. Risk of bias assessment was conducted using Cochrane Risk of Bias 2, ROBINS-I, or National Institutes of Health Quality Assessment Tool, depending on the study design. One reviewer screened titles and abstracts. Two reviewers independently conducted full-text screening, data extraction, and risk of bias assessment.

Results: A systematic search was conducted on October 25, 2023, and was updated on June 12, 2025, yielding a total of 41,172 records after duplicate removal. Sixteen studies were included, involving 4158 participants. Mental health–related conditions were the most studied (10 studies). Most video interventions were delivered via YouTube (12 studies). Studies have reported that video interventions were associated with significant improvements in peri-procedural anxiety, mood, and physical activity levels, although most findings were limited to individual studies with variable methodological quality. Three studies that developed videos with user input and theoretical frameworks significantly impacted study-specific primary outcomes. A subgroup meta-analysis demonstrated a significant moderate impact of online video interventions in improving peri-procedural anxiety (standard mean difference=0.57, 95% CI 0.09 - 1.05). All but one study showed some concern or high risk of bias.

Conclusions: We demonstrated a potential positive impact of social media videos on quantitative health outcomes, notably in improving peri-procedural anxiety. Videos developed with user input and theoretical frameworks significantly impacted study-specific primary outcomes. Nevertheless, there is the need to shift focus toward measuring physical health–related outcomes and to develop better designed, innovative methodologies to measure the impact that can better simulate the social media environment.

Trial Registration: PROSPERO CRD42023474648; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023474648

(JMIR Infodemiology 2026;6:e77578) doi:10.2196/77578

KEYWORDS

social media; online video; health outcome; PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

Introduction

Social media can be defined as "websites and computer programs that allow individuals to communicate and share information, opinions, pictures, videos, and other formats on the internet" [1]. Its growing presence within the health care context has facilitated communication between the public, patients, and health care professionals, as exemplified by its usage during the COVID-19 pandemic for organizations to disseminate information to the public [2,3] and for providing peer support spaces for patients suffering from a range of conditions, such as cancer and mental health illnesses [4,5].

One form of social media communication is through online videos, such as those hosted on YouTube, the largest video–based platform with 2.5 billion monthly active users globally [6]. Videos can simultaneously present complementary visual and auditory information, which optimally activate our cognitive systems and lead to better learning, as shown by previous meta-analyses of the effectiveness of learning with multimedia [7,8]. In addition, videos ensure consistency in the delivery of educational content and can be designed to cater to individuals with low literacy [9]. Viewers watching the videos may have a perceived greater autonomy over learning, as they can control the progress of the video [10], while video creators can incorporate video editing processes to improve learning experience [10,11].

Social media videos differ from traditional formats by their potential for a wide reach and the added level of interactivity from predominantly 1-way to 2-way forms of communication [12,13]. Within various health care disciplines, such videos have been developed to tackle a range of issues, including improving community health literacy, facilitating community building by patients sharing their illness journey [14], communicating research [15], and medical education [2]. An example of a popular health-related social media video includes a video of antibiotic pills singing "keep antibiotics working," released as part of an antimicrobial resistance awareness campaign by UK Health Security Agency in 2017 [16]. While the main intention of the campaign was to garner online pledges from individuals to agree to undertake actions relating to antibiotic usage on the dedicated campaign website, the number of views from the video—over 1.3 million times to date—has exceeded the number of pledges the campaign has so far gained, and social media has directed the greatest level of traffic to the campaign website [9,16,17].

Despite the potential of health-related social media content, including those delivered through video formats, in impacting individuals' health, as well as how they interact with health care services globally, how such impact translates to changes in individuals' health-related outcomes remains under-researched. Health outcomes can be defined by health impacts resulting from a condition, event, or intervention [18]. These can be measured clinically (blood pressure, laboratory testing), self-reported (eg, quality of life measures), or observed (eg, changes in gait) [19]. Discerning such impact will enable more tailored governance, adaptation, and integration of such tools, so as to better understand and harness the value of social media within health care systems [20,21].

Previous systematic reviews have examined the impact of health-related videos with limited inclusion of and emphasis on the rise of social media videos. They have largely assessed the effectiveness of video interventions in terms of improvement in learning outcomes, such as knowledge and skills, and their impact on health-related behaviors, such as attendance to disease screening and lifestyle modifications (eg, smoking cessation) [10,22]. The impact of health-related social media videos beyond the openly available engagement analytics is relatively under-researched [23]. To our knowledge, no systematic review has collectively examined the impact of social media videos on quantitative health outcomes. Challenges in conducting trials to test the impact of online interventions include the difficulties in identifying an appropriate control, ensuring the control group has no access to the intervention on a publicly available platform, and collecting measurable outcomes from an audience who accesses content over the internet [24,25].

This systematic review aims to explore and synthesize evidence on the impact of social media online videos on quantitative health outcomes. We began by synthesizing the methodology used to evaluate the measurable health impact of online social media videos and in turn explored whether there was evidence that such videos could have an impact on quantitative health outcomes.

Methods

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Checklist 1)[26]. The study protocol was registered on PROSPERO (CRD42023474648).

Eligibility Criteria

Studies were eligible for inclusion if they reported the use of an online video intervention, defined as "pre-recorded multimedia that combine moving images and audio" found on a publicly available, freely accessible online social media platform that allows online communication between users, and discussed any health-related conditions, targeting any individuals with or without health conditions [10]. Currently, there are 2 formats of videos on social media—the traditional long-form videos and "Shorts," the latter defined by vertical videos that are less than 180 seconds in duration and are presented to users passively and sequentially, with which users can choose to watch or swipe away [27]. While "Shorts" have gained global popularity, their brief, varied, and uncorrelated features have been associated with the development of addiction and impaired attention span, thus negatively impacting users [28,29]. Given different ways in which audience interact with long-form or "Shorts" videos, which can translate to the heterogeneous mechanisms in which different video forms can impact audience's health-related behaviors and outcomes, this systematic review only included studies that described videos in long-form. In addition, if a study described several interventions, the online video would need to be the main intervention, rather than a part of the intervention package. To

XSL•FO
RenderX

illustrate, studies would be included if they described a video intervention that was sent to participants through an SMS text, or if they described a video that included additional materials that were collectively presented onto the social media platform (eg, linked within the video description section). If the video and other information formats were delivered separately, so that it would be difficult to discern whether the study outcomes resulted from the video alone, the study would not be eligible. Studies were eligible if they measured quantitative health outcome or outcomes, including condition-specific outcomes, such as mini-mental status examination in dementia, and more general health outcomes, such as quantitative measures of quality of life [30,31]. Self-reported outcome measures using validated reporting tools were included. Studies with or without a comparator or comparators were included. There were no restrictions on the publication year and language. Literature published in Chinese was translated by a native Chinese speaker. All other languages were translated by Google Translate [32].

Studies were included if they were original research studies including randomized controlled studies, pre-post study design, nonrandomized controlled trials (RCTs), and surveys regarding the impact of the intervention. Review papers, such as narrative reviews, overviews, systematic reviews, and meta-analyses, and informal publication types, such as case studies, commentaries, letters to the editor, editorials, meeting abstracts, and proceeding papers, were excluded given the focus on primary research and the need for full methodological details.

Information Sources

The following databases were searched: MEDLINE OVID (1946 to date of search), Embase Ovid (1947 to date of search), Web of Science Core Collection (1970 to date of search), and CINAHL EMBSCO (1981 to date of search). In addition, Google Scholar was searched and ranked by relevance. The first 1000 results of this search were reviewed, in line with Google Scholar's capabilities [33].

Search Strategy

Search terms were developed with a librarian experienced in conducting systematic searches. Using the PICO (population, intervention, comparator, outcome) strategy, population was humans of any age groups; intervention was online video as defined in the eligibility criteria; comparator included no intervention, standard intervention, or any other interventions that can impact health outcomes; outcome included any quantitative health–related outcomes. Two main search terms "online video" and "health" were used, which best reflected the PICO strategy. In addition, major online social media video platforms that hosted predominantly long-form videos were included as search terms in place of "online video," including "YouTube," "Vimeo," "Dailymotion," and "Facebook Watch." The full search strategy can be found in Multimedia Appendix 1.

Selection Process

All papers from the systematic search were imported into Covidence, a systematic review reference management system [34]. Screening took place in 2 stages. One reviewer (FC) performed abstract and title screening, selecting papers that investigated online videos and had mentioned the measure of quantitative health–related outcomes as defined in the eligibility criteria above. If it was unclear whether the outcome or outcomes met the eligibility criteria, the papers would be included for full-text screening.

Two reviewers (FC and HT or TT or CN) then performed blinded full-text screening against the full eligibility criteria, selecting papers that discussed online videos on open-sourced social media platforms and measured quantitative health outcomes. Disagreement between the 2 reviewers was resolved by consensus and a third reviewer when necessary.

Study Outcomes

The primary outcome of the systematic review was the impact of the health-related video intervention on quantitative health–related outcomes as specified by each paper.

Additional outcomes included the following: study participant characteristics and sample size; video intervention characteristics, including the method of intervention delivery to the participants, social media platform used, video length, video content, and the use of any theory or frameworks in its development; and the quantitative measure of change in the health-related outcome.

Data Extraction

Data extraction took place on the Covidence platform. Two reviewers (FC and HT or TT) performed the full data extraction process in a blinded and independent manner. Disagreement was resolved by consensus and a third reviewer when necessary.

Data collection followed the study outcomes as described above. In addition, the following data were collected: study information, study setting, study design, protocol registration, ethical approval status, funding sources, conflict of interest, and statistical analysis methods.

Risk of Bias Assessment

As it was expected that the review would include studies of different designs, appropriate tools were selected to assess the risk of bias according to the anticipated methodologies used by the studies. For RCTs, the Cochrane Risk of Bias 2 was used [35]. For nonrandomized studies, the ROBIN-I tool was used [36]. For cross-sectional studies, the National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-Sectional studies was considered more appropriate given it is specifically tailored for observational studies; therefore, selected [37]. This was in addition to the methods described in the registered protocol.

Two reviewers (FC and HT or TT) independently evaluated the risk of bias. Disagreement was resolved by consensus and a third reviewer when necessary.

Narrative Synthesis

Given the anticipated wide range of health conditions and outcomes encompassed within this search, data were synthesized using a narrative synthesis approach according to the study outcomes: the participants described in the study, the health condition of interest, the video intervention, the methods of delivering the video to the participants, video characteristics,

XSL•FO
RenderX

the comparator, the outcome measures of interest, and the effect of the video on the specific outcomes.

Quantitative Synthesis

The included studies were evaluated for their suitability for meta-analysis. If 2 or more studies were sufficiently homogeneous in terms of participants, interventions, and health outcome measures to provide a meaningful summary, meta-analysis would be performed. The DerSimonian and Laird random effects models with the inverse variance method were used to generate the summary measures of effect in the form of standard mean difference (SMD) to account for similar outcomes measured using different assessment tools [38,39]. SMD was calculated using change from baseline or point measure mean values and SDs for intervention and control groups for each study with relevant outcome data [39]. For studies that did not report SD of changes from baseline, this was imputed from baseline and final SDs using the standard methods described, assuming the correlation coefficient of 0.5 [39]. When necessary, mean values were standardized to reflect the direction of the scale [39]. For study outcomes derived from multiple outcome measures, the most reported measure across studies was selected. If there was no clear indication of the relative importance of the measures, these were combined into a single effect size by averaging the SMDs and calculating the variance, accounting for the correlation between outcome, assuming a correlation coefficient of 0.5 [39]. The resulting composite SMD and variance were used in the subgroup meta-analysis. Statistical heterogeneity would be examined using the standard methods described [38,39]. The I^2 was used to quantify the magnitude of statistical heterogeneity between studies, where I^2 of 30% to 60% represents moderate and I^2 of >60% represents substantial heterogeneity [38]. A meta-analysis would be performed using the R *meta* package [40].

For all interventional studies, an albatross plot was constructed to allow the *P* values to be interpreted in the context of study sample size. The contour lines of the albatross plots were formed by hypothetical effect sizes [41]. *P* values were calculated from the SMD values using the Wald test [42]. In addition, different colors were used to facilitate the visualization of outcomes by subgroup. The albatross plot was made using the R *metap* package [43].

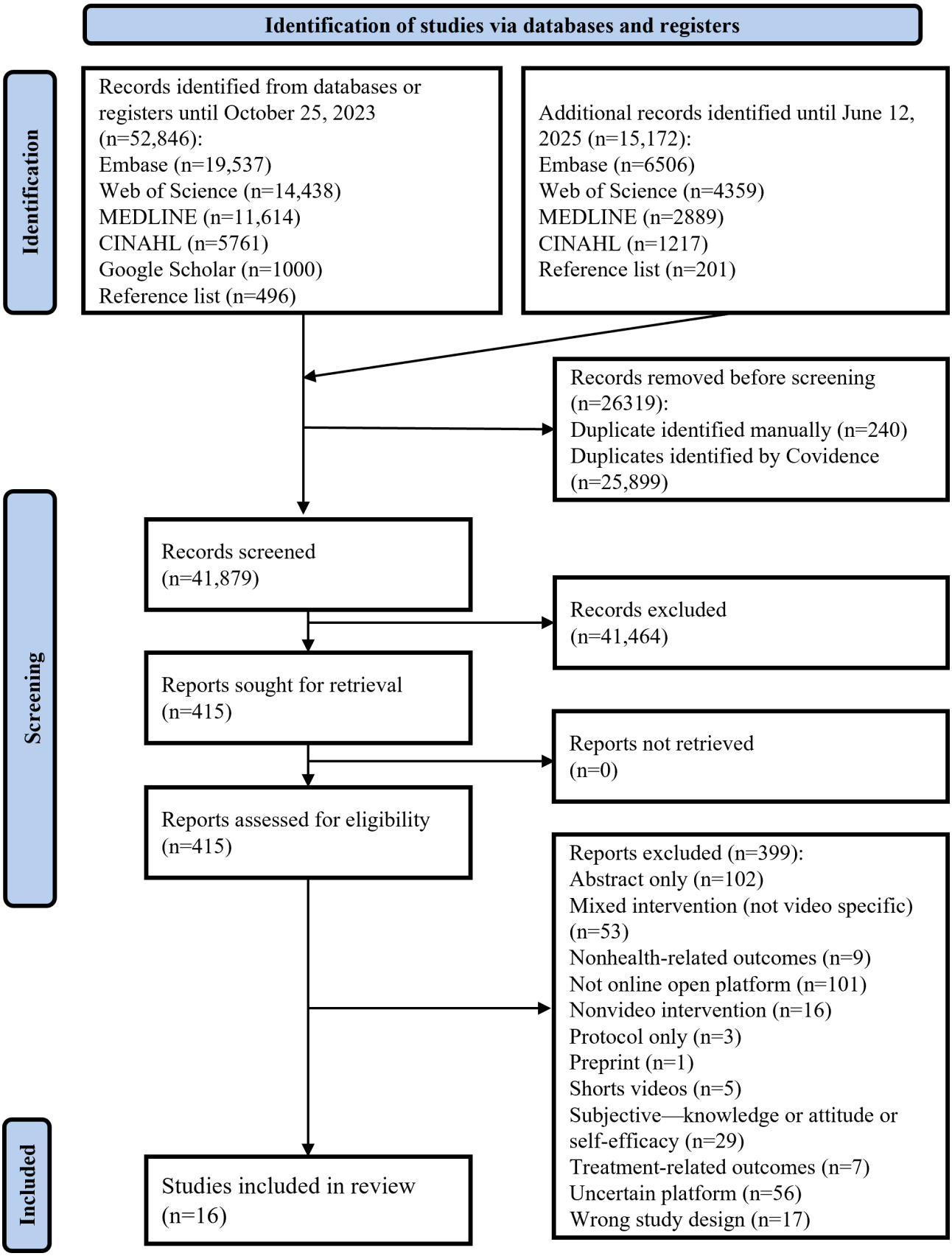
Results

Study Selection

A systematic search was conducted on October 25, 2023, and was updated on June 12, 2025, yielding a total of 41,182 records after duplicate removal. After abstract and title screening, 415 papers were included for full-text review, leading to 15 studies being identified as suitable for inclusion [44-58]. In addition, 1 study was added after screening the reference list of the included studies [59]. The PRISMA flow diagram is shown in Figure 1.

XSL•FO
RenderX

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



Study Characteristics

The characteristics of the 16 included studies are shown in Multimedia Appendices 2 and 3. There were 10 RCTs [44-52,59], one of which was of a cross-over design [52], 1 quasi-experimental study [53], 2 pre-post studies [54,56], and 3 observational studies [55,57,58]. Seven studies delivered interventions within health care facilities [44,47,51,53-55,59],

XSL•FO
RenderX

5 in educational institutions [45,49,50,52,57], and 4 online [46,48,56,58].

Participant Characteristics

Two studies were performed in children [47,59], 1 in children and their parents [51], 5 in adolescents and young adults [45,46,49,50,57], and the remaining in the adult population 18 years or over [44,48,52-56,58].

Overall, 4158 participants were recruited across studies, of which 4097 were analyzed after taking into account participant dropouts. The number of recruits per study ranged from 32 [52] to 1160 participants [57]. The mean age of the participants was 27.1 years (excluding studies that presented age in median only [48] and those that did not present age data [58]). Female participants accounted for 66.7% (n=2734) of the participants across the studies. Detailed participant characteristics are presented in Multimedia Appendices 2 and 4.

Health Conditions

Ten (63%) out of the 16 studies assessed the effectiveness of an online video intervention on a mental health–related condition [44-47,52,54-56,58,59]. Among these, 1 study assessed the impact of videos on health care workers' mental health relating to the COVID-19 pandemic [44]. Four studies used social media videos to alleviate peri-procedural anxiety [47,54,55,59], including 3 that assessed the impact of watching selected videos during the preoperative period on preoperative anxiety [54,55,59]; 1 study examined the efficacy of videos in reducing anxiety relating to dental procedures [47].

Two studies used video interventions to improve participants' lifestyles; 2 studies examined the effects of YouTube video–delivered physical activity intervention on young adults' physical activity [49,50]; and 1 study investigated the impact of watching food videos on social media on food consumption, appetite, and BMI [57].

Other health conditions targeted by video interventions included pain relating to knee osteoarthritis [49], patients with coronary artery disease undergoing cardiac catheterization [54], and fear of topical steroid treatment [51].

Video Intervention

Delivery of the Video Intervention to Participants

Overview

All videos were hosted on publicly available social media platforms, meaning that any individual could access and view the video content online, unlimited to study participants. Twelve studies hosted videos on YouTube [44-47,49,50,52,54-56,58,59]. All except for 3 studies intentionally delivered the videos to a predetermined group of participants and evaluated their health impact on this group [44-54,56,59]. The following methods of delivery to study participants were described in the included studies.

Delivered Through Weblinks

Three studies provided participants with the weblink of the specified social media videos and instructed participants to watch the videos at specific time points [44,50,51]. In 1 study, patients were given a link to the social media website, and they could search the video using keywords such that they could watch it asynchronously at their convenience [53].

Delivered Directly to Participants

In 5 studies, participants were directly shown the video sequence in a controlled environment, including the following ways: shown on large screens in the preoperative waiting room [59], shown prior to dental procedures [47], shown through personal cell phones in a testing room [52], or shown to participants in a clear area of a quiet classroom at a school [45]. In 1 study, patients were shown the video through a variety of tools prior to their surgery, including digital monitor devices connected to a hospital smart bed system, or through any devices with internet access by entering the video URL or with QR code scanning [53].

Embedded Within Surveys

In 3 studies, video interventions were embedded within a randomized or single-arm pre-post online survey [46,48,56], of which 2 studies used the Qualtrics platform [48,56] and 1 used the Social Science Survey platform [46]. Participants were invited to complete the survey through audience-specific social media platforms [46,56], invitation flyers distributed around university departments [46], invitation shown in the YouTube description section of the intervention video [56], and through a consumer network for digital survey–based research [56].

Embedded Within YouTube Channels

In a 2-arm RCT investigating the impact of social media videos on young individuals' physical activity, participants were instructed to subscribe to the YouTube channel for their respective study groups [49]. They would watch 1 video per week that was uploaded onto each channel, and they were instructed not to consume other YouTube videos relating to physical activity during this time period [49].

Observational Studies

Three studies were observational rather than interventional. One prospective study compared the preoperative anxiety of individuals who have watched operation-related videos against those who have not watched such videos within 1 week prior to the operation [55]. The remaining 2 were cross-sectional studies. Tazeoğlu et al [57] investigated the association between self-reported watching of social media food videos and individuals' weight and BMI. Participants were asked to recall the frequency in which they had watched any food videos on social media platforms. Shin et al [58] uploaded a cross-sectional survey onto a specific YouTube channel that produces sleep-aid related content and invited the channel's audience to report how watching videos on the channel has impacted their quality of sleep.

Video Content

Four studies used videos to guide participants through an activity, including Yoga Nidra [44], workout routines [45,49,50], and autonomous sensory meridian response sound [56]. Seven studies placed greater focus on information provision in educating participants about procedures [47,53-55] or on specific health conditions including osteoarthritis [48], suicide prevention

XSL•FO
RenderX

in depression [46], and atopic dermatitis and its treatment [51]. The remaining studies offered participants greater flexibility in selecting the videos to watch. One study investigating preoperative anxiety selected videos from a list of age-appropriate video clips based on participants' individual preference [59]. In Oppenheimer et al [52], participants could watch any videos from a YouTube playlist of 8 nonevocative videos. In 2 studies, participants watched any videos on specified topics (videos on how to perform an impacted tooth extraction [55] and food-related videos [57]) or videos on a specific YouTube channel providing sleep aid [58].

Three studies mentioned the use of theories and/or frameworks in developing videos [48,49,51], including the following: constructivism, social-cognitive theory and information-motivation-behavioral skills model [48], social determination theory [49], storytelling, and behavior change technique taxonomy version 1 [51]. Notably, the videos in these 3 studies were evaluated with their audience prior to their formal delivery to participants. Videos were evaluated through consumer panels [48]; survey and focus groups [49]; and a panel of patients, family members, and health care professionals [51].

Video Format

Of the studies that reported video lengths, they ranged from 1 minute 25 seconds to 30 minutes [44,52]. Six out of the 16 studies used 1 video only [44,46,48,53,54,56], while the remaining incorporated several videos as part of the intervention [45,47,49-52,55,57-59]. Participants were either shown the video at a single time point [46-48,52,54,56,59] or at regular time intervals [44,45,49,50] or at the participant's time of convenience [51,53,55,58].

Risk of Bias

Among the RCTs, 1 study showed low risk of bias [49], while the remaining studies showed some concern [46,47,51,59] or high risk of bias [44,45,48,50,52]. Of the four nonrandomized controlled studies, all have shown moderate [53,54,58], serious [55,57], or critical [56] risk of bias. For the RCTs, the most reported bias was in the selection of the reported results, whereas for the non-RCTs, it was the bias in the measurement of outcomes. The summarized risk of bias assessment of the included studies is shown in Multimedia Appendix 5 for RCTs [43-51,58], and non-RCTs [52-55], and Multimedia Appendix 6 for the full assessment.

Assessing the Impact of Online Videos

The primary outcome of the systematic review was the impact of the health-related video intervention on quantitative health–related outcomes as specified by each paper.

Narrative Synthesis

Two RCTs [47,59] and 1 pre-post [54] study all demonstrated that online video interventions can significantly improve peri-procedural anxiety including perioperative anxiety and dental anxiety ($P \leq .001$). One prospective cohort study, however, reported that watching social media videos on tooth extraction 1 week prior to the procedure may be associated with a greater level of anxiety ($P < .05$) [55]. While 2 studies measured procedural-related anxiety using self-reported measures [54,55], 2 assessed outcomes by independent assessors [47,59], and 1 RCT measured 1 component of preprocedural anxiety using heart rate measured with a finger pulse oximeter [47].

In terms of other mental health–related outcomes, 1 study demonstrated that an online video depicting personal stories of how to cope with depression significantly alleviated suicidal ideation ($P = .04$) [46], and a video showing Yoga Nidra to health care workers during COVID-19 duty period improved insomnia ($P = .02$) [44]. One study that showed videos that triggered autonomous sensory meridian response in participants led to improved mood ($P = .002$) and lower levels of arousal ($P < .001$) [56]. No significant impact was demonstrated on participants' levels of depression [44] or stress [45]. One crossover RCT comparing watching a nonevocative YouTube playlist against browsing social media platforms (Facebook and Instagram) showed that both interventions reduced the level of stress, and there was no significant difference between the 2 interventions [52]. Most studies describing mental health–related outcomes were measured using self-completed questionnaires; 1 study measured the level of stress objectively using arm-band continuous heart rate monitoring and individual pre- and postintervention cortisol levels [52].

Two studies assessed the impact of videos on self-reported quality-of-life outcomes: an RCT of online videos that described atopic dermatitis and its treatment, and 1 quasi-experimental study of showing a video prior to cardiac catheterization. No significant impact was found in either study, although watching online videos prior to cardiac catheterization improved spiritual well-being ($P < .001$) [51,53].

One RCT with low risk of bias found that a video grounded in self-determination theory significantly improved young adults' level of physical activity compared with watching a general health video [49]. The study measured the level of physical activity and sleep quality of its participants using a wrist-worn ActiGraph Link GT9X accelerometer [49]. Another RCT examining the impact of a video intervention on executive function, measured using self-completed tasks by participants, did not demonstrate significance [45].

Two RCTs that incorporated appropriate theories and/or frameworks in developing videos showed a significant impact on improving patients' self-sufficiency in managing pain relating to knee osteoarthritis [48], and in reducing parents' fear of topical corticosteroid, respectively, although the latter study did not have a significant impact upon disease severity or the family's quality of life [51].

Of the two cross-sectional studies, 1 study demonstrated a positive correlation between those who regularly watch food videos and individuals' BMI, while the other reported effectiveness in YouTube-delivered mind-body interventions in improving sleep quality [57,58].

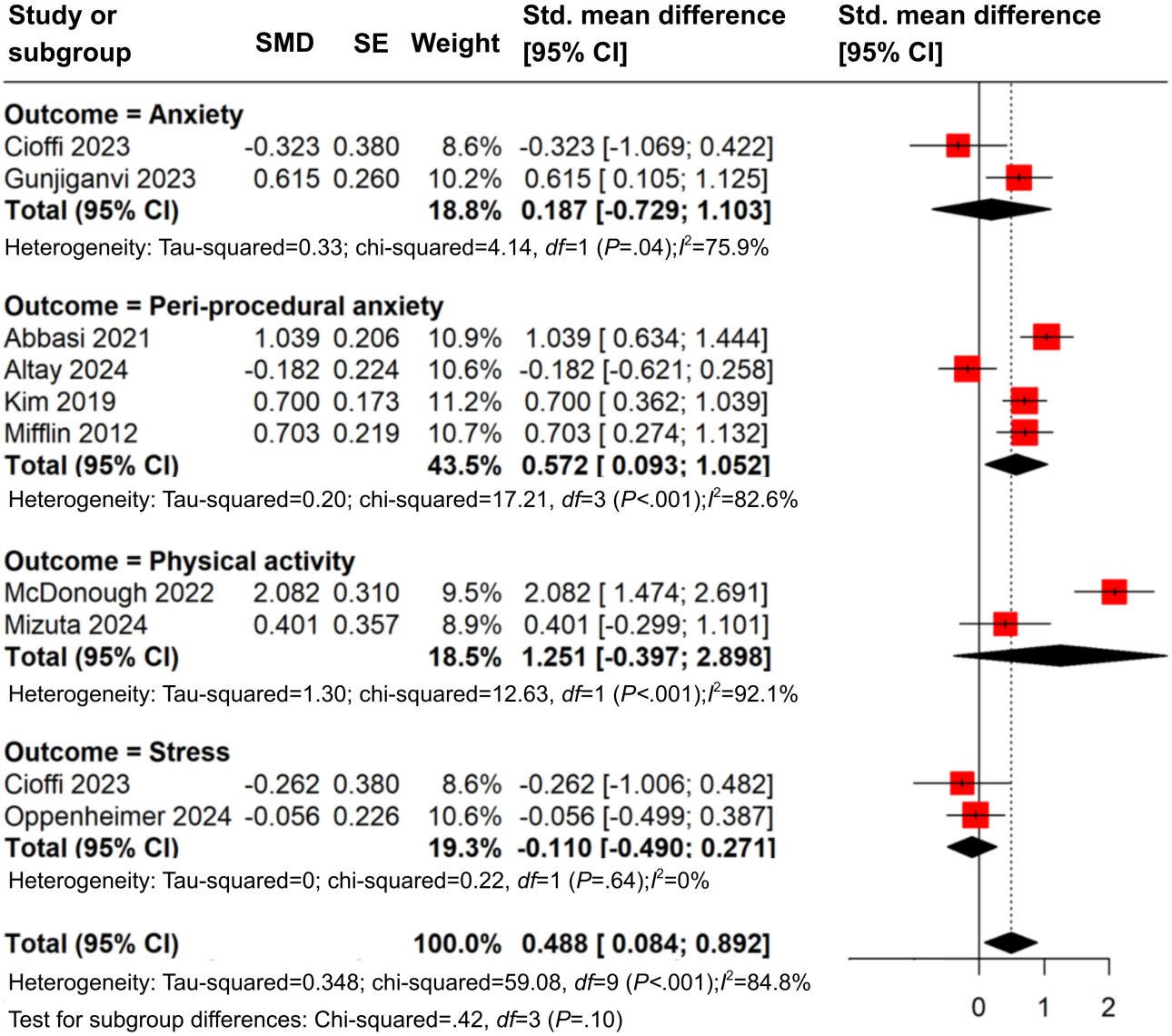
Nine studies used solely self-reported outcome measures [46,48,50,53-56,58]. Five studies incorporated predominantly objectively measured outcomes or observer-determined outcomes [47,49,52,57,59].

XSL•FO
RenderX

Quantitative Synthesis

The primary outcomes and measures of the included studies were categorized based on outcome types, and the numerical values, calculated standard mean difference, and 2-tailed *P* values in the context of the studies' risk of bias are shown in Multimedia Appendix 4. While the included studies displayed variations in terms of study design, participants, and the range of health outcomes, a subgroup meta-analysis was conducted to examine the impact of online video interventions on different health-related outcomes, including anxiety, peri-procedural anxiety, physical activity, and stress (Figure 2) [45,47,49,50,52,54,55,59]. Stratified by outcomes, for peri-procedural anxiety, a significant moderate effect was observed (SMD=0.57, 95% CI 0.09 - 1.05, I^2 82.6%), although studies were heterogeneous in terms of study design (RCTs and non-RCTs) and participant characteristics (both adults and children). For physical activity, a large yet nonsignificant effect was identified (SMD=1.25, 95% CI -0.40 - 2.90, I^2 92.1%), and the 2 included studies were both RCTs with participants of similar age. No significant effects were identified for anxiety and stress. Notably, all but 1 study included in sub-group meta-analysis displayed moderate-to-high risk of bias.

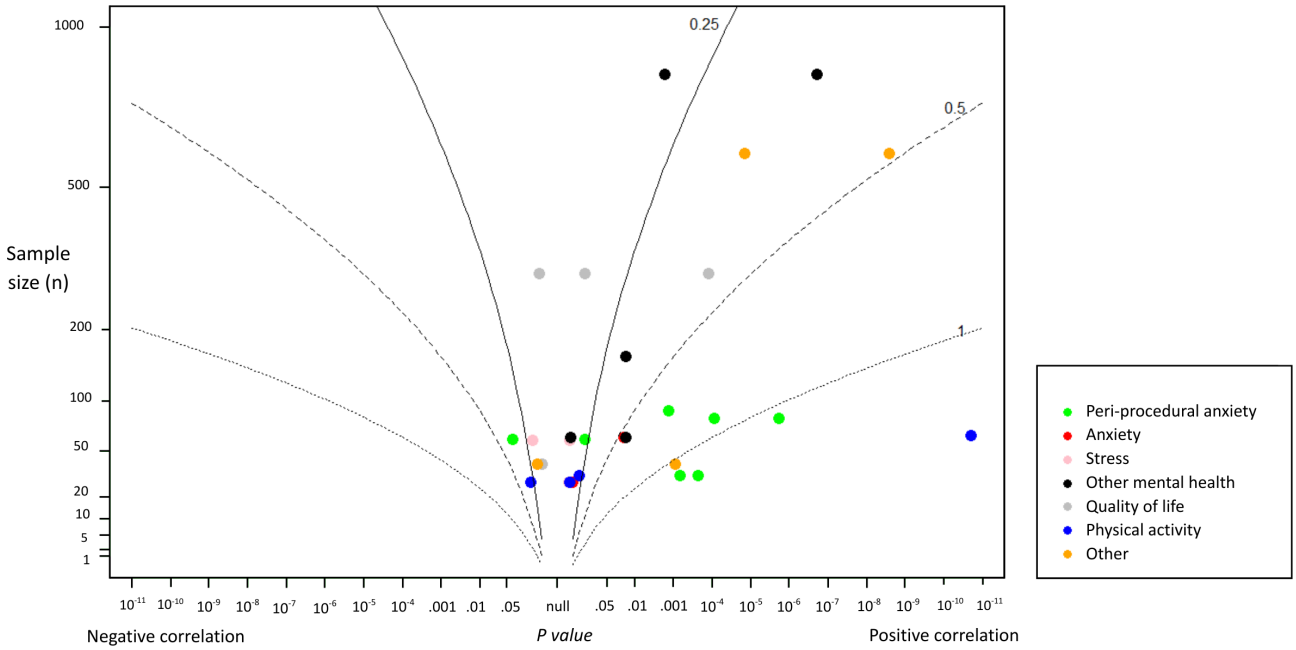
Figure 2. Forest plot of standard mean differences (SMDs) for intervention effects across outcomes. Pooled estimates were calculated using random-effects models. Diamonds represent the overall effect size for each outcome and the total pooled effect size. Horizontal lines indicated 95% CI. Study weights are shown as percentages. Heterogeneity was assessed using the I^2 statistic and Tau^2 [44,45,47,49,50,52,54,55,59].



An albatross plot (Figure 3) was used to visualize the primary outcomes across all interventional studies. Most studies were of sample sizes between 50 and 200. A concentration of points could be observed within the small-to-moderate effect size contours (SMD 0.25 - 0.50). Notably, peri-procedural anxiety appeared to be more significantly associated with watching social media videos across studies, while other outcomes appeared more dispersed. One RCT on the effect of a behavioral framework–informed video on physical activity displayed the greatest effect size among all studies [49].

XSL•FO
RenderX

Figure 3. Albatross plot of *P* values against study sample size. The plot was constructed to visually summarize the distribution of study results by mapping *P* values against sample sizes (**N**). The vertical axis represents sample size on a logarithmic scale (ranging from 1 to 1000), while the horizontal axis displays *P* values on a 2-sided scale, with negative correlations plotted to the left and positive correlations to the right. Contour lines indicate approximate effect sizes expressed as standardized mean differences (SMD=0.25, 0.5, 1.0). Each point corresponds to a particular outcome of an individual study, color-coded according to outcome domain (eg, peri-procedural anxiety, stress, quality of life, physical activity). Studies falling closer to the center line suggested weaker or null effects, whereas those positioned further toward the contour boundaries indicate stronger associations.



Discussion

Principal Results

To our knowledge, this is the first systematic review to comprehensively evaluate the impact of social media videos on quantitative health outcomes. The evidence indicates that social media videos can positively influence health outcomes, particularly by reducing peri-procedural anxiety. When interventions were developed with behavioral or psychological theories and were informed by audience input, they demonstrated stronger and more consistent improvements in study-specific health outcomes.

Most studies utilized self-reported outcome measures, and significant gaps persist in measuring the impact of online videos on physical health–related outcomes. Of the studies that measured physical health–related outcomes, including disease severity and 90-day readmission rate, no significant impact was found [50,51], with the exception of McDonough et al [49], which demonstrated that a theory-based video, tailored to audience preferences, significantly improved physical activity levels.

Notably, all but 1 study demonstrated moderate-to-high risk of bias [49].

Comparison With Prior Work

The most consistent finding across the included studies was the positive impact of online videos in reducing peri-procedural anxiety. This was aligned with a previous systematic review on video-based preoperative patient education—not specific to social media videos—in which the majority of the studies (six out of eight) reported that preoperative videos significantly lowered patients' anxiety scores [60]. The preprocedural period can often be psychologically burdensome for patients. Preparatory information shown visually through videos can potentially help to alleviate anxiety by clarifying what to expect during medical procedures. Interestingly, videos tailored to patients' preferences were as effective in reducing anxiety as those explaining the medical procedures themselves, suggesting mechanisms beyond patient education. Nevertheless, when examining the impact of videos on peri-procedural physical and quality-of-life outcomes, the findings of this systematic review were consistent with previous studies that no significant changes were demonstrated [53,54,60]. This may suggest that a video intervention provided preoperatively may not lead to enough change in individuals' knowledge, attitude, and behaviors, such that individuals' postoperative recovery can be impacted, which can lead to improved physical outcomes.

Several studies in our review described the use of online video as a medium to deliver specific interventions to improve individuals' health outcomes. For instance, Gunjiganvi et al described the use of Yoga Nidra, a form of meditative procedure described by its practitioners to induce a calm inner stillness [44,61]. Previous studies examining the effectiveness of in-person Yoga Nidra, including RCTs, have similarly demonstrated its effectiveness in alleviating anxiety, depression, and insomnia [62-64], as well as physical outcomes such as blood pressure and lipid profiles [65,66]. It appeared that the delivery of Yoga Nidra through the video format can result in similar improvement in the participants' mental health–related outcomes [44]. Similarly, when participants were instructed to follow a series of physical activity videos, described in McDonough et al [49], the level of physical activity of the participants was significantly increased. Nevertheless, it remains uncertain as to whether such improvement persisted postintervention.

XSL•FO
RenderX

Regarding video content, 3 studies developed videos incorporating theories or frameworks, and with input from their target audience [48,49,51]. All have demonstrated significant positive impact on their respective outcomes, including psychological-cognitive outcomes in people with knee pain, levels of physical activity, and fear of topical steroid treatment [48,49,51]. Such findings align with a previous systematic review assessing the effectiveness of video-based intervention in health promotion, which highlighted the importance of incorporating theoretical frameworks to guide message development in future video interventions [67].

Limitations

Our systematic review has several limitations. First, we restricted our search to databases containing peer-reviewed papers and excluded studies lacking full methodological details, such as meeting abstracts and proceedings, to ensure the quality and completeness of our review. However, this may have excluded relevant methodologies and outcomes discussed in the broader literature. Second, to ensure broad coverage of health-related conditions, we used wide-ranging search terms, resulting in a large number of papers for screening. Given resource constraints, this systematic review was limited by having 1 reviewer for the abstract and title screening, which could increase its vulnerability to selection bias. To reduce such bias, papers with unclear eligibility were conservatively included, and the full text screening phase involved 2 reviewers to ensure concordance. Third, we focused exclusively on long-form social media videos, as audience interaction with "Shorts" videos can vary substantially. "Shorts"—brief, algorithmically surfaced video clips designed for quick consumption—have grown in popularity in recent years, reshaping how audiences engage with content, but such format has been associated with the risk of online addiction [28,29]. The findings from this systematic review should thus be generalized to long-form social media videos.

Importantly, all except for 1 included study have displayed moderate-to-high risk of bias, and the most reported bias stemmed from the measurement and reporting of study outcomes [49]. Such findings reemphasized the importance of utilizing more objective methods in tracking health-related outcomes beyond self-reported tools. While the systematic review reported a significant positive impact of social media videos on peri-procedural anxiety, the 4 included studies all showed a moderate-to-high risk of bias and are heterogeneous in terms of study designs (both RCT and non-RCTs) and participant demographics (children and adults); therefore, future studies should validate such findings with better designed studies, more robust form of measuring the outcomes beyond self-reported tools, and in a specified population.

Future Work and Implications

Social media is characterized by its bidirectional, interactive nature, and previous research has shown the paramount importance of engaging community partners when designing public health messaging in order to build trust and ensure its effectiveness [25,68]. Nevertheless, all but 2 studies in this systematic review delivered the videos in a unidirectional manner to a predetermined group of participants in a controlled, experimental setting, without clear opportunities for reciprocal interactions. Two studies attempted to better simulate how individuals typically interact on social media platforms: 1 utilized features on the YouTube platform by creating 2 separate channels for participants in experimental and control groups, yet the audience was preselected with no clear monitoring of their online interactions [49]; 1 cross-sectional study directly posted the online questionnaire onto the YouTube channel containing sleep-aid video interventions, allowing respondents to better reflect how users typically behave on such platforms [58].

The benefit of interactivity on social media has been highlighted in previous studies among individuals suffering from mental illnesses [5,69]. Individuals find ease of connecting with each other anonymously for mutual support, especially when they may have few social contacts offline and may suffer from highly stigmatizing conditions [5,69]. On the other hand, the negative impact of social media cannot be undermined, including the potential for mass misinformation and disinformation, and its link with online addiction, associated with the platforms' algorithmic nature [70-72]. Future studies should focus on assessing the health impact of social media videos incorporating the full range of functionalities native to the platform, such as likes, shares, and comments on the YouTube platform, as well as the platform's algorithmic nature, by which users receive individualized content that reflects their viewing history [73]. While the included studies of this systematic review tended to show 1 or a predetermined series of videos, the users of such platforms may view a series of videos as determined by the algorithm. How such features can positively or negatively impact upon their users' health outcomes is important to be explored, especially in evaluating how social media can impact upon the traditional health care services, and whether, or indeed, how such interventions can be incorporated into the health care workflow.

This systematic review found that the assessment of the health impact of social media videos primarily focused on mental health-related or more subjectively reported outcomes, such as pain and fear. Previous systematic reviews of health-related video interventions have described their use for a broader range of purposes in health care, including health promotion such as disease prevention (eg, nutrition, vaccination), detection (eg, cancer screening and self-examination), and prevention of risky behaviors (eg, smoking, binge drinking) [67], as well as hospital-based education to change outcomes relating to heart disease, cancer, stroke, sleep apnea, and diabetes [74]. Several studies in this systematic review have attempted to measure objective physical health–related outcomes, such as the use of heart rate and physical activity trackers, and the measure of cortisol levels [49,52]. Nevertheless, given the wide-reaching potential of social media to a global audience, our findings highlighted the need to assess a wider range of health outcomes, notably those relating to physical health, utilizing methodological innovations such as the use of digital tracing, passive sensors, and laboratory investigations. Given the existing interactive features on social media platforms, means to directly measure health-related outcomes or their surrogate measures

XSL•FO
RenderX

on the platform should be explored, such as the embedding of validated survey tools and digital tracking of user behaviors.

The variable effectiveness of video interventions identified in this systematic review can be influenced by several factors, including the duration and frequency of exposure to intervention, the content and quality of the videos, participant compliance, and whether behavioral theories or frameworks were applied in developing the intervention. As a visual-audio tool, it is necessary to understand the mechanisms that lead to a change in individuals' health-related outcomes, including the effect of the videos on individuals' health-related behaviors and their determinants such as knowledge and attitudes. Mental health–related outcomes may be more easily affected by a video, while a change in physical health–related outcomes may require more consistent and interactive interventions. Nevertheless, in 1 RCT of low risk of bias, where participants can directly follow the physical activity videos that were grounded in a behavioral theory, which were consistently delivered to participants on a weekly basis, there was a significant increase in the participants' physical activity levels [49]. Health care providers and policymakers should thus design future studies to examine how best to deliver social media videos as behavioral interventions, with in-depth behavioral analysis of the individuals affected by the health conditions, so that more targeted video interventions that are grounded in behavioral theories or frameworks can be developed.

Conclusions

In conclusion, this systematic review has demonstrated a potential positive impact of online social media videos in improving peri-procedural anxiety and the merit of incorporating behavioral theories or frameworks in developing the interventions. Nevertheless, studies have demonstrated a moderate-to-high risk of bias and high heterogeneity in terms of study design, participant demographics, and the range of health conditions. Future studies should focus on the measurement of more objective physical outcomes and the evaluation of video interventions in the context of the interactive and algorithmic features of social media platforms. Health care providers and policymakers should endeavor to incorporate in-depth behavioral analysis of the target populations so as to develop interventions grounded in behavioral theories or frameworks.

Data Availability

All data analyzed in this study are present in Multimedia Appendices 1-6.

Authors' Contributions

Conceptualization: AA (equal), FC (lead), KG (equal),
Data curation: CN (supporting), FC (lead), HT (equal), TT (equal)
Formal analysis: FC (lead), SD (supporting)
Methodology: AA (supporting), FC (lead), KG (supporting)
Supervision: AD (supporting), KG (lead)
Visualization: FC (lead), SD (supporting)
Writing – original draft: AA (supporting), FC (lead), KG (supporting), SD (supporting)
Writing – review & Editing: AA (supporting), AD (supporting), CN (supporting), HT (supporting), KG (lead), SD (supporting), TT (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 21 KB - infodemiology_v6i1e77578_app1.docx]

Multimedia Appendix 2
Summary table of the included studies and study core characteristics.
[XLSX File, 40 KB - infodemiology_v6i1e77578_app2.xlsx]

Multimedia Appendix 3
Detailed extraction information of all included studies.
[XLSX File, 78 KB - infodemiology_v6i1e77578_app3.xlsx]

Multimedia Appendix 4
Study numerical outcomes by subgroups in the context of risk of bias.
[XLSX File, 22 KB - infodemiology_v6i1e77578_app4.xlsx]

XSL•FO
RenderX

Multimedia Appendix 5
Risk of bias of the included randomized controlled trials and nonrandomized controlled trials (except for cross-sectional studies).
[DOCX File, 1286 KB - infodemiology_v6i1e77578_app5.docx]

Multimedia Appendix 6
Risk of bias assessment by items for all included studies.
[XLSX File, 58 KB - infodemiology_v6i1e77578_app6.xlsx]

Checklist 1
PRISMA checklist.
[DOCX File, 298 KB - infodemiology_v6i1e77578_app7.docx]

References

1. Social media. Cambridge Dictionary. URL: https://dictionary.cambridge.org/dictionary/english/social-media [accessed 2025-03-05]
2. Grajales III FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014;16(2):e13. [doi: 10.2196/jmir.2912]
3. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194. [doi: 10.1016/S2589-7500(20)30315-0]
4. Braun LA, Zomorodbakhsch B, Keinki C, Huebner J. Information needs, communication and usage of social media by cancer patients and their relatives. J Cancer Res Clin Oncol 2019 Jul;145(7):1865-1875. [doi: 10.1007/s00432-019-02929-9] [Medline: 31123824]
5. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social media and mental health: benefits, risks, and opportunities for research and practice. J Technol Behav Sci 2020 Sep;5(3):245-257. [doi: 10.1007/s41347-020-00134-x] [Medline: 33415185]
6. How YouTube search works. YouTube. URL: https://www.youtube.com/intl/ALL_uk/howyoutubeworks/product-features/search/ [accessed 2025-01-06]
7. Mayer RE. Applying the science of learning: evidence-based principles for the design of multimedia instruction. Am Psychol 2008 Nov;63(8):760-769. [doi: 10.1037/0003-066X.63.8.760] [Medline: 19014238]
8. Rolfe VE, Gray D. Are multimedia resources effective in life science education? A meta-analysis. Biosci Educ 2011 Dec;18(1):1-14. [doi: 10.3108/beej.18.5]
9. Kesten JM, Bhattacharya A, Ashiru-Oredope D, Gobin M, Audrey S. The Antibiotic Guardian campaign: a qualitative evaluation of an online pledge-based system focused on making better use of antibiotics. BMC Public Health 2017 Jul 11;18(1):5. [doi: 10.1186/s12889-017-4552-9] [Medline: 28693462]
10. Noetel M, Griffith S, Delaney O, et al. Video improves learning in higher education: a systematic review. Rev Educ Res 2021 Apr;91(2):204-236. [doi: 10.3102/0034654321990713]
11. Schneider S, Nebel S, Beege M, Rey GD. The autonomy-enhancing effects of choice on cognitive load, motivation and learning with digital media. Learn Instr 2018 Dec;58:161-172. [doi: 10.1016/j.learninstruc.2018.06.006]
12. Kopf S, Wilk S, Effelsberg W. Bringing videos to social media. Presented at: 2012 IEEE International Conference on Multimedia and Expo; Jul 9-13, 2012. [doi: 10.1109/ICME.2012.86]
13. Al-Quran MWM. Traditional media versus social media: challenges and opportunities. Technium 2022;4(1):145-160. [doi: 10.47577/technium.v4i10.8012]
14. 4 stories of living with multiple sclerosis - #mymsmoment. National MS Society YouTube page. 2022 Mar 14. URL: https://www.youtube.com/watch?v=P5wFE3ZQIfw&ab_channel=NationalMSSociety [accessed 2025-03-05]
15. How can the world solve the climate and biodiversity crises? Grantham Institute, Imperial College London YouTube page. 2022 Nov 2. URL: https://www.youtube.com/watch?v=CNl6fB7wBSA&ab_channel=GranthamImperial [accessed 2025-03-05]
16. Antibiotic resistance advert—keep antibiotics working and take your doctor's advice. Chorley and South Ribble and Greater Preston CCG YouTube page. 2017. URL: https://www.youtube.com/watch?v=zTbLai2GaQM&ab_channel=ChorleyandSouthRibbleandGreaterPrestonCCG [accessed 2025-03-05]
17. Bhattacharya A, Hopkins S, Sallis A, Budd EL, Ashiru-Oredope D. A process evaluation of the UK-wide Antibiotic Guardian campaign: developing engagement on antimicrobial resistance. J Public Health (Oxf) 2017 Jun 1;39(2):e40-e47. [doi: 10.1093/pubmed/fdw059] [Medline: 27451416]
18. Health outcomes. Taylor and Francis. 2025. URL: https://taylorandfrancis.com/knowledge/Medicine_and_healthcare/Epidemiology/Health_outcomes [accessed 2025-03-05]
19. Oleske DM, Islam SS. Role of epidemiology in the biopharmaceutical industry. In: Doan T, Renz C, Bhattacharya M, Lievano F, Scarazzini L, editors. Pharmacovigilance: A Practical Approach: Elsevier; 2019:69-87. [doi: 10.1016/B978-0-323-58116-5.00005-5]

XSL•FO
RenderX

20. Pantaleon L. Why measuring outcomes is important in health care. J Vet Intern Med 2019 Mar;33(2):356-362. [doi: 10.1111/jvim.15458] [Medline: 30784125]
21. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
22. Tuong W, Larsen ER, Armstrong AW. Videos to influence: a systematic review of effectiveness of video-based education in modifying health behaviors. J Behav Med 2014 Apr;37(2):218-233. [doi: 10.1007/s10865-012-9480-7] [Medline: 23188480]
23. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917. [doi: 10.2196/17917] [Medline: 33978589]
24. Claflin SB, Klekociuk S, Fair H, et al. Assessing the impact of online health education interventions from 2010-2020: a systematic review of the evidence. Am J Health Promot 2022 Jan;36(1):201-224. [doi: 10.1177/08901171211039308] [Medline: 34382448]
25. Ghahramani A, de Courten M, Prokofieva M. The potential of social media in health promotion beyond creating awareness: an integrative review. BMC Public Health 2022 Dec 21;22(1):2402. [doi: 10.1186/s12889-022-14885-0] [Medline: 36544121]
26. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Oladipo T. Ask Buffer: what is short-form video, and how can you use it? Buffer. URL: https://buffer.com/resources/short-form-video/ [accessed 2025-03-05]
28. Qin Y, Omar B, Musetti A. The addiction behavior of short-form video app TikTok: the information quality and system quality perspective. Front Psychol 2022;13:932805. [doi: 10.3389/fpsyg.2022.932805] [Medline: 36148123]
29. Chen Y, Li M, Guo F, Wang X. The effect of short-form video addiction on users' attention. Behav Inf Technol 2023 Dec 10;42(16):2893-2910. [doi: 10.1080/0144929X.2022.2151512]
30. Arevalo-Rodriguez I, Smailagic N, Roqué I Figuls M, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). Cochrane Database Syst Rev 2015 Mar 5;2015(3):CD010783. [doi: 10.1002/14651858.CD010783.pub2] [Medline: 25740785]
31. Pequeno NPF, Cabral NLDA, Marchioni DM, Lima SCVC, Lyra CDO. Quality of life assessment instruments for adults: a systematic review of population-based studies. Health Qual Life Outcomes 2020 Jun 30;18(1):208. [doi: 10.1186/s12955-020-01347-7] [Medline: 32605649]
32. Google Translate. URL: https://translate.google.co.uk [accessed 2025-03-05]
33. Bramer WM. Variation in number of hits for complex searches in Google Scholar. J Med Libr Assoc 2016 Apr;104(2):143-145. [doi: 10.3163/1536-5050.104.2.009] [Medline: 27076802]
34. Covidence. URL: https://www.covidence.org/ [accessed 2026-02-05]
35. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
36. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919. [doi: 10.1136/bmj.i4919] [Medline: 27733354]
37. Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res 2020 Feb 29;7(1):7. [doi: 10.1186/s40779-020-00238-8] [Medline: 32111253]
38. Higgins JPT, Deeks JJ, Altman DG. Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons Ltd; 2019:241-284. [doi: 10.1002/9781119536604]
39. Higgins JPT, Li T, Deeks JJ. Choosing effect measures and computing estimates of effect. In: Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons, Ltd; 2019:143-176. [doi: 10.1002/9781119536604]
40. Meta: general package for meta-analysis. The Comprehensive R Archive Network. 2025. URL: https://CRAN.R-project.org/package=meta [accessed 2025-09-15]
41. Harrison S, Jones HE, Martin RM, Lewis SJ, Higgins JPT. The albatross plot: a novel graphical tool for presenting results of diversely reported studies in a systematic review. Res Synth Methods 2017 Sep;8(3):281-289. [doi: 10.1002/jrsm.1239] [Medline: 28453179]
42. Mansournia MA, Nazemipour M, Etminan M. P-value, compatibility, and S-value. Global Epidemiology 2022 Dec;4:100085. [doi: 10.1016/j.gloepi.2022.100085]
43. Meta-analysis of significance values. The Comprehensive R Archive Network. 2017. URL: https://CRAN.R-project.org/package=metap [accessed 2025-09-15]
44. Gunjiganvi M, Rai S, Awale R, Mishra P, Gupta D, Gurjar M. Efficacy of yoga nidra on depression, anxiety, and insomnia in frontline COVID-19 healthcare workers: a pilot randomized controlled trial. Int J Yoga Therap 2023 Jan 1;33(2023):Article 3. [doi: 10.17761/2023-D-22-00011] [Medline: 37327384]
45. Cioffi R, Lubetzky AV. BOXVR versus guided YouTube boxing for stress, anxiety, and cognitive performance in adolescents: a pilot randomized controlled trial. Games Health J 2023 Jun;12(3):259-268. [doi: 10.1089/g4h.2022.0202] [Medline: 36745402]

XSL•FO
RenderX

46. Niederkrotenthaler T, Till B. Effects of awareness material featuring individuals with experience of depression and suicidal thoughts on an audience with depressive symptoms: randomized controlled trial. J Behav Ther Exp Psychiatry 2020 Mar;66:101515. [doi: 10.1016/j.jbtep.2019.101515] [Medline: 31610437]
47. Abbasi H, Saqib M, Jouhar R, et al. The efficacy of little lovely dentist, dental song, and tell-show-do techniques in alleviating dental anxiety in paediatric patients: a clinical trial. Biomed Res Int 2021;2021:1119710. [doi: 10.1155/2021/1119710] [Medline: 34124238]
48. Egerton T, Bennell KL, McManus F, Lamb KE, Hinman RS. Comparative effect of two educational videos on self-efficacy and kinesiophobia in people with knee osteoarthritis: an online randomised controlled trial. Osteoarthr Cartil 2022 Oct;30(10):1398-1410. [doi: 10.1016/j.joca.2022.05.010]
49. McDonough DJ, Helgeson MA, Liu W, Gao Z. Effects of a remote, YouTube-delivered exercise intervention on young adults' physical activity, sedentary behavior, and sleep during the COVID-19 pandemic: randomized controlled trial. J Sport Health Sci 2022 Mar;11(2):145-156. [doi: 10.1016/j.jshs.2021.07.009] [Medline: 34314877]
50. Mizuta R, Maeda N, Tashiro T, et al. Effectiveness of metaverse space-based exercise video distribution in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2024 Jan 16;12:e46397. [doi: 10.2196/46397] [Medline: 38227355]
51. Brunner C, Schlüer AB, Znoj H, et al. Video-based education with storytelling reduces parents' fear of topical corticosteroid use in children with atopic dermatitis: a randomized controlled trial (the EduDerm study part II). Adv Skin Wound Care 2023 Aug 1;36(8):414-419. [doi: 10.1097/ASW.0000000000000002] [Medline: 37471446]
52. Oppenheimer S, Bond L, Smith C. Social media does not elicit a physiological stress response as measured by heart rate and salivary cortisol over 20-minute sessions of cell phone use. PLoS ONE 2024;19(4):e0298553. [doi: 10.1371/journal.pone.0298553] [Medline: 38568926]
53. Peng JR, Su HC, Lin CP, et al. Role of an e-health intervention in holistic healthcare: a quasiexperiment in patients undergoing cardiac catheterization in Taiwan. J Healthc Eng 2021;2021:6692952. [doi: 10.1155/2021/6692952] [Medline: 33815732]
54. Kim MJ, Oh HK, Lee KC, et al. Effects of an internet-based informational video on preoperative anxiety in patients with colorectal cancer. Ann Surg Treat Res 2019 Jun;96(6):290-295. [doi: 10.4174/astr.2019.96.6.290] [Medline: 31183333]
55. Altay B, Kale Ş, Basiry MN, Çoban E. The association of social media videos and patients' preoperative anxiety. J Oral Maxillofac Surg 2024 Feb;82(2):228-234. [doi: 10.1016/j.joms.2023.10.006] [Medline: 37925167]
56. Smejka T, Wiggs L. The effects of autonomous sensory meridian response (ASMR) videos on arousal and mood in adults with and without depression and insomnia. J Affect Disord 2022 Mar 15;301:60-67. [doi: 10.1016/j.jad.2021.12.015] [Medline: 34915083]
57. Tazeoğlu A, Kuyulu Bozdogan FB. The effect of watching food videos on social media on increased appetite and food consumption. Nutr Clín Diet Hosp 2022;42(2):73-79. [doi: 10.12873/422tazeoglu]
58. Shin JW, Yoo J, Kim H. User experiences and effects of expert-led YouTube mind-body interventions on insomniacs: a survey study. J Korean Med Sci 2025 Jun 2;40(21):e91. [doi: 10.3346/jkms.2025.40.e91] [Medline: 40461139]
59. Mifflin KA, Hackmann T, Chorney JM. Streamed video clips to reduce anxiety in children during inhaled induction of anesthesia. Anesth Analg 2012 Nov;115(5):1162-1167. [doi: 10.1213/ANE.0b013e31824d5224] [Medline: 23051880]
60. Tom K, Phang PT. Effectiveness of the video medium to supplement preoperative patient education: a systematic review of the literature. Patient Educ Couns 2022 Jul;105(7):1878-1887. [doi: 10.1016/j.pec.2022.01.013] [Medline: 35101307]
61. Pandi-Perumal SR, Spence DW, Srivastava N, et al. The origin and clinical relevance of yoga nidra. Sleep Vigil 2022;6(1):61-84. [doi: 10.1007/s41782-022-00202-7] [Medline: 35496325]
62. Moszeik EN, von Oertzen T, Renner KH. Effectiveness of a short yoga nidra meditation on stress, sleep, and well-being in a large and diverse sample. Curr Psychol 2022 Aug;41(8):5272-5286. [doi: 10.1007/s12144-020-01042-2]
63. Musto S, Hazard Vallerand A. Exploring the uses of yoga nidra: an integrative review. J Nurs Scholarsh 2023 Nov;55(6):1164-1178. [doi: 10.1111/jnu.12927] [Medline: 37489597]
64. Datta K, Tripathi M, Verma M, Masiwal D, Mallick HN. Yoga nidra practice shows improvement in sleep in patients with chronic insomnia: a randomized controlled trial. Natl Med J India 2021;34(3):143-150. [doi: 10.25259/NMJI_63_19] [Medline: 34825538]
65. Ahuja N, Bhardwaj P, Pathania M, et al. Yoga Nidra for hypertension: a systematic review and meta-analysis. J Ayurveda Integr Med 2024;15(2):100882. [doi: 10.1016/j.jaim.2023.100882] [Medline: 38484438]
66. Anjana K, Archana R, Mukkadan JK. Effect of om chanting and yoga nidra on blood pressure and lipid profile in hypertension—a randomized controlled trial. J Ayurveda Integr Med 2022;13(4):100657. [doi: 10.1016/j.jaim.2022.100657] [Medline: 36375220]
67. Xiao X, Wong RM, Yang W. Effectiveness of video-based health promotion: a systematic review and meta-analysis. Patient Educ Couns 2024 Feb;119:108095. [doi: 10.1016/j.pec.2023.108095] [Medline: 38070298]
68. de Vere Hunt I, Linos E. Social media for public health: framework for social media-based public health campaigns. J Med Internet Res 2022 Dec 14;24(12):e42179. [doi: 10.2196/42179] [Medline: 36515995]
69. Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube. PLoS ONE 2014;9(10):e110171. [doi: 10.1371/journal.pone.0110171] [Medline: 25333470]

XSL•FO
RenderX

70. Richards D, Caldwell PHY, Go H. Impact of social media on the health of children and young people. J Paediatr Child Health 2015 Dec;51(12):1152-1157. [doi: 10.1111/jpc.13023] [Medline: 26607861]
71. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85. [doi: 10.2196/jmir.1933] [Medline: 23615206]
72. Pellegrino A, Stasi A, Bhatiasevi V. Research trends in social media addiction and problematic social media use: a bibliometric analysis. Front Psychiatry 2022;13:1017506. [doi: 10.3389/fpsyt.2022.1017506] [Medline: 36458122]
73. Zollo F, Baronchelli A, Betsch C, Delmastro M, Quattrociocchi W. Understanding the complex links between social media and health behaviour. BMJ 2024 Jun 20;385:e075645. [doi: 10.1136/bmj-2023-075645] [Medline: 38901860]
74. Dahodwala M, Geransar R, Babion J, de Grood J, Sargious P. The impact of the use of video-based educational interventions on patient outcomes in hospital settings: a scoping review. Patient Educ Couns 2018 Dec;101(12):2116-2124. [doi: 10.1016/j.pec.2018.06.018] [Medline: 30087021]

Abbreviations

PICO: population, intervention, comparison, outcome
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standard mean difference

Edited by T Mackey; submitted 16.May.2025; peer-reviewed by A Ijeoma, A Ghahramani, C Obianyo, M Gasmi, ME Heidari, T Akinsanya-Ibrahim; revised version received 22.Nov.2025; accepted 26.Jan.2026; published 19.Feb.2026.

Please cite as:
Chen F, Tay T, Thould H, Nangsue C, Dryden S, Acharya A, Darzi A, Grailey K
The Impact of Social Media Videos on Quantitative Health Outcomes: Systematic Review
JMIR Infodemiology 2026;6:e77578
URL: https://infodemiology.jmir.org/2026/1/e77578
doi:10.2196/77578

© Fangyue Chen, Tricia Tay, Hannah Thould, Chatchaya Nangsue, Simon Dryden, Amish Acharya, Ara Darzi, Kate Grailey. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 19.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identifying Public Response Topics to the Centers for Disease Control and Prevention's COVID-19 Communications on Social Media: Infoveillance Study Using Large Language Model–Based Rephrasing

Wangjiaxuan Xin[1], MSc; Shuhua Yin[2], PhD; Albert Park[3], PhD; Yaorong Ge[4], PhD; Shi Chen[2,4], PhD

[1]College of Computing and Informatics, The University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC, United States

[2]Department of Public Health and Health Administration, The University of North Carolina at Charlotte, Charlotte, NC, United States

[3]School of Health Information Science, University of Victoria, Victoria, BC, Canada

[4]School of Data Science, The University of North Carolina at Charlotte, Charlotte, NC, United States

Corresponding Author:
Wangjiaxuan Xin, MSc
College of Computing and Informatics, The University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC, United States

Abstract

Background: Public health agencies increasingly use social media platforms such as Twitter (subsequently rebranded X, X Corp) to monitor public responses to official health communications and support timely decision-making during health crises. Public replies to communications from official agencies provide valuable insights into population-level discourse and engagement with health policies. However, analyzing large-scale short-text responses remains challenging because social media replies are often brief, informal, ambiguous, and linguistically fragmented. Topic modeling offers a scalable approach for identifying thematic patterns in such data, but its performance is often limited when applied to social media short texts. Recent advances in large language models (LLMs) provide new opportunities to improve text normalization before topic modeling, but their value and limitations for public health infoveillance remain underexplored.

Objective: This study develops and evaluates TM-Rephrase, a model-agnostic LLM-based rephrasing framework designed to improve the performance of topic models for short public health social media texts while ensuring rephrasing preserves the original semantic meaning, stance or intent, and tone.

Methods: We analyzed 25,027 public replies to official Centers for Disease Control and Prevention tweets on Twitter collected between May 2020 and November 2022. TM-Rephrase transformed informal and noisy short texts into more standardized expressions using 2 prompt-guided schemes: general rephrasing and colloquial-to-formal rephrasing. Original and rephrased texts were analyzed using multiple topic models, with rephrasing generated by Gemini 2.5 Flash, GPT-4o mini, and Mistral-7B-Instruct. Topic quality was evaluated using coherence, uniqueness, redundancy, and diversity. We also conducted an expert semantic-fidelity validation study in which 4 expert raters evaluated whether rephrased texts preserved the original meaning, stance or intent, and tone.

Results: TM-Rephrase generally improved topic quality across models, rephrasing schemes, and LLM backbones. For latent Dirichlet allocation, Cv coherence increased from 0.3094 without rephrasing to 0.5004 with colloquial-to-formal rephrasing. For BERTopic (BERT-based topic modeling), Cv coherence improved from 0.4078 to 0.4734. Diversity-related metrics also improved in most rephrasing settings. Additional checks using CNPMI, CUCI, and UMASS generally supported the main improvement pattern. Sensitivity analysis across different topic numbers further suggested that the coherence gains were not limited to the main setting. Expert validation showed favorable semantic preservation overall, with a mean rating of 3.90 (SD 1.01) and median of 4.00 (IQR 3.00-5.00) of 5.00, and acceptable interrater reliability, intraclass correlation coefficient =0.736. General rephrasing received stronger semantic-fidelity ratings than colloquial-to-formal rephrasing, suggesting that stronger formalization may improve topic readability while introducing greater risk of changes to the original meaning, stance or intent, and tone.

Conclusions: TM-Rephrase can enhance the coherence, distinctiveness, and interpretability of topic modeling outputs for short, noisy public health social media data. However, it is a text-normalization preprocessing strategy for topic enhancement rather than a fully neutral replacement for original public discourse.

(JMIR Infodemiology 2026;6:e98319) doi:10.2196/98319

XSL•FO
RenderX

KEYWORDS

text mining; topic modeling; public health infoveillance; COVID-19; social media analytics; large language models

Introduction

Background

The wide adoption of social media has fundamentally transformed public discourse, particularly during public health crises (eg, the COVID-19 pandemic) [1]. Twitter, characterized by rapid information dissemination and a concise conversational format, has emerged as a critical yet methodologically challenging data source for understanding and analyzing public discussion topics and the dynamics of information exchange during the COVID-19 pandemic [2]. In addition to enabling real-time communication, these platforms support large-scale public engagement, where individuals actively express opinions, share experiences, and respond to official health communications. For public health agencies such as the US Centers for Disease Control and Prevention (CDC), analyzing short-text public replies that express opinions and concerns provides valuable insights for informing more effective communication strategies to the public [3]. Such innovative surveillance and analyses can help capture evolving public concerns, assess responses to current health policies, and improve the alignment between institutional messaging and public needs. However, from a public health perspective, effectively leveraging these data remains challenging because of the unstructured nature of short, user-generated social media texts, which complicates the systematic monitoring and timely interpretation of population-level responses. In particular, the inherent brevity, informality, and ambiguity of individual tweets, along with the 280-character limit [4], pose significant challenges for traditional social media data mining and analytics. These characteristics are particularly problematic for topic modeling, which aims to identify latent thematic structures in textual corpora, because limited textual context constrains reliable semantic interpretation.

The contextual sparsity of social media short texts often leads to topics of keywords that are less coherent and less diverse (ie, more overlapping topics with redundant or unrelated keywords), thereby reducing interpretability and limiting their utility for understanding various public opinions during the pandemic. This limitation is further compounded by the heterogeneous nature of online public discourse, where variations in writing and expression styles introduce additional complexity into textual data. This problem becomes particularly pronounced when attempting to capture emerging and latent thematic structures in communications during public health emergencies, where semantic ambiguity, brevity, and informal linguistic styles (eg, abbreviations, slang, layman expressions, and misspellings) are prevalent [5]. Moreover, even when topic models identify clusters of topic keywords, they often lack meaningful interpretability without additional contextualization or external knowledge from experts, making it difficult to translate model outputs into coherent narratives for public health decision support. Consequently, these challenges limit the effectiveness of topic modeling in supporting systematic analysis of public health discourse. These issues suggest the need to standardize and formalize short social media texts into more contextually formal content before the application of topic modeling in health communications, thereby improving semantic clarity and enhancing the interpretability of downstream analytical results.

Topic modeling, as a widely used method for discovering latent thematic structures in textual corpora, is exemplified by the classic probabilistic model, latent Dirichlet allocation (LDA), proposed by Blei et al [6] in 2003. However, despite its broad adoption, applying LDA or its variants to short texts from social media platforms poses substantial challenges. The brevity, informality, and noise of tweets weaken statistical signals and exacerbate the context sparsity, leading to incoherent, redundant, or less interpretable topics generated from these probabilistic models [7].

To address these challenges, some enrichment strategies have been proposed. One line of research augments short texts with external knowledge or auxiliary contextual signals to enhance statistical robustness [8]. Another direction leverages distributed representations of words and documents, incorporating embeddings or neural architectures to map texts into richer semantic representations. Prominent examples include BERTopic (BERT-based topic modeling) [9] and Top2Vec [10], which use more recent transformer-based embeddings and clustering methods to generate more semantically meaningful topics. Recent work has explored more advanced neural network architectures, such as FASTopic (fast, adaptive, stable, and transferable topic model) [11] and topic-semantic contrastive learning [12], which demonstrate improved handling of context sparsity and have the potential to be applied in health communications.

However, all these approaches focus on algorithmic model redesign with parameterized inference strategies for short texts, which may adapt poorly to real-life large-scale data (eg, public health-related discourse on social media). In addition, such topic modeling algorithms typically output sets of top-ranked keywords that can be difficult for analysts or readers without additional domain knowledge to interpret.

Regarding the public discourse during COVID-19, the pandemic fundamentally transformed social media into the primary platform for real-time public discourse, where Twitter has been extensively used to study public attitudes and evolving concerns toward the pandemic and various health policies [13,14]. Scholars have analyzed public reactions to official communications from health agencies such as the CDC, focusing on opinion and thematic feedback to health policies [3,15]. Additional research has explored geospatial and demographic variations in public concerns, demonstrating the heterogeneous impact of the pandemic on different communities [16]. Despite the active research and the value of social media for effective public health infoveillance during emergencies, it often overlooks the interpretability issues arising from short and noisy texts in online discourse, especially on social media platforms.

XSL•FO
RenderX

Motivated by Maini et al [17], we leverage large language models (LLMs) as a text preprocessing approach for rephrasing short texts for topic modeling, aiming to standardize and formalize public replies to CDC tweets on Twitter. This framework, termed TM-Rephrase, builds on recent advances in LLMs, which have demonstrated powerful performance across diverse text-processing tasks, including text summarization [18], sentiment analysis [19], and information retrieval [20].

TM-Rephrase differs from conventional text preprocessing approaches. Standard text normalization and spelling correction primarily address surface-level noise, such as misspellings, inconsistent casing, punctuation, or informal abbreviations. Data augmentation usually creates additional training examples to improve model learning. In contrast, TM-Rephrase uses LLM-based, prompt-constrained rephrasing to transform each short social media reply into a more standardized and contextually explicit expression for topic modeling. Its distinguishing characteristic lies in treating rephrasing as a model-agnostic, data-centric preprocessing strategy for improving topic-model outputs rather than as a replacement for topic modeling algorithms or as generic text rewriting.

In this study, we examine the impact of TM-Rephrase on the quality of topics derived from 4 representative topic models, using a dataset of public replies to CDC Twitter tweets during the COVID-19 pandemic. We hypothesize that rephrasing original short-text public replies on social media through TM-Rephrase enhances topic quality in terms of both intratopic semantic coherence and inter–topic metrics of topic model outputs, which are represented as sets of keywords. Specifically, we investigate two rephrasing schemes: (1) general rephrasing and (2) colloquial-to-formal rephrasing, and evaluate their effectiveness across multiple topic modeling algorithms, including probabilistic algorithms (eg, LDA [6]) and more advanced neural network-based topic models (eg, BERTopic [9] and FASTopic [11]).

Research Questions and Objectives

The research questions (RQs) of this study are as follows: (RQ1) whether and how effectively TM-Rephrase improves the quality and interpretability of topic model outputs for short texts in health communications, (RQ2) how different rephrasing schemes influence topic model performance, and (RQ3) how to interpret public discussions (ie, replies to CDC during the COVID-19 pandemic) based on the topics and inform public health practice.

This study aims to improve the analysis of public responses to official public health communications on social media to support more effective public health surveillance. To achieve this, we develop and evaluate TM-Rephrase, an LLM-based model-agnostic framework that enhances the quality and interpretability of topic modeling for short-text public replies to CDC communications during the COVID-19 pandemic.

Methods

Study Design

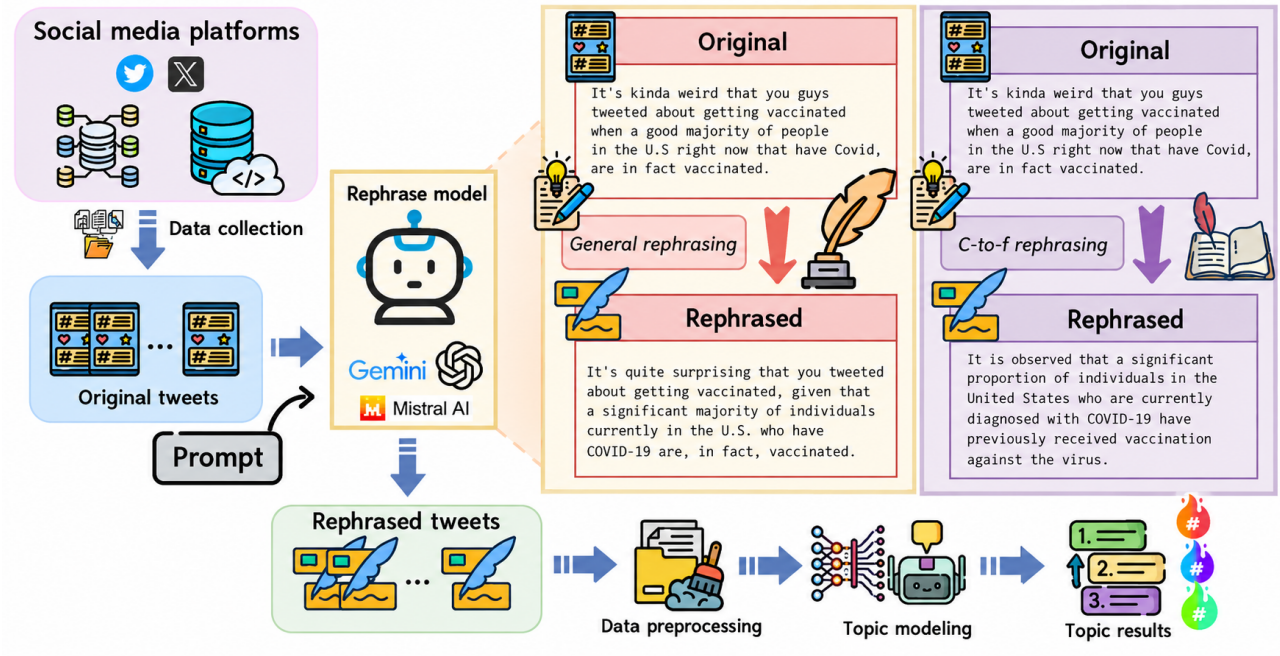
TM-Rephrase is designed as a multistage pipeline to improve topic modeling performance in public health communications, as shown in Figure 1. The pipeline is composed of four major stages: (1) data collection, (2) LLM-based rephrasing, (3) data preprocessing, and (4) topic modeling with a selected algorithm and evaluation of the resulting topics, as detailed in the following subsections. We used this pipeline to study how rephrased text of public responses to CDC tweets on Twitter impacts topic modeling outcomes compared to nonrephrased ones.

Before introducing the model details, we begin by formulating the problem setting with preliminary notations that are used throughout this paper. Let D={d1,d2,…,dN} denote a corpus of *N* short-text tweets (ie, replies to CDC tweets). Each document di is typically brief and context-limited because of the inherent sparsity of social media discourse.

The goal of topic modeling in this setting is to extract and model a set of *K* latent topics, denoted by T={T1,T2,…,TK}, where each topic Tk is represented by a set of its top-*T* representative keywords, extracted or generated by the corresponding topic modeling algorithm, that is, Tk={w1,w2,…,wT},k=1,2,…,K. These representative keywords serve as the semantic expression of the topic and form the basis for downstream topic interpretation, evaluation, and further analysis.

XSL•FO
RenderX

Figure 1. Overview of the TM-Rephrase framework. The pipeline integrates data collection, LLM–based rephrasing, data preprocessing, and topic modeling algorithm application with evaluation.



Data Collection

We constructed a dataset of public responses to official CDC communications on Twitter during the COVID-19 pandemic, covering the period from May 2020 to November 2022. Data were collected via the Twitter Academic API and search queries (Section 1: data search queries in Multimedia Appendix 1), tracking 7 official CDC accounts: @CDCgov, @CDCDirector, @CDCGlobal, @CDCMMWR, @CDCtravel, @DrReddCDC, and @CDCEmergency. The initial retrieval yielded 180,090 public interactions associated with these CDC accounts, including replies and direct mentions. The retrieved metadata included tweet posting dates, tweet text, tweet account ID, account type, public engagement metrics, referenced tweet type, and referenced tweet IDs. Amazon Web Services was used as the computing or storage environment for managing the collected data. For the purpose of this study, only tweets that could be linked to an original CDC tweet were retained. Approximately 49,000 direct mentions that referenced a CDC account but were not replies to a specific CDC tweet were removed. For example, a tweet that mentions @CDCEmergency independently from a user's own account was not considered a reply to the CDC communication. Another 104,000 reply-like interactions were further excluded as they could not be reliably linked to an accessible original CDC tweet, primarily because the original CDC tweet was deleted, unavailable, or otherwise inaccessible at the time of data processing. This step was necessary to preserve conversational context between each public reply and the corresponding CDC communication.

The final analytic dataset consisted of 25,027 public replies linked to 1,512 unique original CDC tweets. These replies capture public opinions, concerns, and discussions in response to official CDC communications during a major public health emergency. We retained only English-language tweets. Statistics of the final original and rephrased datasets are presented in Table 1.

Table . Statistics of the original and rephrased datasets of public replies to CDC[a] communications during the COVID-19 pandemic.

Quantity of original and rephrased datasets	Original	General rephrased	C-to-f[b] rephrased
Mean number of words	27.33	27.76	28.14
SD of word counts	14.23	15.12	14.11
Minimum number of words	1	1	1
25th percentile number of words	15	16	16
Number of words, median (IQR)	27 (15-40)	27 (16-41)	29 (16-42)
75th percentile number of words	40	41	42
Maximum number of words	64	66	68

[a]CDC: US Centers for Disease Control and Prevention.

[b]C-to-f: colloquial to formal.

XSL•FO
RenderX

Formally, the dataset D={d1,d2,…,dN} denotes the original corpus, where each short-text reply is denoted as:

(1)di=(w1i, w2i, …,wLii), i=1,2,…, N

which contains a limited number of tokens Li of tweet di as each tweet reply is limited to 280 characters [4]. The short and informal nature of these short texts results in sparse lexical co-occurrence signals, motivating the need for semantic enrichment before topic modeling algorithms are applied.

LLM-Based Rephrasing

The second stage introduces an LLM-based rephrasing operator that transforms each short-text document (ie, a public reply to a CDC tweet) into a semantically and contextually refined version that is similar to CDC's official communication. Let Rθ denote a rephrase model (ie, an LLM) parameterized by θ. For each short-text reply di, the rephrasing process is defined as:

(2)d^i=Rθ(π,di), i=1,2,…,N

where di^ refers to the rephrased short-text reply, and π denotes a carefully designed prompt that constrains the generation process to preserve semantic fidelity and ensure that rephrasing is executed under a certain scheme that will be introduced in this subsection. This rephrasing stage is crucial for the downstream topic modeling of the public replies and for better understanding of public responses to official health communications during emergencies.

For the rephrasing stage, we experimented with 3 different LLMs: Google Gemini (Gemini 2.5 Flash [21]), OpenAI generative pretrained transformer (GPT-4o mini [22]), and an open-source LLM (Mistral-7B-Instruct [23]), given their strong natural language generation capabilities and complementary trade-offs between performance and cost for large-scale applications. We used a low sampling temperature (0.2) to ensure semantic fidelity and minimize stochastic variation during automated rephrasing, consistent with findings from prior work [24].

To systematically investigate how different styles of linguistic refinement influence topic modeling performance, TM-Rephrase incorporates 2 distinct prompt-guided rephrasing schemes that we developed: general rephrasing and colloquial-to-formal rephrasing. Although both schemes are designed to preserve semantic fidelity, they differ in the degree and style of linguistic transformation applied to the original short texts of public replies. Prompts for these 2 schemes are shown in Table S1 of Multimedia Appendix 1 in Section 2: prompt information for rephrasing schemes.

The general rephrasing scheme performs lightweight linguistic refinement while strictly maintaining the meaning and informational content of the original text. Its primary objective is to improve grammatical correctness, syntactic clarity, and overall readability without altering domain-specific terminology, named entities, hashtags, usernames, or technical details. Concretely, the general rephrasing scheme fixes typographical errors, incomplete sentence fragments, inconsistent verb tenses, and ambiguous phrasing, all of which are common in social media discourse. It may restructure fragmented clauses into complete sentences and resolve minor ambiguities that arise from informal writing conventions. However, it avoids introducing new concepts, removing key terms, or substituting specialized vocabulary. As such, general rephrasing can be viewed as a minimal-intervention normalization process that enhances structural regularity while preserving the lexical distribution of domain-relevant tokens.

The colloquial-to-formal rephrasing scheme transforms informal, conversational, or slang-based expressions into formal, professional English suitable for public health communication contexts. The rephrased outputs are structured to resemble the tone and style of public health reports or professional summaries, using complete sentences and standardized grammar while avoiding slang, contractions, and casual phrasing. During this process, implicit references—such as omitted subjects, abbreviated expressions, or context-dependent meanings (eg, "this shot," "they said," or fragmentary statements lacking clear referents)—are made explicit by clarifying the intended subject, action, or relationship within the sentence. Fragmented or elliptical constructions are rewritten into fully articulated and logically coherent statements to improve readability and interpretability. Strict constraints are applied to preserve the original meaning, including retaining all named entities, hashtags, usernames, and domain-specific terminology. No additional claims, interpretations, or external information are introduced.

These two rephrasing schemes enable a controlled comparison between moderate linguistic normalization and more substantial stylistic formalization for health communications. This design allows us to evaluate whether incremental grammatical refinement alone suffices to improve topic modeling performance, or whether deeper formalization is required for optimal topic quality. In order to ensure semantic alignment between the original and rephrased texts, the prompt explicitly instructs the LLMs to retain meaning and avoid modifying domain-specific content.

The resulting rephrased corpus is defined in equation 3, where di^ represents each individual rephrased short-text reply:

(3)D^={d^1,d^2,…,d^N}

Expert Validation of Semantic Fidelity

To evaluate whether LLM-based rephrasing preserved the meaning, stance or intent, and tone of the original public replies, we further conducted a human expert validation study. A sample of 30 original CDC-reply tweets was randomly selected for validation. For each original tweet, 2 rephrased versions were evaluated: one generated using the general rephrasing scheme and one generated using the colloquial-to-formal scheme, resulting in 60 original-rephrased text pairs across 2 schemes.

Four expert raters with complementary expertise in health communication, social media research, linguistics, and public health independently evaluated the rephrased texts. Each rater compared the original tweet with the corresponding rephrased version in a paired format and provided a single holistic semantic-fidelity rating using a 5-point Likert scale. Raters were instructed to consider whether the rephrased version preserved the original meaning, stance or intent, and tone, but these

XSL•FO
RenderX

dimensions were not rated separately. As the task required direct comparison between original and rephrased texts, complete blinding to the rephrasing scheme was not feasible. Raters completed the evaluations independently. We summarized the validation results using the mean, median, and distribution of expert ratings. To assess interrater reliability, we calculated a 2-way random-effects, absolute-agreement intraclass correlation coefficient (ICC[2,4]) for the aggregated expert ratings.

Data Preprocessing

Before applying topic modeling, both the original and rephrased corpora undergo a consistent and standardized preprocessing procedure to ensure comparability and reduce extraneous noise. The preprocessing step includes the removal of URLs, punctuation marks, and stop words, which do not contribute meaningful semantic information to topic inference during topic modeling.

At the same time, semantically meaningful elements are carefully preserved. Emojis are retained through their textual descriptions to maintain their affective or contextual signals. For example, the emoji is converted to a textual representation “smiling face,” and is mapped to “face with medical mask.” This conversion ensures that nontextual, yet informative elements remain encoded within the lexical space rather than be discarded as noise.

All texts are further normalized by converting all characters to lowercase, followed by tokenization and lemmatization. Lowercasing eliminates case-based lexical duplication, tokenization segments text into analyzable units, and lemmatization reduces inflected forms of tokens to their base form, thereby consolidating semantically equivalent variants into unified lexical representations. All these operations promote lexical consistency and stabilize downstream word co-occurrence statistics in topic modeling.

Formally, let $P_\gamma(\cdot)$ denote the preprocessing operator parameterized by γ, which encapsulates the sequence of normalization, tokenization, and lemmatization in the data preprocessing step. After preprocessing, each original document di and rephrased document di^ are respectively converted into the preprocessed format:

$$(4) p_i = P_\gamma(d_i),\ \hat{p}_i = P_\gamma(\hat{d}_i),\ i = 1,2,\ldots,N$$

Here $p_i \in R^{|V|}$ and $\hat{p}_i \in R^{|V|}$ denote the tokenized representations of the original and the rephrased documents (ie, short-text replies), respectively, under a shared vocabulary space V. Maintaining identical preprocessing across both corpora ensures that any observed differences in topic modeling performance can be credited to the rephrasing rather than discrepancies in preprocessing or representation learning.

Topic Models

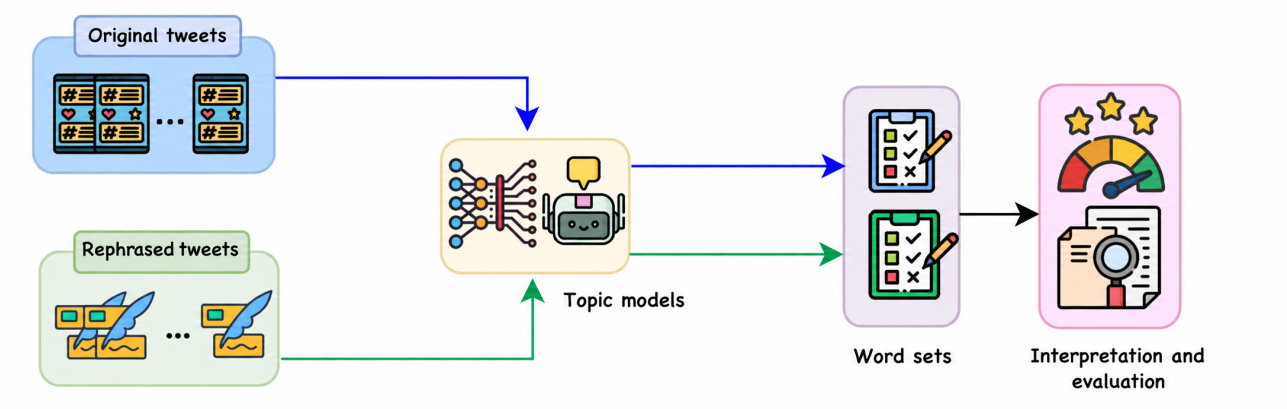
In the final stage of the pipeline, topic modeling algorithms are applied, and their performance is systematically evaluated. As illustrated in Figure 2, both the original and rephrased replies are modeled by the same topic models, and the resulting topics are compared and evaluated through quantitative metrics and expert semantic fidelity evaluation. The original corpus D and rephrased corpus D^ are independently input into the same specific topic modeling algorithm MTM to ensure methodological consistency:

$$(5) T = MTM(D),\ \hat{T} = MTM(\hat{D})$$

This parallel modeling design yields 2 corresponding sets of topic outputs, that is, one derived from the original texts and the other from the rephrased texts. This enables a direct and controlled comparison. Due to the same topic modeling procedure, hyperparameter settings for topic models and preprocessing configurations are maintained across both corpora. Therefore, any observed differences in topic quality can be credited to the rephrasing scheme rather than the specific topic modeling algorithm.

To comprehensively evaluate the effects of TM-Rephrase, 4 representative topic models were applied, including conventional statistical and neural network-based topic models. This suite of different topic models enables a systematic analysis of how different topic modeling algorithms respond to TM-Rephrase.

Figure 2. Outline of the topic evaluation process. Both the original and rephrased tweets were modeled using the same topic models, and the resulting topics were compared and evaluated through quantitative metrics and expert semantic fidelity evaluation.



LDA [6] is a foundational probabilistic topic model serving as the benchmark in topic modeling studies. It is effective for classic topic discovery but is usually challenged by data sparsity in short social media texts.

BERTopic [9] is a neural topic model using contextual embeddings (eg, Sentence-BERT [25]) and density-based clustering. It is considered well-suited for noisy, informal social media data.

XSL•FO
RenderX

FASTopic [11] is based on a transformer backbone that directly captures semantic relationships between document embeddings and learnable topic-word embeddings. It uses an embedding transport plan as an optimization objective to enhance topic-word and document-topic associations.

Topic-semantic contrastive topic model (TSCTM) [12] addresses short-text sparsity via contrastive learning, generating dense semantic vectorized representations and topic distributions.

Together, these different topic models ensure that the evaluation of TM-Rephrase is not limited to a single topic modeling paradigm, but captures methodological robustness across multiple topic models. For all 4 topic models, we fixed the number of topics (*K*) at 8, based on an exploratory analysis to balance thematic detail and interpretability for the study dataset. For each topic, represented as a set of keywords, we retained the top 15 keywords , ranked according to the model's topic-word probability. Details of the experimental implementation are available in Section S3: implementation details in Multimedia Appendix 1.

Topic Modeling Performance Evaluation

Overview

Four quantitative evaluation metrics were computed to quantify topic coherence, uniqueness, redundancy, and diversity, with details shown below.

Topic Coherence (Cv)

Cv [26] measures the semantic coherence among the topic keywords, which are determined by topic-word probabilities. This metric was chosen based on the mathematical analysis and justifications in the study by Wu [27]. This metric fundamentally relies on normalized pointwise mutual information (NPMI) to quantify pairwise word semantic associations. For any 2 words wi and wj, their NPMI score is defined as:

$$(6)\text{NPMI}(w_i, w_j)=\log(P(w_i, w_j)+ P(w_i)P(w_j)+)-\log(P(w_i, w_j)+)$$

Here, P(wi) and P(wj) are the probabilities of observing words wi and wj, respectively, and P(wi,wj) is the co-occurrence probability of the word pair (wi,wj) within a defined external corpus (the Wikipedia [Wikimedia Foundation, Inc] corpus was used in this study [28]). A small constant (eg, 10-12) is added to avoid division by 0 or taking the logarithm of 0. Building on the NPMI formulation, the Cv coherence score can be formally expressed as follows:

$$(7)\text{Cv}=1T\sum i=1T\cos(\text{vNPMI}(xi), \text{vNPMI}(\{xj\}j=1T))$$

Here, vNPMI(xi) represents the vector of NPMI scores between a word xi and all other words in the topic keyword list, defined as:

$$(8)\text{vNPMI}(xi)=\{\text{NPMI}(xi, xj)\}j=1,\ldots,T$$

Similarly, the aggregated NPMI vector across all words in the topic keyword list is defined as:

$$(9)\text{vNPMI}(\{xj\}j=1T)=\{\sum i=1T\text{NPMI}(xi, xj)\}j=1,\ldots,T$$

where *T* is the number of top keywords in the topic (*T*=15 in this study). The cosine similarity is then computed between the word-specific and aggregated vectors, and the final Cv score is obtained by averaging across all words. A higher Cv score indicates stronger semantic coherence and better interpretability of the topics.

Topic Uniqueness

Nan et al [29] proposed topic uniqueness (*TU*), a metric that quantifies how distinct the top keywords within each topic are across the entire topic set T. This helps identify whether individual topics use distinctive vocabulary. Given *K* topics and the top *T* keywords of each topic, *TU* is computed as:

$$(10)\text{TU}=1K\sum k=1K(1T\sum xi\in Tk1\#(xi))$$

where Tk denotes the top keyword set of the *k*-th topic, and #(xi) denotes the occurrence count of word xi in the top *T* words of all topics. *TU* ranges from *1* /K to 1, and a higher *TU* score indicates more unique topics.

Topic Redundancy

Topic redundancy (*TR*) [30] was developed to measure the overlap of top keywords across different topics. The formula for *TR* is given as:

$$(11)\text{TR}=1K\sum k=1K(1T\sum xi\in Tk\#(xi)-1K-1)$$

A lower *TR* score suggests that the topics are more distinct.

Topic Diversity

The formula for the topic diversity *(TD)* [31], which measures the proportion of unique top keywords across all topics, is defined as:

$$(12)\text{TD}=1K\sum k=1K1T\sum xi\in TkI(\#(xi))$$

A higher *TD* score indicates more diverse topics with less word overlap. This metric works by identifying words that are unique to a single topic's top keyword list. The indicator function I(·) is defined by a stepwise function. This function returns 1 if and only if a word xi appears in the top keyword list of exactly 1 topic, and 0 otherwise.

These 4 metrics, Cv, *TU*, *TR*, and *TD*, evaluate topic quality across complementary dimensions. In brief, Cv measures semantic coherence among a topic's top keywords, *TU* assesses topic uniqueness across topics, *TR* quantifies topic redundancy due to topical overlap, and *TD* evaluates whether the model produces a diverse set of topics. These metrics provide a comprehensive set of measures for evaluating and comparing topics generated from original nonrephrased tweets and those rephrased through 2 different TM-Rephrase schemes (general and colloquial-to-formal rephrasing).

Through systematic comparisons between generated topics based on original and rephrased texts respectively, we evaluated the extent to which the model-agnostic TM-Rephrase framework improves the intratopic coherence and intertopic quality.

Ethical Considerations

This study analyzed publicly accessible online tweets from Twitter that were replies to official CDC communications during the COVID-19 pandemic. To comply with the platform's privacy policy, data collection was limited to publicly accessible tweets. The social media data component did not involve direct interaction with Twitter users, recruitment of social media users, intervention, or collection of private information beyond

XSL•FO
RenderX

publicly available tweet content. Informed consent from social media users was not obtained because the analysis was based solely on publicly accessible online tweets.

This study, including the human expert validation component, was reviewed by the Office of Research Protections and Integrity at the University of North Carolina at Charlotte and received a Notice of Determination of Exemption. This study was determined to meet the exempt category cited under 45 CFR 46.104(d), Exemption Category 2. The institutional review board study number is IRB-26 - 1225, titled "Human Evaluation of Topic Modeling Performance for Text Analysis of Public Health-Related Social Media Short Texts," with an approval date of July 8, 2026 (Figure S1 in Multimedia Appendix 1 in Section S4: notice of determination of exemption from IRB Office in UNC Charlotte).

To minimize privacy risks, no usernames, profile information, or other direct user identifiers are reported in this paper. Example tweets, when presented, are used only for methodological illustration, and the main findings are reported in aggregate form. Expert validation results are also summarized in aggregate, without identifying individual raters.

Results

Overview

This section presents the results of quantitative evaluation of TM-Rephrase, followed by illustrative examples. We report results across four topic models, FASTopic, BERTopic, TSCTM, and LDA, under three conditions, as follows: (1) original tweets without any rephrasing, (2) general rephrasing, and (3) colloquial-to-formal rephrasing, with three different LLMs used for rephrasing (ie, Gemini 2.5 Flash, GPT-4o mini, and Mistral-7B-Instruct). Topic quality was quantitatively assessed using 4 metrics: Cv, *TU*, *TR*, and *TD*, as described in the Topic Modeling Performance Evaluation section.

Before introducing the quantitative results, we refer to Table 1 to assess potential text expansion or compression caused by TM-Rephrase. Both general and colloquial-to-formal rephrasing schemes hardly change token lengths: the mean increases marginally from 27.33 (original, SD 14.23) to 27.76 (general, SD 15.12) and 28.14 (colloquial-to-formal, SD 14.11), and the median changes slightly from 27 (IQR 15-40) to 29 (colloquial-to-formal, IQR 16-42). Quartiles and maxima show similar minor differences, indicating that rephrasing preserves the overall text length distribution, and any effects on topic modeling are unlikely to stem from token counts.

Quantitative Results

Overview

Table 2 reports the quantitative evaluation results under a 3-factor design, including 4 topic modeling algorithms, 3 LLM backends, and 2 rephrasing schemes plus the original baseline, resulting in a total of 28 configurations. This design enables a systematic assessment of the main effects of rephrasing, as well as its interaction with model architecture and LLM choice.

XSL•FO
RenderX

Table . Quantitative metric results for various topic models, both with and without TM-Rephrase by three LLMs[a,b].

	Cv(0 - 1) ↑[c]	*TU*[d] (0.125 - 1) ↑	*TR*[e] (0 - 1) ↓	*TD*[f] (0 - 1) ↑
FASTopic[g] w/o[h] rephr[i,j]	.3388	.9917	.0024	.975
Gemini 2.5 Flash				
FASTopic w/[k] general rephr[j]	.3723	.9917	.0024	.9917
FASTopic w/ c-to-f rephr[j]	.3301	1[l]	0[l]	1[l]
GPT-4o mini				
FASTopic w/ general rephr	.3746[l]	.9917	.0024	.9917
FASTopic w/ c-to-f rephr	.3422	1[l]	0[l]	1[l]
Mistral-7B-Instruct				
FASTopic w/ general rephr	.3728	.9917	.0024	.9917
FASTopic w/ c-to-f rephr	.3321	.9917	.0024	.9917
BERTopic[m] w/o rephr	.4078	.4667	.3976	.4667
Gemini 2.5 Flash				
BERTopic w/ general rephr	.4564	.525[l]	.3357[l]	.525[l]
BERTopic w/ c-to-f rephr	.4734[l]	.525[l]	.3452	.525[l]
GPT-4o mini				
BERTopic w/ general rephr	.4612	.4667	.3357[l]	.4667
BERTopic w/ c-to-f rephr	.4687	.525[l]	.3452	.525[l]
Mistral-7B-Instruct				
BERTopic w/ general rephr	.4574	.4667	.3357[l]	.4667
BERTopic w/ c-to-f rephr	.4713	.4667	.3976	.4667
TSCTM[n] w/o rephr	.3094	.9917	.0024	.975
Gemini 2.5 Flash				
TSCTM w/ general rephr	.3145	.9833	.0048	.9833
TSCTM w/ c-to-f rephr	.3394[l]	1[l]	0[l]	1[l]
GPT-4o mini				
TSCTM w/ general rephr	.3267	.9917	.0024	.9917
TSCTM w/ c-to-f rephr	.3331	1[l]	0[l]	1[l]
Mistral-7B-Instruct				
TSCTM w/ general rephr	.3104	.9833	.0024	.9833
TSCTM w/ c-to-f rephr	.3213	.9917	.0024	.975

XSL•FO
RenderX

	Cv(0 - 1) ↑[c]	*TU*[d] (0.125 - 1) ↑	*TR*[e] (0 - 1) ↓	*TD*[f] (0 - 1) ↑
LDA[o] w/o rephr	.3094	.575[l]	.3095	.575[l]
Gemini 2.5 Flash				
LDA w/ general rephr	.4206	.5583	.3214	.5583
LDA w/ c-to-f rephr	.5004[l]	.575[l]	.3048[l]	.575[l]
GPT-4o mini				
LDA w/ general rephr	.4197	.5583	.3214	.5583
LDA w/ c-to-f rephr	.4917	.575[l]	.3095	.575[l]
Mistral-7B-Instruct				
LDA w/ general rephr	.4008	.5583	.3095	.5583
LDA w/ c-to-f rephr	.4762	.5583	.3095	.5583

[a]LLM: large language model.

[b]Table rows of TM-Rephrase are results based on Gemini 2.5 Flash, GPT-4o mini, and Mistral-7B-Instruct, respectively.

[c]Arrows in the header indicate the desired direction for each metric (higher ↑ or lower ↓ is better).

[d]TU: topic uniqueness.

[e]TR: topic redundancy.

[f]TD: topic diversity.

[g]FASTopic: fast, adaptive, stable, and transferable topic model.

[h]w/o: without.

[i]Rephr stands for the rephrasing.

[j]General rephr stands for the general rephrasing scheme, while c-to-f rephr stands for the colloquial-to-formal rephrasing scheme.

[k]w: with.

[l]The best results within each model group.

[m]BERTopic: BERT-based topic modeling.

[n]TSCTM: topic-semantic contrastive topic model.

[o]LDA: latent Dirichlet allocation.

Effect of Rephrasing Schemes

Across most configurations, TM-Rephrase improves topic quality, with the most pronounced and stable gains observed in topic coherence (Cv). However, the magnitude and consistency of improvement vary across topic models, evaluation metrics, and rephrasing schemes. In particular, the colloquial-to-formal rephrasing scheme generally yields the highest coherence scores. For example, with LDA, coherence increases substantially from 0.3094 (without rephrasing) to 0.5004 under Gemini 2.5 Flash, with similarly large improvements observed for GPT-4o mini (0.4917) and Mistral-7B-Instruct (0.4762). These results indicate that rephrasing effectively mitigates contextual sparsity in short texts and enhances semantic consistency across topics.

As Cv coherence relies on word co-occurrence statistics derived from an external reference corpus, its values may be influenced by how closely the topic keywords align with the lexical patterns of that corpus. In this study, the reference corpus was Wikipedia, which may better represent standardized or encyclopedic expressions than informal social media language. Therefore, colloquial-to-formal rephrasing may partly improve Cv by shifting topic keywords toward lexical patterns more characteristic of Wikipedia, rather than solely by improving the substantive representation of public meaning.

Effect Across Topic Models

The magnitude of improvement varies across topic models. Probabilistic models such as LDA exhibit the largest gains, reflecting their strong dependence on lexical co-occurrence patterns. BERTopic shows consistent but more moderate improvements, with coherence increasing across all LLMs and rephrasing schemes, and reaching its highest value under Gemini 2.5 Flash (Cv=0.4734). In contrast, FASTopic demonstrates relatively smaller and less consistent gains, with slight decreases in coherence under the colloquial-to-formal scheme in some cases (eg, Cv=0.3301 under Gemini 2.5 Flash). This suggests that embedding-based models may already partially capture semantic relationships despite informal language, reducing the effectiveness of input standardization. TSCTM, meanwhile, benefits substantially from rephrasing in terms of diversity-related metrics, achieving perfect topic separation (*TU*=1, *TD*=1, and *TR*=0) under colloquial-to-formal rephrasing with both Gemini 2.5 Flash and GPT-4o mini.

Effect of LLM Backbones

The improvements from TM-Rephrase are broadly similar across all 3 LLM backbones. Although minor variations in absolute performance are observed, the patterns remain clear and consistent: rephrasing improves coherence and diversity metrics across topic models regardless of the underlying LLM. For example, LDA coherence gains and TSCTM diversity

XSL•FO
RenderX

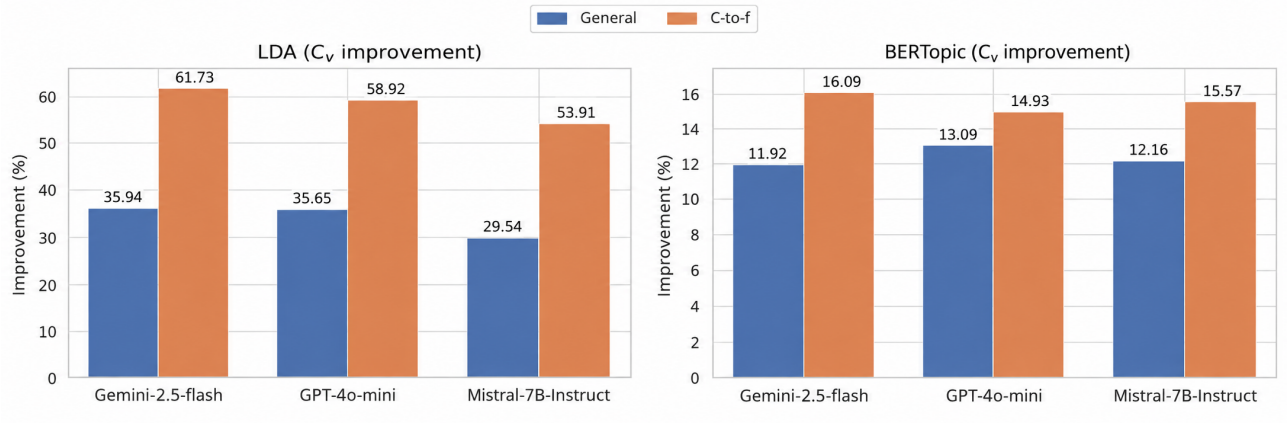
improvements are observed consistently across Gemini 2.5 Flash, GPT-4o mini, and Mistral-7B-Instruct. This stability suggests that the effectiveness of TM-Rephrase is not dependent on a specific LLM, but rather reflects a generalizable rephrasing strategy.

Diversity and Redundancy Metrics

In addition to improving intratopic coherence, rephrasing also improves intertopic quality. Across FASTopic, BERTopic, and TSCTM, *TU* and *TD* generally increase, while *TR* decreases or remains similar. The most notable improvement is observed in TSCTM under colloquial-to-formal rephrasing, where perfect diversity scores are achieved, indicating fully distinct topic clusters. These findings suggest that rephrasing enhances not only intratopic coherence but also intertopic separation.

Figure 3 shows that colloquial-to-formal rephrasing often achieves larger coherence gains than general rephrasing for LDA and BERTopic, while the overall improvement patterns are observed across Gemini 2.5 Flash, GPT-4o mini, and Mistral-7B-Instruct. These results demonstrate the robustness of the TM-Rephrase framework as well.

Figure 3. Cv improvement (%) across large language models (LLMs) and rephrasing schemes for LDA and BERTopic. LDA: latent Dirichlet allocation.



Additionally, a nuanced trade-off between intratopic coherence and inter-*TD* emerges, especially for the colloquial-to-formal rephrasing scheme. While this scheme often delivers the highest coherence and perfect diversity scores, it can occasionally reduce *TU* in certain topic models. For example, colloquial-to-formal rephrasing maximizes coherence across all LLMs but slightly lowers *TU* compared to the nonrephrased baseline in LDA. This suggests that aggressive formalization might homogenize lexical choices across topics, highlighting the importance of matching rephrasing strength with model characteristics. In addition, the comparison across LLM backbones reveals a high degree of robustness in TM-Rephrase. While Gemini 2.5 Flash often achieves the strongest absolute results, GPT-4o mini and Mistral-7B-Instruct closely track its performance across all metrics and models. The relative ranking of rephrasing schemes and topic models remains largely unchanged, indicating that TM-Rephrase is not tightly tied to a specific LLM but instead constitutes a generalizable topic-model-agnostic paradigm.

To demonstrate whether the observed improvements were dependent on the use of Cv coherence alone, we further evaluated topic coherence using 3 additional metrics: CNPMI [26], CUCI [32], and UMASS [33]. These metrics provide complementary views of topic coherence based on word co-occurrence patterns, with higher values indicating better coherence for all 3 metrics. The results are reported in Table 3.

XSL•FO
RenderX

Table . Quantitative coherence metric results for various topic models, both with and without TM-Rephrase by 3 LLMs[a,b].

	CNPMI (-1, 1) ↑[c]	CUCI (-∞, +∞) ↑	UMASS (-∞, 0] ↑
FASTopic[d] w/o[e] rephr[f]	−0.033	−1.477	−5.035
Gemini 2.5 Flash			
FASTopic w/[g] general rephr[h]	−0.021[i]	−1.461	−4.671[i]
FASTopic w/ c-to-f rephr[h]	−0.036	−1.456[i]	−4.818
GPT-4o mini			
FASTopic w/ general rephr	−0.027	−1.471	−4.833
FASTopic w/ c-to-f rephr	−0.031	−1.466	−4.874
Mistral-7B-Instruct			
FASTopic w/ general rephr	−0.024	−1.469	−4.726
FASTopic w/ c-to-f rephr	−0.035	−1.465	−4.819
BERTopic[j] w/o rephr	.028	.38	−2.358
Gemini 2.5 Flash			
BERTopic w/ general rephr	.049[i]	.401	−2.18
BERTopic w/ c-to-f rephr	.045	.386	−2.163[i]
GPT-4o mini			
BERTopic w/ general rephr	.031	.411[i]	−2.212
BERTopic w/ c-to-f rephr	.042	.387	−2.319
Mistral-7B-Instruct			
BERTopic w/ general rephr	.044	.385	−2.338
BERTopic w/ c-to-f rephr	.03	.381	−2.314
TSCTM[k] w/o rephr	−0.021	−1.065	−3.512
Gemini 2.5 Flash			
TSCTM w/ general rephr	−0.017	−0.955	−3.141
TSCTM w/ c-to-f rephr	−0.013[i]	−0.876[i]	−2.949
GPT-4o mini			
TSCTM w/ general rephr	−0.015	−1.032	−2.941[i]
TSCTM w/ c-to-f rephr	−0.016	−0.944	−2.997
Mistral-7B-Instruct			
TSCTM w/ general rephr	−0.02	−0.912	−3.217
TSCTM w/ c-to-f rephr	−0.018	−0.998	−3.111
LDA[l] w/o rephr	.017	−0.084	−2.457
Gemini 2.5 Flash			
LDA w/ general rephr	.038	.282	−2.401
LDA w/ c-to-f rephr	.071[i]	.848[i]	−2.146
GPT-4o mini			
LDA w/ general rephr	.069	.193	−2.39
LDA w/ c-to-f rephr	.058	.764	−2.418
Mistral-7B-Instruct			
LDA w/ general rephr	.047	.242	−2.06[i]

XSL•FO
RenderX

	CNPMI (-1, 1) ↑[c]	CUCI (-∞, +∞) ↑	UMASS (-∞, 0] ↑
LDA w/ c-to-f rephr	.066	.656	−2.279

[a]LLM: large language model.

[b]Table rows of TM-Rephrase are results based on Gemini 2.5 Flash, GPT-4o mini, and Mistral-7B-Instruct, respectively.

[c]Arrows in the header indicate the desired direction for each metric (higher ↑ or lower ↓ is better).

[d]FASTopic: fast, adaptive, stable, and transferable topic model.

[e]w/o: without.

[f]Rephr stands for rephrasing.

[g]w/: with.

[h]General rephr stands for the general rephrasing scheme, while c-to-f rephr stands for the colloquial-to-formal rephrasing scheme.

[i]The best results within each model group.

[j]BERTopic: BERT-based topic modeling.

[k]TSCTM: topic-semantic contrastive topic model.

[l]LDA: latent Dirichlet allocation.

The results support the main finding that TM-Rephrase generally improves topic coherence, although the magnitude of improvement varies across topic models, LLM backbones, and rephrasing schemes. The greatest and most consistent improvements are observed for LDA. For example, under Gemini 2.5 Flash, LDA improves from 0.017 to 0.071 on CNPMI, from −0.084 to 0.848 on CUCI, and from −2.457 to −2.146 on UMASS after colloquial-to-formal rephrasing. Similar improvements are also observed for GPT-4o mini and Mistral-7B-Instruct, suggesting that the coherence gains for LDA are not specific to a single LLM backbone.

For BERTopic and TSCTM, the results also show improved coherence after rephrasing, although the best-performing rephrasing scheme differs by metric and LLM. For BERTopic, rephrased texts improve CNPMI and CUCI in most settings, with Gemini 2.5 Flash general rephrasing achieving the highest CNPMI score and GPT-4o mini general rephrasing achieving the highest CUCI score. For TSCTM, both general and colloquial-to-formal rephrasing improve coherence relative to the baseline without rephrasing across most metrics, with colloquial-to-formal rephrasing under Gemini 2.5 Flash performing best on CNPMI and CUCI.

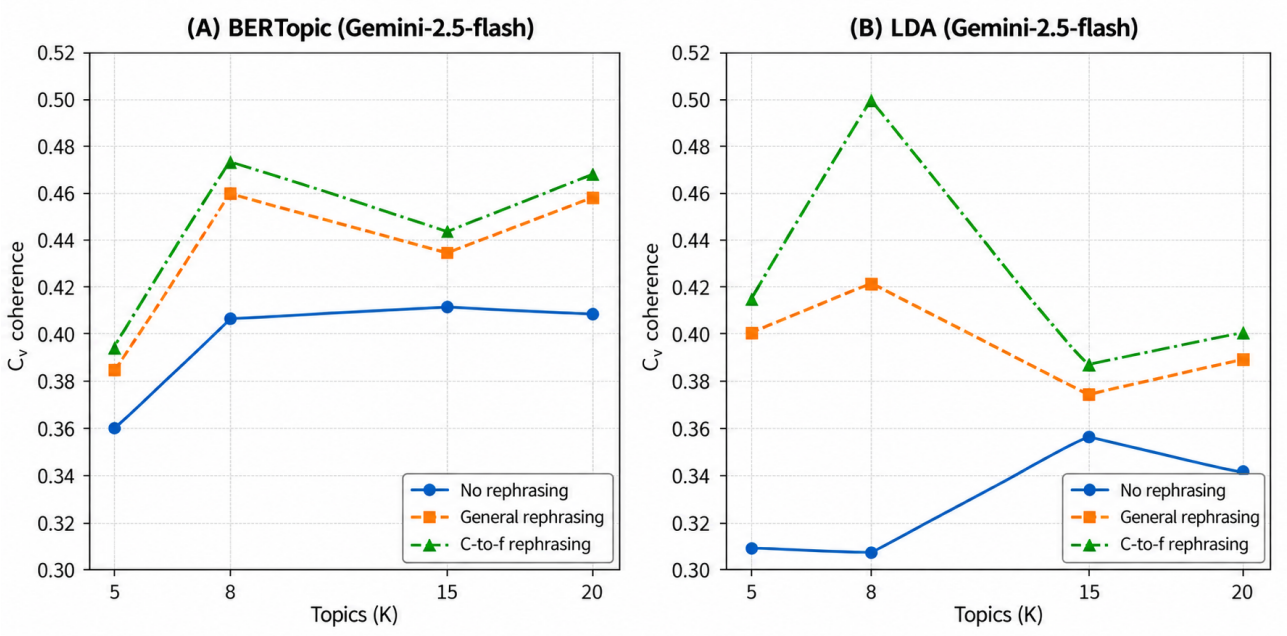
The effects are less consistent for FASTopic. General rephrasing improves CNPMI and UMASS in several settings, while colloquial-to-formal rephrasing does not consistently improve CNPMI relative to the nonrephrased baseline. This suggests that embedding-based or neural topic models may already capture some semantic relationships in noisy short texts, making them less sensitive to input-level formalization than LDA. Therefore, these results support a more cautious interpretation: TM-Rephrase generally improves topic coherence, especially for LDA, BERTopic, and TSCTM, but the magnitude of effects is topic model-dependent.

Sensitivity Analysis Across Different Topic Numbers

As shown in Figure 4, both general and colloquial-to-formal rephrasing improved Cv coherence across all tested K values for both models. For BERTopic, colloquial-to-formal rephrasing improved Cv from 0.3445 to 0.3741 at 5 and from 0.4096 to 0.4682 at 20. For LDA, colloquial-to-formal rephrasing improved Cv from 0.3102 to 0.4093 at 5 and from 0.3411 to 0.4002 at 20. These results suggest that the coherence gains from TM-Rephrase are not limited to the original 8 setting, although the magnitude of improvement varies across topic numbers and topic models.

XSL•FO
RenderX

Figure 4. Sensitivity of topic coherence to number of topics and rephrasing schemes for (A) BERTopic and (B) LDA, based on Gemini 2.5 Flash rephrasing backbone. BERTopic: BERT-based topic modeling; LDA: latent Dirichlet allocation.



These quantitative results suggest that TM-Rephrase generally enhances intratopic coherence and can improve intertopic quality metrics for short, noisy texts in health communications, although the effects are model- and metric-dependent. The 2 different rephrasing schemes that we developed have similar overall performance. The general rephrasing scheme offers balanced and reliable improvements, while the colloquial-to-formal scheme has a slight edge, especially for lexically sensitive topic models such as LDA and BERTopic. These findings demonstrate the effectiveness and robustness of TM-Rephrase as an LLM-based model-agnostic framework for topic modeling in health communications.

Results of Expert Ratings of Rephrasing Semantic Fidelity

The expert validation results (Table 4) provided additional evidence regarding the semantic fidelity of the 2 LLM-based rephrasing schemes. Across all 240 expert ratings, the mean semantic fidelity score was 3.90 (SD 1.01), and the median score was 4.00 (IQR 3.00-5.00) out of 5.0, suggesting that rephrasing generally preserved the original meaning.

Table . Results of expert validation on rephrasing semantic fidelity.

Validation scope	Ratings	Mean (SD)	Median (IQR)	Ratings ≥4 (%)	ICC(2,4)[a], average-measure
Overall	240	3.90 (1.01)	4.00 (3.00-5.00)	65.42	0.736
General rephrasing	120	4.38 (0.78)	5.00 (4.00-5.00)	87.50	0.613
C-to-f[b] rephrasing	120	3.42 (0.99)	3.00 (3.00-4.00)	43.33	0.553

[a]ICC: intraclass correlation coefficient.

[b]C-to-f: colloquial-to-formal.

However, the 2 rephrasing schemes showed different fidelity patterns. General rephrasing received higher expert ratings, with a mean score of 4.38 (SD 0.78), a median score of 5.00 (IQR 4.00-5.00), and 87.50% (105/120) of ratings of 4 or above. In contrast, the colloquial-to-formal scheme received a lower mean score of 3.42 (SD 0.99), a median score of 3.00 (IQR 3.00-4.00), and 43.33% (52/120) of ratings of 4 or above. This result suggests that general rephrasing better preserved the original meaning, stance, or intent, and tone, whereas colloquial-to-formal rephrasing introduced greater changes because of its stronger formalization.

As meaning, stance, or intent, and tone were evaluated through a single holistic rating rather than separate dimension-specific scores, the validation results cannot determine quantitatively which aspect of fidelity contributed most to the lower ratings for colloquial-to-formal rephrasing. Therefore, the lower colloquial-to-formal ratings should be interpreted as indicating a greater risk of fidelity change overall, rather than as definitive evidence that tone or stance or intent alone was the primary source of reduced fidelity. However, 1 expert noted that meaning and stance or intent were often similar while tone differed, suggesting that professionalization of tone may be one important contributor.

The overall average interrater reliability was ICC(2,4)=0.736, indicating acceptable agreement among expert raters for the aggregated semantic fidelity ratings. Scheme-specific intraclass

XSL•FO
RenderX

correlation coefficient values were lower, with ICC(2,4)=0.613 for general rephrasing and ICC(2,4)=0.553 for colloquial-to-formal rephrasing. This pattern suggests that expert judgments were more variable when evaluating the degree to which formalized rephrasing preserved tone and stance, which is consistent with the greater linguistic transformation involved in the colloquial-to-formal scheme. Generally, these findings support a more nuanced interpretation: general rephrasing provides stronger preservation of the original public voice, whereas colloquial-to-formal rephrasing offers clearer formalization but involves a possible trade-off in tone and rhetorical fidelity.

Illustrative Topic-Level Examples

To contextualize the quantitative findings, Tables 5 and 6 provide illustrative examples of topic keywords and topic assignments before and after rephrasing.

XSL•FO
RenderX

Table . Top 15 keywords for LDA[a] topics based on Gemini 2.5 Flash (K=8)[b].

Topic ID	Keywords
w/o[c] rephrasing	
1	vaccine, covid, get, people, dont[d], mask, kid, cdc, child, need, school, vaccinated, like[d], work, trump
2	mask, covid, wear, virus, vaccine, wearing, people, stop, dont[d], flu, need, face, work, social, spread
3	covid, death, vaccine, cdc, people, case, day[d], many[d], child, one[d], dont,[d] symptom, week[d], died, year[d]
4	vaccine, effect, dose, people, covid, know, covaxin, dont[d], pfizer, shot, like[d], mrna, child, long, one[d]
5	covid, death, rate, case, cdc, data, infection, people, immunity, vaccination, study, show[d], variant, state, number
6	vaccine, cdc, covid, pfizer, stop, fda, people, transmission, please[d], health, year[d], child, approved, public, prevent
7	covid, ivermectin, dont[d], mask, cdc, people, work, vaccine, stop, positive, tested, like[d], getting[d], dose, doctor
8	vaccine, covid, vaccinated, get[d], flu, still[d], shot, people, child, year[d], got[d], even[d], booster, fully[d], getting[d]
w/[e] general rephrasing scheme	
1	covid, cdc, vaccine, death, please[d], mask, pandemic, trump, virus, could[d], american, health, public, regarding, trust
2	mask, vaccinate, wear, covid, please[d], personal_protective_equipment, get[d], vaccine, school, individual, society, social_distance, virus, wearing, still[d]
3	covid, vaccine, individual, flu, vaccinated, people, case, virus, immunity, child, stop, positive, tested, get[d], one[d]
4	vaccine, covid, ivermectin, mrna, pfizer, people, lie, individual, death, effective, health, covaxin, received, moderna, treatment
5	vaccine, covid, child, shot, booster, effective, mrna, prevent, flu, please, receive, cdc, people, covaxin, variant
6	mask, covid, day[d], cases, test, people, virus, individual, cdc, one[d], work, vaccine, hospital, positive, wear
7	covid, child, vaccine, death, risk, vaccination, data, year[d], people, cdc, rate, individual, adverse, virus, cause
8	vaccine, risk, side_effect, testing, tweet, pfizer, would[d], covid, child, vaccination, kid, also[d], hospital, concern, original
w/ c-to-f[f] rephrasing scheme	
1	mask, covid, individual, vaccine, public, operation, vaccinated, state, virus, health, wear, public_measure, may, work, mandate
2	covid, child, vaccine, individual, vaccination, regarding, death, concern, risk, year[d], virus, age, variant, may, significant
3	covid, individual, vaccine, positive, test, statement, author, mask, vaccination, testing, result, current, case, information, president
4	disease, cdc, prevention, control, center, covid, tweet, ivermectin, individual, health, professional, public, treatment, travel, coronavirus

XSL•FO
RenderX

Topic ID	Keywords
5	vaccine, covid, effect, associated, data, symptom, adverse, cdc, child, efficiency, vaccination, side_effect, potential, disease, death
6	individual, vaccine, vaccination, covid, vaccinated, regarding, may, immunity, infection, health, efficacy, influenza, public, risk, concern
7	vaccine, vaccination, covid, individual, booster, pfizer, received, pharmaceutical, dose, trust, cdc, company, fda, financial, administration
8	public, health, regarding, covid, concern, pandemic, vaccine, social_distance, significant, information, current, mask, trump, cdc, individual

[a]LDA: latent Dirichlet allocation.

[b]Results with rephrasing are based on the input texts rephrased through Gemini 2.5 Flash.

[c]w/o: without.

[d]Irrelevant keywords (off-topic words identified through author interpretation).

[e]w/: with.

[f]C-to-f: colloquial-to-formal.

XSL•FO
RenderX

Table . Illustrative examples of original, general, and c-to-f[a] rephrased tweet replies with corresponding LDA[b] and BERTopic[c] topic assignments[d].

Text type	Text content	Topic assigned by LDA	Topic assigned by BERTopic
Example 1			
Original	Are you kidding me with this? Don't be stupid enough to let your perfect baby be a guinea pig for these people. People are getting bells palsy after getting the vaccine.	vaccine[e], covid, get, people, dont, mask, kid[e], cdc, child[e], need, school, vaccinated[e], like, work, trump	covaxin, novavax, vaccine[e], approve, mrna, booster, ichoosecovaxin, need, want, fda, safe[e], people, child[e], wait, ocugen
General rephrased	Are you serious? Don't be foolish and allow your healthy child to be a test subject for these individuals. Reports indicate people are developing Bell's palsy after receiving the vaccine.	vaccine[e], risk[e], side_effect[e], testing[e], tweet, pfizer, would, covid, child[e], vaccination[e], kid[e], also, hospital, concern[e], original	vaccine[e], covaxin, omicron, novavax, variant, mrna, delta, booster, approve, child[e], vaccinate, side_effect[e], people, risk[e], concern[e]
C-to-f rephrased	Concerns have been raised regarding potential adverse effects, specifically Bell's palsy, following vaccination. It is imperative to approach public health recommendations and medical interventions with informed decision-making.	vaccine[e], covid, effect, associated, data, symptom, adverse[e], cdc, child[e], efficiency, vaccination[e], side_effect[e], potential, disease, death	vaccine[e], covaxin, novavax, mrna, approval, subject[e], receive, adverse[e], booster, vaccination, risk[e], efficacy, child[e], side_effect[e], concern[e]
Example 2.			
Original	Yes - let's trust Pfizer - same company that entered a $2.3 billion criminal plea deal for lying! Same company that is exempt from liability and CANT be sued for any injuries - yup sign me right up *eye roll*	vaccine[e], effect, dose, people, covid, know, covaxin, dont, pfizer[e], shot, like, mrna, child, long, one	pfizer[e], vaccine[e], effect, mrna, covid, cdc, moderna, people, shot, child, test, get, know, stop, receive
General rephrased	Should we trust Pfizer? The same company that previously entered a $2.3 billion criminal plea deal for lying is now exempt from liability and cannot be sued for any injuries. So, yes, sign me right up. *eye roll*	vaccine[e], covid, ivermectin, mrna, pfizer[e], people, lie[e], individual, death, effective, health, covaxin, received, moderna, treatment	vaccine[e], pfizer[e], individual, vaccination, trial, mrna, concern[e], lie[e], report, utility, public, moderna, cdc, fda, information
C-to-f rephrased	Concerns exist regarding public trust in Pfizer, given its history, including a $2.3 billion criminal plea deal for deceptive practices. Furthermore, the company's exemption from liability and inability to be sued for injuries raises significant questions regarding accountability.	vaccine[e], vaccination[e], covid, individual, booster, pfizer[e], received, pharmaceutical, dose, trust[e], cdc, company[e], fda, financial, administration	vaccine[e], pfizer[e], individual, vaccination[e], responsibility[e], mrna, concern[e], profit[e], business, trust[e], public, moderna, cdc, fda, information

[a]C-to-f: colloquial-to-formal.

[b]LDA: latent Dirichlet allocation.

[c]BERTopic: BERT-based topic modeling.

[d]Results with rephrasing are based on the input texts rephrased through Gemini 2.5 Flash.

[e]Words that better describe topics.

These examples are intended to demonstrate how TM-Rephrase may affect interpretability. In interpreting these examples, keywords were considered more informative when they were specific, domain-relevant, and clearly aligned with public health themes, whereas generic terms, colloquial artifacts, and weakly related words were treated as less informative.

In Table 5, LDA-derived topics without rephrasing exhibit limited semantic coherence, particularly for topic 2 (public health measures), which is dominated by generic and less informative terms (eg, "mask," "wear," and "stop") and includes off-topic noise such as "dont." General rephrasing improves topical specificity by introducing more contextually meaningful health-related terms (eg, "personal_protective_equipment" and "social_distance"), although some colloquial artifacts persist (eg, "please" and "could").

In contrast, colloquial-to-formal rephrasing produces more coherent and structured topics, characterized by standardized and policy-relevant terminology (eg, "mandate" and "public_measure"). A similar pattern is observed in vaccine-related topics (eg, topic 5), where baseline outputs contain vague terms (eg, "like" and "get"), while rephrasing, particularly for colloquial-to-formal, yields more precise and

XSL•FO
RenderX

analytically meaningful keywords (eg, “adverse,” “associated,” and “efficiency”). These examples suggest that increasing linguistic formality and contextual explicitness can reduce less informative terms and make selected topic outputs easier to interpret in relation to public health discourse.

Table 6 further illustrates how rephrasing may affect the apparent semantic alignment between selected replies and their assigned topics. In Table 6, rephrasing improves the semantic alignment between tweet content and assigned topics across both LDA and BERTopic. In example 1, the original tweet expresses concerns about vaccine side effects using informal and fragmented language, resulting in generic or weakly focused topic keywords (eg, “vaccine” and “child”). General rephrasing introduces explicit health-related terms such as “side effect,” “risk,” and “concern,” improving thematic alignment. Colloquial-to-formal rephrasing further strengthens this alignment by incorporating more formal and domain-specific terminology (eg, “adverse”), enabling both models to produce more coherent and semantically grounded topics. A similar pattern is observed in example 2, where the original tweet contains colloquial and emotionally charged expressions regarding Pfizer.

Topic assignments from the original text include noisy and less informative terms (eg, “people,” “dont,” and “get”). General rephrasing improves relevance by introducing clearer semantic cues (eg, “concern”), while colloquial-to-formal rephrasing further refines the discourse with more structured and policy-relevant vocabulary (eg, “trust,” “responsibility,” and “profit”), allowing topic models, particularly BERTopic, to capture higher-level institutional themes.

Discussion

Principal Findings

This study developed and evaluated TM-Rephrase, a model-agnostic framework that uses LLM-based rephrasing to standardize short, informal public health texts while aiming to preserve their original meaning. Using 25,027 public replies to CDC tweets on Twitter during the COVID-19 pandemic, we found that rephrasing improved topic coherence and topic distinctiveness while generally preserving semantic fidelity across multiple topic modeling approaches.

For RQ1 (whether and how effectively TM-Rephrase improves the quality and interpretability of topic model outputs for short texts in health communications), the results show that TM-Rephrase improves topic quality and interpretability for short, noisy social media texts. The strongest coherence gains were observed for LDA, where Cv increased from 0.3094 to 0.5004, suggesting that rephrasing helps mitigate lexical sparsity and fragmented wording. Improvements were also observed for embedding-based approaches such as BERTopic, indicating that the benefits of rephrasing are not limited to one type of topic model. Across multiple evaluation metrics, including Cv, CNPMI, CUCI, UMASS, *TU*, *TR*, and *TD*, the results suggest that TM-Rephrase produces topic-word distributions that are more coherent, less redundant, and easier to interpret.

However, these improvements should be interpreted cautiously. Automated coherence metrics may partly reward the lexical regularity introduced by rephrasing, and higher topic coherence does not necessarily mean that all meaningful public-response signals are preserved. In infoveillance contexts, sarcasm, distrust, anger, uncertainty, and oppositional framing are not merely noise; they may represent important public health signals. Therefore, we interpret TM-Rephrase as improving the readability and interpretability of topic-model outputs, rather than as fully validating the substantive meaning of public responses.

For RQ2 (how different rephrasing schemes influence topic model performance), the 2 rephrasing schemes showed different effects. General rephrasing produced balanced and consistent improvements while better preserving the original public voice. In contrast, the colloquial-to-formal scheme often produced stronger gains in coherence and topic distinctiveness, especially for lexically sensitive models such as LDA and TSCTM. At the same time, colloquial-to-formal rephrasing involves greater linguistic transformation and may alter tone, rhetorical force, or framing. This suggests that prompt design should be aligned with the analytic goal: general rephrasing may be preferable when preserving public expression is important, while colloquial-to-formal rephrasing may be useful when the goal is to maximize topical clarity and model interpretability.

For RQ3 (how to interpret public discussions, that is, replies to CDC during the COVID-19 pandemic, based on the topics and inform public health practice), the improved topic representations helped clarify public responses to CDC communications. Rephrased topics more clearly captured recurring concerns related to vaccination attitudes, perceived risks, preventive behaviors, mistrust of official guidance, and uncertainty about public health recommendations. These findings suggest that TM-Rephrase can support public health social listening by helping analysts organize large volumes of noisy public replies into more interpretable thematic structures.

Practically, TM-Rephrase may serve as a human-in-the-loop preprocessing strategy for public health social listening rather than a replacement for expert interpretation. By reducing lexical fragmentation and improving the readability of topic-model outputs, it may help analysts organize large volumes of noisy public replies and identify themes that warrant closer review, such as vaccine safety concerns, mistrust of official guidance, confusion about preventive measures, or perceived gaps between public messaging and lived experience. However, this study does not directly evaluate whether TM-Rephrase improves public health decisions, communication outcomes, or detection of otherwise missed concerns. Therefore, its public health utility should be interpreted as a potential application that requires further validation in real-world infoveillance and communication workflows.

Generally, this study provides evidence that LLM-based rephrasing can enhance topic modeling of short public health texts, while also highlighting the need for careful validation. Future work should further evaluate semantic fidelity through larger-scale human expert review, topic intrusion tasks, posttopic

XSL•FO
RenderX

alignment assessments, and cross-platform testing across different public health contexts.

Limitations and Future Work

This study has several limitations that should be addressed in future research. First, during dataset curation, we did not perform additional bot or spam classification beyond the filtering described in Data Collection, because the primary objective was to evaluate the effect of rephrasing on topic modeling outputs rather than to infer public opinion. The filtering process may also introduce selection bias. In particular, replies linked to deleted or inaccessible CDC tweets were excluded, and these excluded interactions may differ systematically from the retained replies. Therefore, the final corpus should be interpreted as a dataset of public replies with recoverable CDC conversational context, rather than as a complete representation of all public interactions with CDC accounts or all COVID-19 discourse on Twitter. In addition, because the analysis focuses on replies to official CDC communications on Twitter during the COVID-19 pandemic, future research should test TM-Rephrase across other platforms, health topics, population groups, and public health communication contexts before applying it in operational surveillance or social listening systems.

Second, although LLM-based rephrasing is designed to preserve semantic meaning while improving linguistic clarity, it may introduce an important trade-off for public health infoveillance. Rephrasing can reduce lexical fragmentation and improve topic coherence, but it may also soften or alter meaningful public-discourse signals, including sarcasm, anger, distrust, misinformation cues, uncertainty, oppositional framing, and colloquial expressions. These features should not be treated merely as noise, because they may reflect public resistance, confusion, risk perception, or distrust toward health institutions.

The expert semantic-fidelity validation provides partial evidence regarding this trade-off. Overall, expert ratings suggested generally favorable preservation of meaning, stance or intent, and tone; however, general rephrasing received stronger semantic-fidelity ratings than colloquial-to-formal rephrasing. This suggests that stronger formalization may be more likely to alter tone, rhetorical force, or stance expression. Therefore, automated improvements in coherence should not be interpreted as definitive evidence of improved substantive interpretability. In infoveillance applications, original and rephrased texts should be examined together, especially when sarcasm, distrust, anger, uncertainty, misinformation cues, or oppositional framing are central to the RQ. In addition, the expert validation used a single holistic semantic-fidelity rating that asked raters to consider meaning, stance or intent, and tone together; therefore, this study cannot quantitatively determine which specific dimension of fidelity was most affected by rephrasing. Future validation should rate meaning, stance or intent, tone, uncertainty, sarcasm, and distrust as separate dimensions to better identify which aspects of public discourse are most vulnerable to meaning drift or tone normalization.

Recent AI alignment research suggests that differences across LLM backbones may extend beyond linguistic performance. Lau et al [34] found that LLMs can vary in value-priority profiles and alignment with human judgments. This is relevant to TM-Rephrase because rephrasing public discourse may involve implicit choices about tone, framing, emphasis, and socially sensitive meanings. Thus, although our results show broadly similar topic-modeling improvements across the LLMs tested, future work should examine whether different LLMs preserve meaning, stance, tone, uncertainty, and distrust in comparable ways, especially for sensitive public-health replies.

Moreover, this study used a fixed number of topics and a selected set of representative topic modeling approaches to enable consistent comparison between original and rephrased texts. Although this design supports methodological evaluation, public health communication is dynamic and context-dependent. Future work could adapt TM-Rephrase to better capture evolving public concerns across different stages of health emergencies.

Conclusions

This study highlights the value of improving input text quality for online public health infoveillance and analytics, particularly those based on social media. By leveraging LLM-based rephrasing, TM-Rephrase enables more interpretable, consistent, and reliable detection of thematic patterns in short, noisy public responses to official health communications during health emergencies.

The data-centric and model-agnostic TM-Rephrase framework also offers a practical and scalable solution to enhance the usability of large-scale social media data for infoveillance. The findings suggest that linguistic standardization plays an important role in strengthening the interpretability of analytical outputs and supporting a clearer understanding of public concerns from large volumes of raw online data.

As social media continues to serve as a key channel for public engagement on various health topics, novel analytical frameworks, such as TM-Rephrase, enable more effective monitoring of public discourse and better-informed public health communication strategies, especially during health emergencies.

Acknowledgments

The authors gratefully acknowledge Dr Sijia Qian and Dr Ron Lunsford for participating as expert validators in the semantic fidelity assessment of original and large language model–rephrased social media tweets. Their feedback helped evaluate whether the rephrased versions preserved the meaning, stance or intent, and tone of the original tweets. The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy [35]; 2025), the following tasks were delegated to GenAI tools under full human supervision: selection of research methods, proofreading and editing, and reformatting. The GenAI tool used was ChatGPT-5.5 (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the outcomes.

XSL•FO
RenderX

Funding

This study was supported by the National Science Foundation (DMS-2436227). The funding organization had no role in this study's design, data collection and analysis, decision to publish, or preparation of this paper. The opinions, findings, conclusions, and recommendations presented in this paper are those of the authors alone and do not necessarily represent the views of the funding organization.

Data Availability

The datasets analyzed during this study consist of public replies to Centers for Disease Control and Prevention communications on Twitter (subsequently rebranded as X). The raw tweet-level social media datasets are not publicly shared because redistribution of user-generated social media content on Brandwatch may be restricted by platform terms and because public tweets may still contain content that could raise privacy or reidentification concerns. Aggregated results and derived materials supporting the findings are available from the corresponding author upon reasonable request, subject to applicable platform policies and ethical considerations.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Queries, prompts, implementation details, and notice of the institutional review board exemption determination.
[PDF File, 363 KB - infodemiology_v6i1e98319_app1.pdf]

References

1. Rauchfleisch A, Vogler D, Eisenegger M. Public sphere in crisis mode: how the COVID-19 pandemic influenced public discourse and user behaviour in the Swiss Twitter-sphere. Javnost 2021;28(2):129-148. [doi: 10.1080/13183222.2021.1923622] [Medline: 34393592]
2. Zhang C, Xu S, Li Z, Hu S. Understanding concerns, sentiments, and disparities among population groups during the COVID-19 pandemic via Twitter data mining: large-scale cross-sectional study. J Med Internet Res 2021 Mar 5;23(3):e26482. [doi: 10.2196/26482] [Medline: 33617460]
3. Yin S, Chen S, Ge Y. Dynamic associations between Centers for Disease Control and Prevention social media contents and epidemic measures during COVID-19: infoveillance study. JMIR Infodemiology 2024 Jan 23;4(1):e49756. [doi: 10.2196/49756] [Medline: 38261367]
4. Timeline of X. Wikipedia. URL: https://en.wikipedia.org/wiki/Timeline_of_Twitter [accessed 2026-09-10]
5. Laureate CDP, Buntine W, Linger H. A systematic review of the use of topic models for short text social media analysis. Artif Intell Rev 2023 May 1;56:1-33. [doi: 10.1007/s10462-023-10471-x] [Medline: 37362887]
6. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
7. Hong L, Davison BD. Empirical study of topic modeling in Twitter. Presented at: Proceedings of the First Workshop on Social Media Analytics; Jul 25-28, 2010; Washington, DC p. 80-88. [doi: 10.1145/1964858.1964870]
8. Rajagopal D, Olsher D, Cambria E, Kwok K. Commonsense-based topic modeling. Presented at: Proceedings of the second international workshop on issues of sentiment discovery and opinion mining; Aug 11-12, 2013; Chicago, IL p. 1-8. [doi: 10.1145/2502069.2502075]
9. BERTopic. URL: https://maartengr.github.io/BERTopic/ [accessed 2026-09-10]
10. Angelov D. Top2Vec: distributed representations of topics. arXiv. Preprint posted online on Aug 19, 2020. [doi: 10.48550/arXiv.2008.09470]
11. Wu X, Nguyen T, Zhang DC, Wang WY, Luu AT. FASTopic: pretrained transformer is a fast, adaptive, stable, and transferable topic model. Presented at: Advances in Neural Information Processing Systems 37; Dec 10-15, 2024; Vancouver, BC, Canada p. 84447-84481. [doi: 10.52202/079017-2683]
12. Wu X, Luu AT, Dong X. Mitigating data sparsity for short text topic modeling by topic-semantic contrastive learning. Presented at: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; Dec 7-11, 2022; Abu Dhabi, United Arab Emirates p. 2748-2760. [doi: 10.18653/v1/2022.emnlp-main.176]
13. Yang JZ, Liu Z, Wong JC. Information seeking and information sharing during the COVID-19 pandemic. Commun Q 2022 Jan 1;70(1):1-21. [doi: 10.1080/01463373.2021.1995772]
14. AbuRaed AGT, Prikryl EA, Carenini G, Janjua NZ. Long COVID Discourse in Canada, the United States, and Europe: topic modeling and sentiment analysis of Twitter data. J Med Internet Res 2024 Dec 9;26:e59425. [doi: 10.2196/59425] [Medline: 39652387]
15. James L, McPhail H, Foisey L, Donelle L, Bauer M, Kothari A. Exploring communication by public health leaders and organizations during the pandemic: a content analysis of COVID-related tweets. Can J Public Health 2023 Aug;114(4):563-583. [doi: 10.17269/s41997-023-00783-4] [Medline: 37349662]

XSL•FO
RenderX

16. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill 2020 May 29;6(2):e19273. [doi: 10.2196/19273] [Medline: 32427106]
17. Maini P, Seto S, Bai R, Grangier D, Zhang Y, Jaitly N. Rephrasing the web: a recipe for compute and data-efficient language modeling. Presented at: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1; Bangkok, Thailand p. 14044-14072. [doi: 10.18653/v1/2024.acl-long.757]
18. Zhang Y, Jin H, Meng D, Wang J, Tan J. A comprehensive survey on automatic text summarization with exploration of LLM-based methods. Neurocomputing 2026 Jan;663:131928. [doi: 10.1016/j.neucom.2025.131928]
19. Miah MSU, Kabir MM, Sarwar TB, Safran M, Alfarhood S, Mridha MF. A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. Sci Rep 2024 Apr 26;14(1):9603. [doi: 10.1038/s41598-024-60210-7] [Medline: 38671064]
20. Liu Z, Zhou Y, Zhu Y, et al. Information retrieval meets large language models. Presented at: Companion Proceedings of the ACM Web Conference 2024; May 13-17, 2024; Singapore, Singapore p. 1586-1589. [doi: 10.1145/3589335.3641299]
21. Gemini 25 flash model. Google Cloud. URL: https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-flash [accessed 2026-09-10]
22. GPT-4o mini: advancing cost-efficient intelligence. OpenAI. URL: https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/ [accessed 2026-09-10]
23. Mistral-7B-instruct-v02. Hugging Face. URL: https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.2 [accessed 2026-09-10]
24. Renze M. The effect of sampling temperature on problem solving in large language models. Presented at: Findings of the Association for Computational Linguistics: EMNLP 2024; Nov 12-16, 2024; Miami, FL p. 7346-7356. [doi: 10.18653/v1/2024.findings-emnlp.432]
25. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using siamese BERT-networks. Presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); Hong Kong, China p. 3982-3992. [doi: 10.18653/v1/D19-1410]
26. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. Presented at: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining; Feb 2-6, 2015. [doi: 10.1145/2684822.2685324]
27. Wu X. Towards effective neural topic modeling. : Nanyang Technological University; 2024 URL: https://dr.ntu.edu.sg/bitstreams/2b7465e1-7444-4cda-9078-9e7163cd4006/download [accessed 2026-09-10]
28. English Wikipedia database dump. Wikimedia Foundation. URL: https://dumps.wikimedia.org/enwiki/latest/ [accessed 2026-09-10]
29. Nan F, Ding R, Nallapati R, Xiang B. Topic modeling with Wasserstein autoencoders. Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Jul 28 to Aug 2, 2019; Florence, Italy p. 6345-6381. [doi: 10.18653/v1/P19-1640]
30. Burkhardt S, Kramer S. Decoupling sparsity and smoothness in the Dirichlet variational autoencoder topic model. J Mach Learn Res 2019;20(131):1-27 [FREE Full text]
31. Dieng AB, Ruiz FJR, Blei DM. Topic modeling in embedding spaces. Trans Assoc Comput Linguist 2020 Dec;8:439-453. [doi: 10.1162/tacl_a_00325]
32. Newman D, Lau JH, Grieser K. Automatic evaluation of topic coherence. Presented at: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics; Jun 2-4, 2010 URL: https://aclanthology.org/N10-1012.pdf [accessed 2026-09-10]
33. Mimno D, Wallach H, Talley E. Optimizing semantic coherence in topic models. Presented at: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing; Jul 27-31, 2011. [doi: 10.5555/2145432.2145462]
34. Lau GR, Low WY, Koh SM, Nah FFH, Hartanto A. Evaluating AI alignment in LLMs: output analysis of value priorities across 75 models with human benchmarking. arXiv. Preprint posted online on May 16, 2026. [doi: 10.48550/arXiv.2506.12617]
35. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (Generative AI Delegation Taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res 2026 Apr;33(3):2544331. [doi: 10.1080/08989621.2025.2544331] [Medline: 40781729]

Abbreviations

BERTopic: BERT-based topic modeling
CDC: Centers for Disease Control and Prevention
FASTopic: fast, adaptive, stable, and transferable topic model
ICC: intraclass correlation coefficient
LDA: latent Dirichlet allocation
LLM: large language model
NPMI: normalized pointwise mutual information
RQ: research question

XSL•FO
RenderX

TD: topic diversity
TR: topic redundancy
TSCTM: topic-semantic contrastive topic model
TU: topic uniqueness

Edited by T Mackey; submitted 15.Apr.2026; peer-reviewed by GR Lau, H Son, N Gevorgyan; revised version received 25.Aug.2026; accepted 31.Aug.2026; published 08.Oct.2026.

Please cite as:
Xin W, Yin S, Park A, Ge Y, Chen S
Identifying Public Response Topics to the Centers for Disease Control and Prevention's COVID-19 Communications on Social Media: Infoveillance Study Using Large Language Model–Based Rephrasing
JMIR Infodemiology 2026;6:e98319
URL: https://infodemiology.jmir.org/2026/1/e98319
doi:10.2196/98319

© Wangjiaxuan Xin, Shuhua Yin, Albert Park, Yaorong Ge, Shi Chen. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 8.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Temporal Dynamics of Influenza-Associated Anxiety Symptom Linguistic Markers on Weibo (2023-2024): Observational Study

Yifan Ou[1,2], MPhil; Gert-Jan de Bruijn[1], PhD; Peter Johannes Schulz[3,4,5], PhD

[1]Department of Communication Science, University of Antwerp, Sint-Jacobstraat 2, Antwerp, Belgium

[2]School of Social Sciences, University of Manchester, Manchester, England, United Kingdom

[3]Faculty of Communication, Culture, and Society, Università della Svizzera italiana, Lugano, Switzerland

[4]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

[5]Department of Communication & Media, Ewha Womans University, Seoul, Republic of Korea

Corresponding Author:
Yifan Ou, MPhil
Department of Communication Science, University of Antwerp, Sint-Jacobstraat 2, Antwerp, Belgium

Abstract

Background: Influenza seasons may be associated with increased anxiety-related expressions on social media. Social media can reflect population-level emotional expression patterns in real time.

Objective: The aim of the study is to characterize diurnal and full-season dynamics of anxiety-related language during the 2023 - 2024 influenza season in China and its association with influenza activity.

Methods: We retrieved Sina Weibo posts in February 2025 covering September 4, 2023, to April 28, 2024. Posts containing *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*)–based anxiety terms were cleaned and deduplicated (N=169,728 → 106,440). We first linked weekly influenza incidence with anxiety-related postings. Then, we plotted diurnal patterns by epidemiologic phase, conducted supplementary within-sample hourly normalization analyses, and modeled longitudinal symptom trajectories using ARIMA (autoregressive integrated moving average) and supplementary ARIMAX (autoregressive integrated moving average with exogenous regressors) time-series models.

Results: Anxiety-related posts closely followed influenza activity, surging during the outbreak and peak phases and remaining elevated even after influenza declined. Initial Spearman correlation analyses showed significant negative associations for irritability (r=−0.413; P=.02) and restlessness or feeling keyed up or on edge (r=–0.396; P=.02). However, supplementary ARIMAX analyses further revealed that being easily fatigued and difficulty concentrating or mind going blank exhibited more stable positive temporal associations with influenza activity after controlling for autocorrelation and lagged effects. Diurnal patterns shifted across stages, showing mild early-evening variation during the outbreak, clear morning peaks with secondary afternoon and evening rises during prevalence and decline, and morning-afternoon concentration in the end stage. Supplementary within-sample hourly normalization analyses showed that the major diurnal structures remained generally stable after normalization. ARIMA time-series analysis revealed that irritability and being easily fatigued consistently dominated the discussions, whereas others remained at relatively low levels. Out-of-sample forecasting based on a chronological 80% training and 20% testing split suggested generally stable short-term trajectories, with being easily fatigued showing a slight increase.

Conclusions: This study demonstrates how social media can capture diurnal and seasonal fluctuations of anxiety symptoms associated with influenza activity, advancing understanding of affective dynamics in population health.

(JMIR Infodemiology 2026;6:e88849) doi:10.2196/88849

KEYWORDS

influenza season; Weibo; anxiety symptoms; linguistic marker; affective dynamics

Introduction

In the digital era, the collection, use, and analysis of social media data have created new opportunities across diverse domains. For example, it can be used to monitor public opinion, track consumer behavior, and follow emerging social issues. In the field of mental health, social media likewise provides a novel pathway for the early identification of problems. Given the limited accessibility of traditional mental health services and the reluctance of many individuals to actively seek professional help when experiencing psychological distress, exploring alternative approaches to detection and intervention is particularly important. The emergence of social media platforms offers a promising avenue to address this challenge [1].

Social media has become an important channel for individuals to express emotions [2,3], document daily life, and reflect

XSL•FO
RenderX

psychological states [4], with its texts often embedding psychological cues such as emotional fluctuations, cognitive biases, and social behavior patterns [4-9]. For instance, De Choudhury et al [5], using Twitter data, demonstrated that users' linguistic features and behavioral patterns can reveal their psychological states, highlighting the potential of social media texts for understanding depressive symptoms.

Subsequent studies have extended this line of research. Zhu et al [8] and Yu et al [7] analyzed discussions on anxiety and depression disorders on Sina Weibo and identified their linguistic features, further supporting the capacity of social media data to reflect collective perceptions and attitudes toward mental health issues.

Compared with traditional surveys or clinical interviews, social media data offer unique advantages for identifying linguistic markers of public mental health: (1) real-time monitoring, which captures daily fluctuations; (2) naturalistic expression, which reduces social desirability bias; and (3) accessibility, as platform APIs enable large-scale and longitudinal tracking [10]. These features make social media an important complement to conventional screening tools and provide new possibilities for the early identification of mental health problems [11].

Some early studies have demonstrated this potential. For example, Hung and Tadius [12] used Twitter data labeled as "depressed" and "non-depressed" to compare various machine learning methods for distinguishing between different types of texts. The results showed that one approach achieved superior accuracy and was able to automatically detect potential depression-related expressions in users' writing. This suggests that social media text data can serve as linguistic markers of mental health symptoms [4].

Mental health symptoms, particularly anxiety, exhibit pronounced temporal dynamics. Anxiety does not emerge instantaneously but rather unfolds through processes of accumulation, fluctuation, and recovery in response to environmental changes [13]. Accordingly, longer observation periods are essential to reveal its developmental mechanisms and interactions with contextual factors. Among the possible temporal frameworks, the influenza season offers unique advantages. Unlike transient social events, influenza seasons typically last for several months and proceed through distinct stages, from preoutbreak to recession. This cyclical progression provides a natural context for examining the dynamic shifts in linguistic markers and exploring the interplay between infectious disease outbreaks and collective anxiety-related expressions. In the Chinese context, the 2023 - 2024 influenza season holds significance that extends beyond traditional seasonal influenza research. First, it represents the first complete influenza cycle following China's full withdrawal of nonpharmaceutical interventions against COVID-19. The strict social distancing and lockdown policies implemented during the pandemic [14,15] markedly reduced influenza incidence and population immunity between 2019 and 2022 [16], thereby heightening the intensity and health risks of this influenza season compared with previous years. Second, in the postpandemic phase of psychological recovery, the public may perceive influenza—a disease that is both familiar and threatening—in novel ways, potentially eliciting distinctive patterns of cognition and emotional response [17]. This psychological tension, situated between normalized disease risks and the residual memory of a global health crisis, makes the 2023 - 2024 influenza season an ideal setting for investigating the dynamic evolution of anxiety symptoms [18]. In sum, this unique context offers not only a rare opportunity to examine how collective psychology adapts to new large-scale health threats in the aftermath of COVID-19 but also valuable empirical evidence to inform public health interventions.

Existing literature suggests a close association between influenza and anxiety. Wang and Sabran [19], in a survey conducted during the 2023 - 2024 influenza season, found that the anxiety levels of young individuals were significantly associated with the impact of influenza on physical and mental health as well as quality of life. Similarly, Bults et al [20] reported that during the early phase of the 2009 H1N1 outbreak, public anxiety levels were relatively high but declined over time. Tausczik et al [21], through analyses of textual and browsing behaviors, further showed that the use of anxiety-related words surged at the onset of the pandemic and quickly subsided thereafter. However, most existing studies have focused primarily on the 2009 H1N1 influenza outbreak, while research on the 2023 - 2024 influenza season remains scarce. In addition, prior studies often relied on cross-sectional data or concentrated on peak periods, lacking longitudinal tracking across the recurring stages of the influenza season. Consequently, 2 major limitations emerge: first, the inability to capture the temporal dynamics of emotional changes; and second, the restricted capacity for causal inference, making it difficult to determine whether fluctuations in anxiety levels are directly attributable to influenza itself or influenced by other environmental factors.

Building on this, this study adopts a longitudinal perspective to systematically investigate the evolution of public anxiety symptoms during the 2023 - 2024 influenza season. We collected Sina Weibo data from September 4, 2023, to April 28, 2024, covering the entire influenza cycle—thereby enabling us to comprehensively trace the dynamic changes in anxiety symptoms across different stages through linguistic markers and to analyze their intrinsic connection with the course of influenza.

In addition, this study innovatively incorporates a diurnal variation perspective to provide a more comprehensive understanding of the temporal dynamics of anxiety symptoms. Existing research has shown that diurnal fluctuations are closely related to anxiety [22-26]. Individual chronotypes, reflecting differences in diurnal preferences, further shape these patterns [26,27]. For instance, Cox et al [25], using high-frequency tracking, found that anxiety levels exhibit pronounced within-day variation and differ between morning-type and evening-type individuals. Although based on limited samples, these findings indicate that anxiety symptoms fluctuate along diurnal cycles.

Similar patterns are also reflected in online contexts. For instance, Ten Thij et al [28] observed that individuals with depression expressed significantly more negative emotions—such as hopelessness, anxiety, and sadness—on social media during the midnight-to-early-morning hours,

XSL•FO
RenderX

reflecting the psychological stress and emotional distress they experienced at night. However, most of these studies have focused on depression or comorbid mental health symptoms and disorders, and evidence specifically addressing diurnal variation in anxiety symptoms remains limited. Therefore, exploring the diurnal patterns of anxiety during the influenza season not only deepens our understanding of the temporal dynamics of anxiety but also helps fill gaps in the existing literature.

In summary, this study seeks to advance the field through 3 main contributions. Rather than identifying psychological effects of influenza or establishing predictive mental health signals, the study provides a descriptive temporal analysis of anxiety-related linguistic expressions on social media during a large-scale public health period and aims to characterize how anxiety-related expressions fluctuate across influenza stages and diurnal periods. First, it integrates large-scale social media data with the phased progression of the influenza season to longitudinally trace the dynamics of anxiety symptoms. Second, it introduces a novel diurnal perspective to examine the temporal expression of anxiety. Third, it applies time-series modeling to describe temporal covariation patterns between anxiety-related expressions and influenza activity. Accordingly, the specific research questions are as follows:

- RQ1: How are public anxiety symptoms dynamically associated with influenza incidence during the 2023-2024 influenza season?
- RQ2: During the 2023-2024 influenza season, how do public anxiety symptoms change in terms of diurnal rhythm patterns?
- RQ3: What are the evolutionary trends of different categories of anxiety symptoms over the course of the 2023-2024 influenza season?

By addressing the above questions, this study not only fills gaps in the existing literature but also provides empirical support for public health interventions and real-time mental health monitoring, underscoring the critical value of social media in mental health research [4].

Methods

Overview

This study used multiple steps: first, Weibo anxiety symptom data were cleaned and preprocessed. Second, statistical analysis and visualization indicated anxiety symptom diurnal trends. Third, anxiety symptom progression and influenza incidence rates were examined dynamically. Finally, time-series analysis revealed long-term anxiety symptom dynamics, providing novel research insights.

Data Collection and Preprocessing

The description of generalized anxiety disorder in *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (*DSM-IV*) indicates that anxiety symptoms are primarily manifested in the following aspects: (1) restlessness or feeling keyed up or on edge, (2) being easily fatigued, (3) difficulty concentrating or mind going blank, (4) irritability, (5) muscle tension, and (6) sleep disturbance[29]. We selected these terms, translated them into Chinese, and used them as search keywords, as shown in Table 1.

Therefore, we selected specific terms, translated them into Chinese, and used them as search keywords. We crawled all Weibo posts containing these keywords from September 4, 2023, to April 28, 2024, fine-grained on an hourly basis, resulting in 169,728 posts as the raw data.

Data were collected using a Python-based web crawler that retrieved publicly accessible Sina Weibo posts through keyword-based searches. Keyword matching was based on substring matching rather than exact phrase matching. Only posts primarily written in simplified Chinese were retained. All timestamps were converted to China Standard Time (UTC+8) before hourly aggregation. Exact duplicate posts and reposted or highly repetitive irrelevant content were removed during preprocessing.

To examine the textual features and thematic structures of the posts, we applied Python's re module to remove non-Chinese elements through iterative steps: deleting emojis, Unicode special characters, user mentions (@), hashtags (#), special markers (eg, 【】 and "fold"), redundant whitespace, duplicates, and overly short posts.

XSL•FO
RenderX

Table . Keywords to web crawl[a].

Symptoms name	Translation in Chinese
Restlessness or feeling keyed up or on edge	焦躁不安, 感到紧张不安
Being easily fatigued	容易疲劳
Difficulty concentrating or mind going blank	注意力难以集中, 脑子一片空白
Irritability	易怒
Muscle tension	肌肉紧张
Sleep disturbance	睡眠障碍

[a]To improve linguistic coverage, approximately 200 candidate synonymous expressions across the 6 generalized anxiety disorder symptom categories were initially compiled. A supplementary pilot test using 10 synonymous expressions per category over a 3-day period (January 1 - 3, 2024) yielded 57,062 posts after basic deduplication, illustrating the substantial increase in dataset complexity associated with exhaustive synonym expansion (Multimedia Appendix 1). This approach prioritized conceptual consistency with *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*)–based symptom categories rather than exhaustive coverage of all colloquial expressions and was intended to capture population-level temporal patterns rather than identify clinical anxiety at the individual level.

All preprocessing rules were predefined before the main statistical analyses and applied consistently across the full dataset. The iterative cleaning process primarily involved refinement of rule clarity and quality checks to improve consistency between research assistants, rather than introducing new selection criteria based on observed temporal patterns or study results. Core inclusion and exclusion criteria remained stable throughout the process and were not adjusted for specific periods or outcomes.

Since automated preprocessing alone could not sufficiently filter out highly similar or irrelevant posts, an additional manual verification step was conducted. Additional details of the data collection and preprocessing procedures are provided in Multimedia Appendix 1. Manual verification primarily focused on whether posts reflected genuine anxiety-related expressions rather than metaphorical, commercial, or contextually unrelated uses of the keywords. Manual verification was primarily used as a quality-control step to assess consistency of rule implementation rather than to selectively modify the dataset based on study findings.

Frequency analysis was used to exclude high-frequency posts and irrelevant topics such as commercial advertisements, celebrity promotions, astrology, or unrelated events (eg, World Sleep Day). After several rounds of manual cleaning, we obtained a high-quality dataset of 106,440 posts.

Detailed preprocessing procedures, rule descriptions, and quality-control workflows are provided in Multimedia Appendix 2 to improve transparency and reproducibility.

Ethical Considerations

This study used only publicly accessible Sina Weibo posts and did not involve interaction with human participants or access to private identifiable information. Before conducting the study, the research team consulted the university regarding ethics requirements. Based on the institutional guidance in effect at that time, the study was determined to be exempt from ethics review because it analyzed only publicly available, anonymized social media data. Therefore, institutional review board approval and informed consent were not required. To further protect participant privacy, all collected data were anonymized during preprocessing. No personally identifiable information, including user IDs or IP addresses, was collected or stored during the data collection process.

Analysis Plan

RQ1: Step 1

We first collected and visualized weekly influenza data and anxiety symptom posting frequencies to organize and analyze dynamic changes in anxiety symptom data during the influenza season and identify the underlying correlations with influenza incidence.

PyCharm (JetBrains sro) and Python (version 3.12, Python Software Foundation; os, pandas, matplotlib, datetime, and seaborn) were used. After reading the Microsoft Excel file, we translated the “Post Publication Time” timestamp strings into “Week” and “Date” columns for analysis. We then analyzed and visualized the following:

- We counted weekly anxiety symptom posts by “Week” and “Search Criteria.”
- Using data from the *Influenza Weekly Report* of the Chinese National Influenza Center [30], we computed and graphed the influenza A positivity rate (Multimedia Appendix 3).
- We used matplotlib and seaborn to construct line charts to compare and depict the weekly posting frequencies of anxiety symptoms and influenza positivity rates over time.

Furthermore, to examine the relationship between influenza positivity rates and the posting frequency of various anxiety symptoms, we conducted a Spearman rank correlation analysis in R (version 4.4.1; R Foundation for Statistical Computing, with the *psych* package installed). This method was used to assess the monotonic relationships between variables and to reduce the influence of outliers, given the relatively small sample size (32 weeks). Statistical significance was determined at the threshold of $P<.05$.

To further examine the temporal associations between influenza positivity rates and anxiety-related symptom expressions, while controlling for autocorrelation and potential lag effects within the time series, we additionally constructed ARIMAX (autoregressive integrated moving average with exogenous

XSL•FO
RenderX

regressors) models by incorporating influenza positivity rates as exogenous variables.

The ARIMAX analyses were conducted in R (version 4.4.1) using the auto.arima() function from the *forecast* package. Optimal model orders (p, d, q) were automatically selected based on the Akaike information criterion. Stationarity was assessed using the augmented Dickey-Fuller (ADF) test prior to modeling (Multimedia Appendix 4). As the analyses were based on weekly-level data, the time-series frequency was set to 52. No additional logarithmic or Box-Cox transformations were applied to the raw data [31].

To evaluate potential lagged effects, influenza positivity rates at contemporaneous (lag 0), 1-week lag (lag 1), and 2-week lag (lag 2) intervals were separately incorporated into the models. Model validation was performed using a chronological train or test split, with the first 80% of the data used for training and the remaining 20% reserved as an unseen test set [32]. The auto.arima() procedure for model order selection and parameter estimation was conducted exclusively on the training set, while the test set was used solely for out-of-sample evaluation. Regression coefficients of the exogenous variables and their corresponding 95% CIs were subsequently extracted to construct coefficient effect plots, illustrating the direction, strength, and statistical significance of the associations between influenza positivity rates and anxiety-related symptom expressions under different lag conditions.

RQ2: Step 2

The second step was to divide the influenza process into 3 stages and analyze the diurnal fluctuations of 6 anxiety symptoms within each stage.

Based on turning points in influenza A positivity rates reported by Chinese National Influenza Center weekly surveillance, we defined the outbreak stage as November 6 - 26, 2023, when positivity began to rise sharply; the prevalence-decline stage as November 27, 2023, to February 29, 2024, when positivity remained high before declining; and the end stage as March 1 to April 28, 2024, when positivity returned to baseline [17]. Additionally, we investigated the number of anxiety-related posts per 0 - 24 hours in different stages. These steps are specific:

- Data preprocessing: We read the Excel data file from the provided directory, transformed the “post publication time” column time strings to timestamps, and then sorted by hour to count posts in each group.
- Data statistics: We segmented the data by 1 - 24 hours to count the number of posts for each anxiety symptom, and then summarized the data for the 3 phases to get the overall number of posts in each stage.
- Chart optimization and visualization: We set Matplotlib font options to display Chinese characters correctly, and then drew line charts of diurnal post quantity changes in Seaborn to intuitively portray anxiety symptom post variation trends.

To further examine whether the observed circadian patterns might be influenced by hourly fluctuations in posting activity within the anxiety-related dataset, we additionally conducted a robustness analysis using within-sample hourly normalization [32]. Specifically, within each influenza stage and corresponding hour, the number of posts for a given anxiety symptom category was expressed as the proportion of all anxiety-related posts identified during the same hour. For example, if 100 anxiety-related posts were identified within a given hour and 30 of them belonged to the “irritability” category, the relative expression frequency of irritability for that hour was calculated as 30%.

This within-sample normalization approach helps control, to some extent, for hourly fluctuations in posting activity within the anxiety-related dataset and was used to assess whether the observed circadian patterns remained stable after normalization. The normalized circadian variation plots are provided in Multimedia Appendix 5.

PyCharm was used to write and run all code, assuring analytical efficiency and reproducibility.

RQ3: Step 3

The steps are as follows:

1. Data preprocessing: Timestamps were converted to date format using the as.date () function from the *tidyverse* package, along with the *readxl* package.
2. Generating time-series data: The number of posts for each category was counted by date, and time-series data were generated using the group_by() and summarise() functions.

To characterize the temporal evolution of different anxiety symptoms during the influenza season, we used ARIMA (autoregressive integrated moving average) models for time-series analysis and short-term forecasting. The analyses were conducted in R (version 4.4.1) using the *forecast*, *xts*, and *ggprism* packages. First, timestamp data were converted into date format using the *tidyverse* and *readxl* packages, and posting counts for different anxiety symptom categories were aggregated by date to generate time-series data.

ARIMA models were fitted using the auto.arima() function from the *forecast* package, with optimal model orders (p, d, q) automatically selected based on the Akaike information criterion. The model effectively captures autocorrelated structures in time-series data, enables residuals to approximate white noise, and provides statistically interpretable short-term forecasts [32]. Prior to modeling, stationarity was assessed using the ADF test (Multimedia Appendix 6). The ARIMA models were constructed using daily-level data, with the time-series frequency set to 7. No additional transformations were applied to the raw data [32].

The time-series data were chronologically divided into training and testing sets, with the first 80% used for model training and the remaining 20% reserved as an unseen test set. Model order selection and parameter estimation using the auto.arima() function were performed exclusively on the training set, while the test set was used solely for out-of-sample evaluation of forecasting performance. In addition, a fixed-window backtesting strategy was applied to assess model robustness. Forecasting performance was evaluated using root mean square error, mean absolute error, mean absolute percentage error, and

XSL•FO
RenderX

mean absolute scaled error. Furthermore, the ARIMA models were used for 30-day short-term forecasting to describe the potential temporal trajectories of different anxiety symptom categories [32].

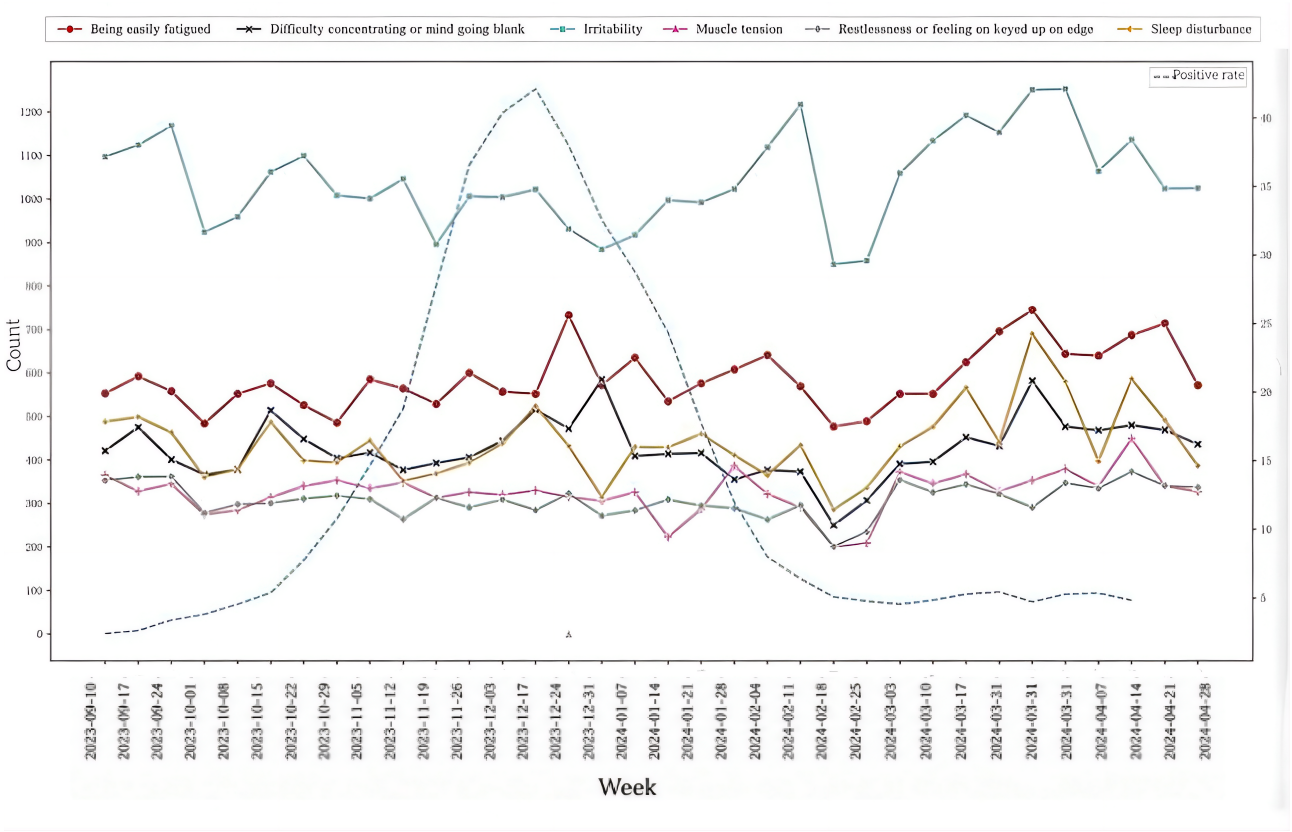
Results

RQ1: The Relationship Between Posting Volumes of Various Anxiety Symptoms and Influenza Infection Rates

Overview

Figure 1 shows that during the 2023 - 2024 influenza season, different anxiety symptoms exhibited distinct characteristics but overall maintained relatively moderate fluctuation patterns. Most symptoms remained at moderate levels from September 2023 to March 2024, with a downward trend observed in April 2024. Irritability and being easily fatigued were consistently the 2 most frequently mentioned symptoms, showing a fluctuating upward trend starting in late November and peaking at the end of January. Even after influenza positivity declined, these symptoms remained at relatively high levels. Sleep disturbance fluctuated repeatedly between November and March without forming a single peak, while difficulty concentrating or mind going blank showed a slight increase around December and another minor fluctuation in March. In contrast, muscle tension and restlessness or feeling keyed up or on edge remained at the lowest overall levels with limited variation.

Figure 1. The relationship between influenza positive rate and 6 kinds of symptoms posts.



Temporal Trend and Correlation Analysis

To further examine the above trends, we conducted a Spearman rank correlation analysis (Table 2). The results showed that only irritability (Spearman r=–0.413; P=.02) and restlessness or feeling keyed up or on edge (Spearman r=–0.396; P=.02) exhibited significant negative correlations with influenza positivity rates. Although the overall Spearman correlation between influenza positivity and irritability was negative (r=–.413; P=.02), irritability increased in parallel with influenza positivity during the outbreak and peak periods. However, irritability remained elevated even after influenza positivity declined, suggesting a lagged and persistent psychological response. This temporal mismatch may partially explain why the overall Spearman correlation coefficient appeared negative across the full study period. The other symptoms, while showing some local consistency with influenza fluctuations, did not reach statistical significance in their correlation coefficients.

XSL•FO
RenderX

Table . Results of Spearman rank correlation analysis between influenza positivity and anxiety symptom postings.

Symptom name	*r*	*P* value	95% CI
Irritability	–0.413	.02	–0.659 to –0.086
Difficulty concentrating or mind going blank	0.128	.47	–0.220 to 0.447
Restlessness or feeling keyed up or on edge	–0.396	.02	–0.647 to –0.066
Sleep disturbance	–0.221	.21	–0.520 to 0.127
Being easily fatigued	0.093	.60	–0.252 to 0.418
Muscle tension	–0.225	.20	–0.523 to 0.123

Robustness Analysis: ARIMAX Analysis of Temporal Associations

To further control for autocorrelation and potential lag effects within the time series, we additionally constructed ARIMAX models using influenza positivity rates as exogenous variables, incorporating contemporaneous (lag 0), 1-week lag (lag 1), and 2-week lag (lag 2) conditions separately. Compared with Spearman correlation analysis, ARIMAX models are better able to capture the dynamic temporal associations between influenza activity and anxiety-related symptom expressions. Detailed ARIMAX coefficient estimates are presented in Table 3, while full ARIMAX model specifications, including model orders (p, d, q), ADF stationarity test results, and diagnostic statistics, are provided in Multimedia Appendix 4.

Table . ARIMAX[a] results for anxiety-related symptom expressions under different lag conditions.

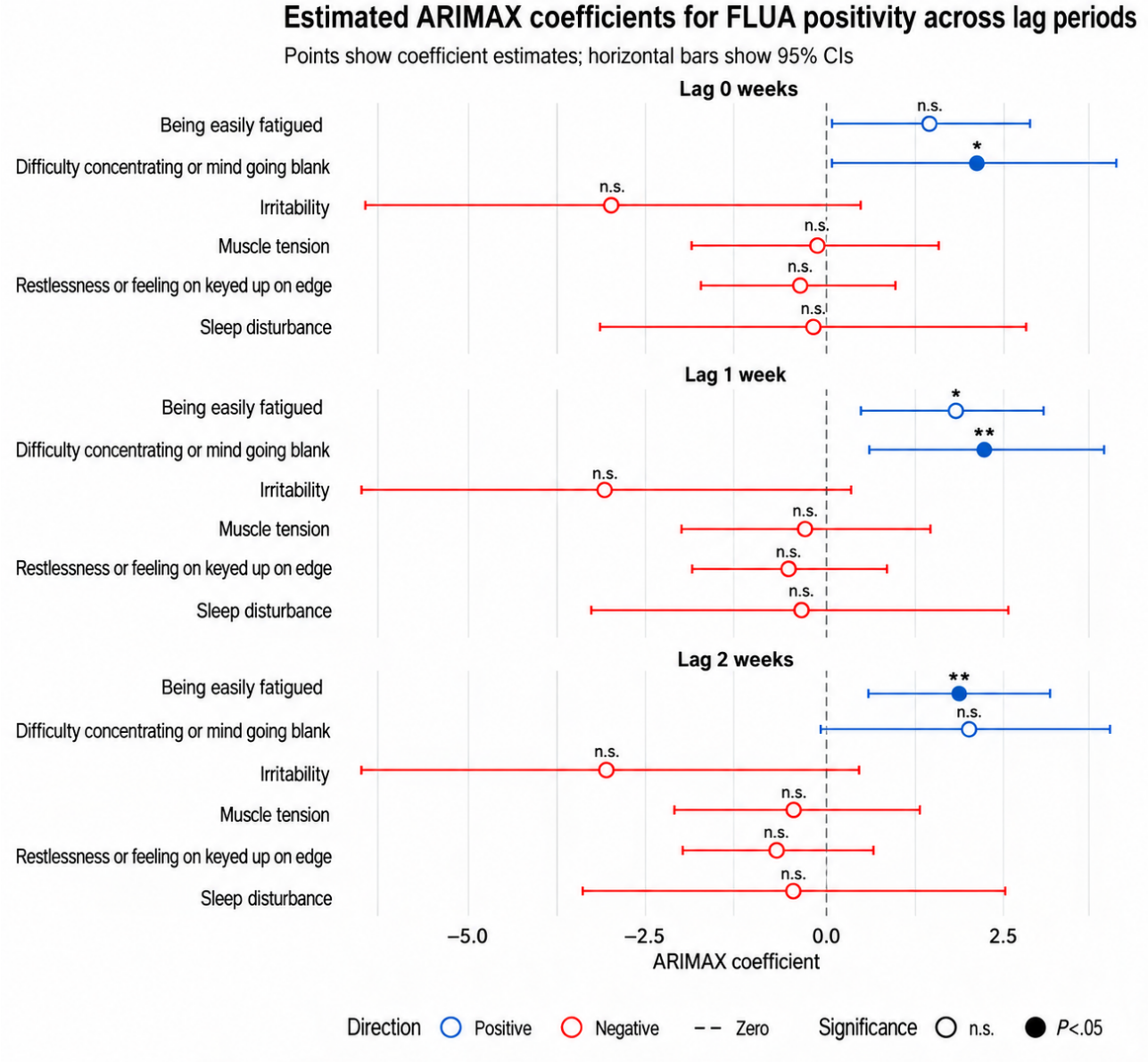
Symptom name and lag_model	Lag	β	*P* value	95% CI
Irritability				
lag0_concurrent	0	–3.0348	.08	–6.4734 to 0.4038
lag1_one_week	1	-3.1407	.10	–6.5525 to 0.2711
lag2_two_weeks	2	-3.0325	.08	–6.4384 to 0.3734
Difficulty concentrating or mind going blank				
lag0_concurrent	0	2.0798	.05	0.0308 to 4.1288
lag1_one_week	1	2.1684	.01	0.5218 to 3.815
lag2_two_weeks	2	1.9615	.06	–0.101 to 4.024
Restlessness or feeling keyed up or on edge				
lag0_concurrent	0	–0.412	.56	–1.7999 to 0.9759
lag1_one_week	1	–0.5406	.43	–1.8973 to 0.8161
lag2_two_weeks	2	–0.6861	.31	–2.0048 to 0.6326
Sleep disturbance				
lag0_concurrent	0	–0.2284	.88	–3.2556 to 2.7988
lag1_one_week	1	–0.3805	.80	–3.3695 to 2.6085
lag2_two_weeks	2	–0.469	.76	–3.4443 to 2.5063
Being easily fatigued				
lag0_concurrent	0	1.4068	.06	–0.0473 to 2.8609
lag1_one_week	1	1.7447	.01	0.3574 to 3.132
lag2_two_weeks	2	1.8699	.007	0.5157 to 3.2241
Muscle tension				
lag0_concurrent	0	–0.1324	.88	–1.8923 to 1.6275
lag1_one_week	1	–0.3161	.72	–2.0574 to 1.4252
lag2_two_weeks	2	–0.4388	.62	–2.1669 to 1.2893

[a]ARIMAX: autoregressive integrated moving average with exogenous regressors.

XSL•FO
RenderX

The results showed that, after controlling for temporal dependency structures, the association patterns between different anxiety symptoms and influenza positivity rates differed substantially (Figure 2). In the contemporaneous model (lag 0), influenza positivity rates were significantly positively associated with difficulty concentrating or mind going blank, whereas being easily fatigued showed a positive trend that did not reach statistical significance. The remaining symptoms, including irritability, sleep disturbance, muscle tension, and restlessness or feeling keyed up or on edge, generally did not exhibit significant associations.

Figure 2. ARIMAX coefficient estimates under different lag conditions. ARIMAX: autoregressive integrated moving average with exogenous regressors; FLUA: Influenza A virus; n.s.: not significant.



In the 1-week lag model (lag 1), influenza positivity rates were significantly positively associated with both being easily fatigued and difficulty concentrating or mind going blank, with the strength of association for difficulty concentrating or mind going blank becoming further enhanced. These findings suggest that certain anxiety-related expressions may not occur entirely synchronously with influenza activity, but instead may reflect delayed psychological or behavioral responses following peaks in influenza activity.

In the 2-week lag model (lag 2), being easily fatigued remained significantly positively associated with influenza positivity rates, whereas difficulty concentrating or mind going blank, although still showing a positive trend, no longer remained statistically significant. Meanwhile, the remaining symptoms did not demonstrate stable significant temporal associations across different lag conditions.

Overall, the ARIMAX findings differed somewhat from the Spearman correlation results. Spearman analyses indicated significant negative correlations between influenza positivity rates and both irritability and restlessness or feeling keyed up or on edge. However, after controlling for autocorrelation and lag effects, the ARIMAX models did not identify stable significant associations for these symptoms. In contrast, the ARIMAX analyses further revealed more stable positive

XSL•FO
RenderX

temporal dynamic associations between influenza activity and both being easily fatigued and difficulty concentrating or mind going blank, particularly within the lagged models.

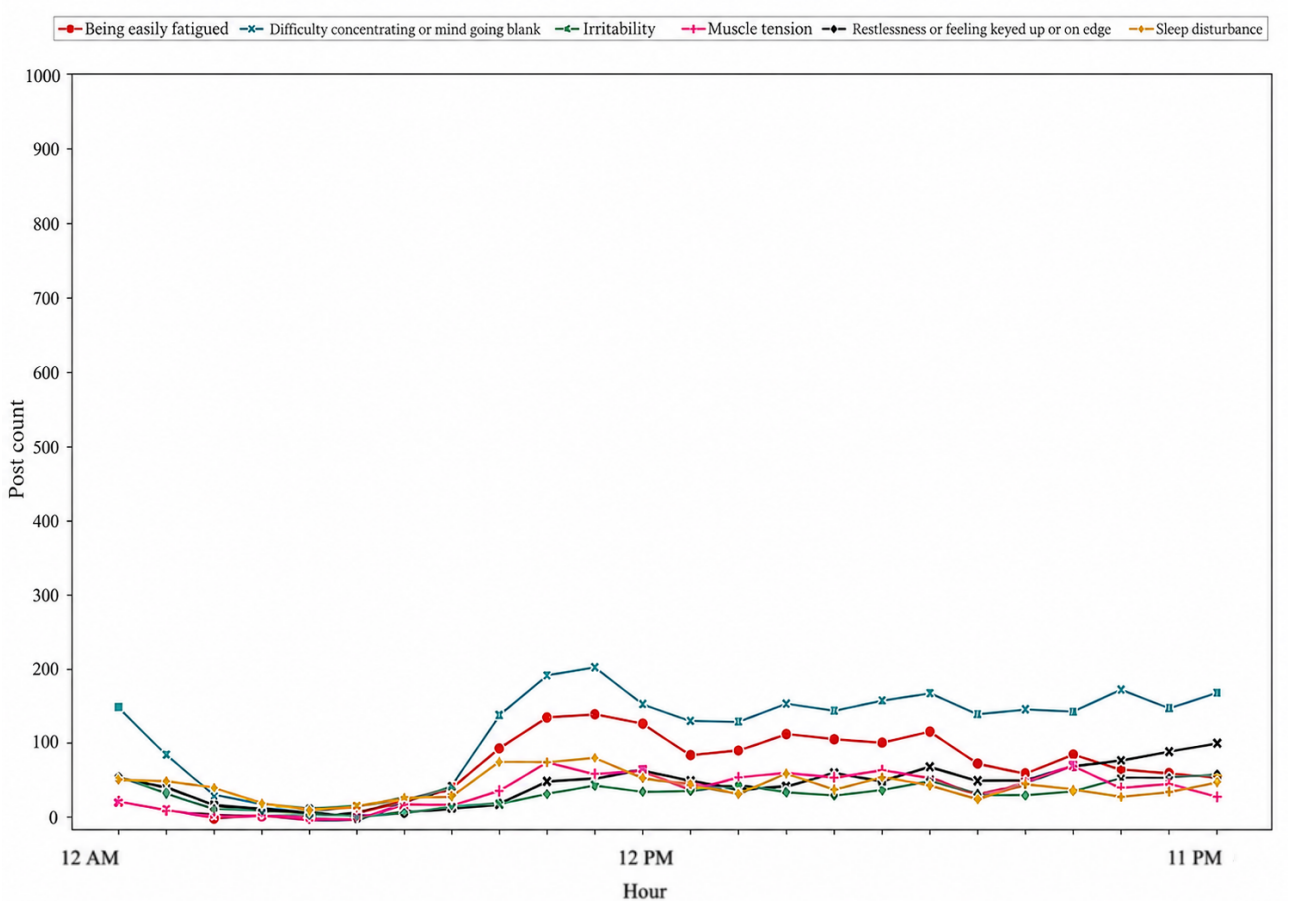
These findings suggest that simple correlation analyses may be influenced by overall temporal trends and stage-specific fluctuations, whereas ARIMAX models, after accounting for time-series dependency structures, are better able to capture the dynamic relationships between influenza activity and anxiety-related expressions. Although the 2 analytical approaches differed regarding which specific symptoms reached statistical significance, both overall supported the core conclusion of this study: anxiety-related expressions dynamically fluctuated alongside influenza activity during the influenza season and exhibited certain persistent and stage-dependent characteristics.

RQ2: Diurnal Variation Chart of Posting Volumes for Various Anxiety Symptoms

Overview

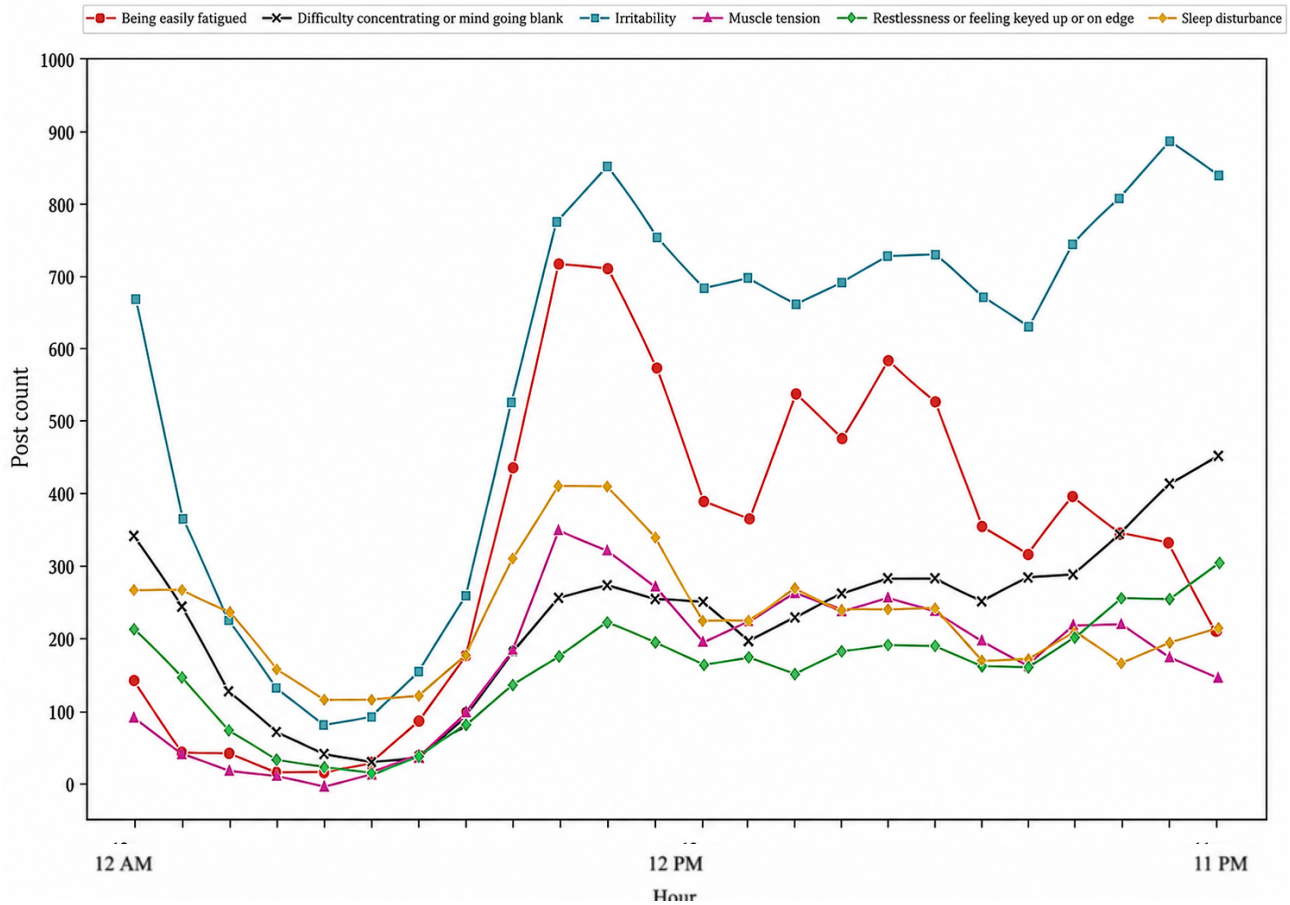
Figures 3-5 show the total number of linked posts for each phase in each 0 - to 24-hour period determined using Excel and Python.

Figure 3. Diurnal variation chart of posting volumes for various anxiety symptoms of the first stage.



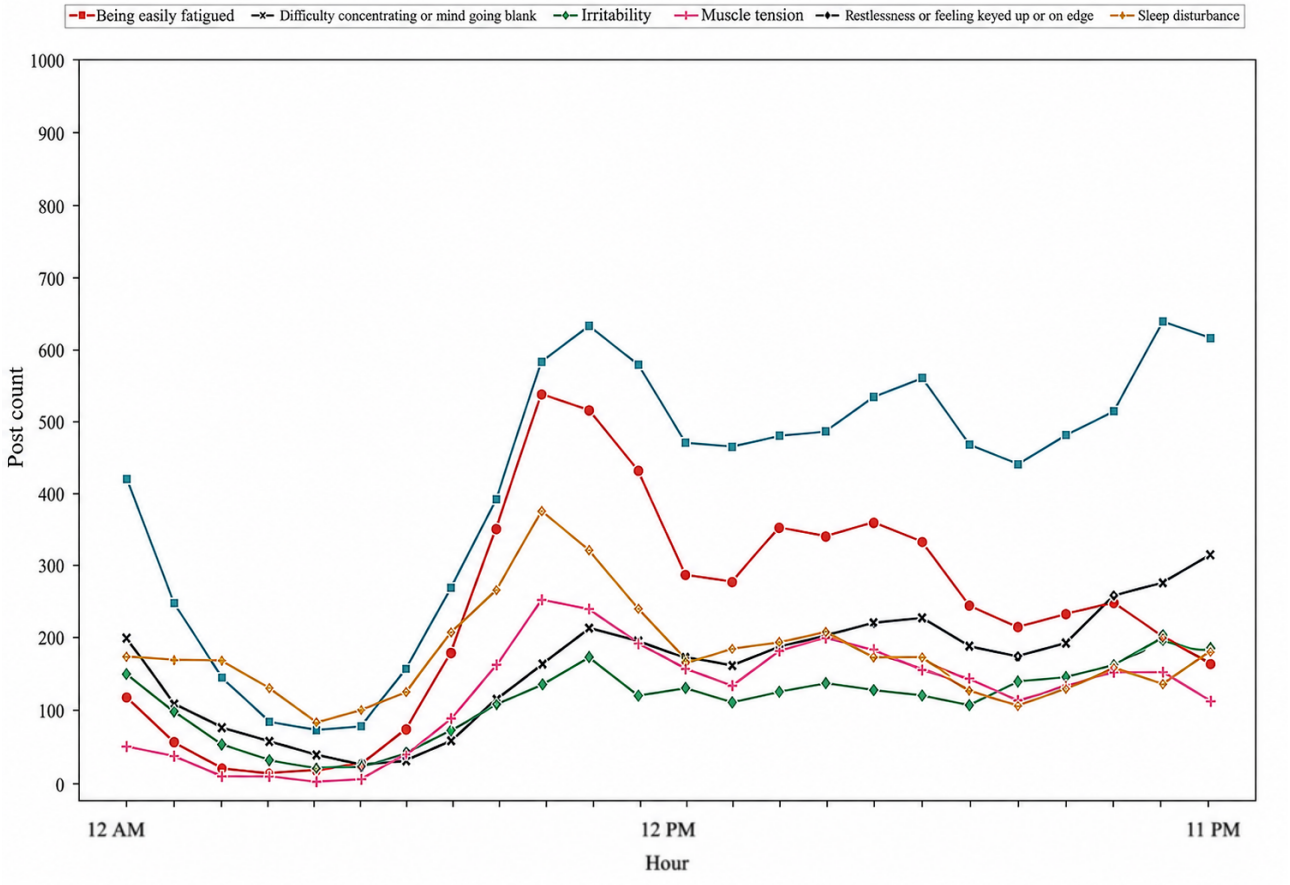
XSL•FO
RenderX

Figure 4. Diurnal variation chart of posting volumes for various anxiety symptoms of the second stage.



XSL•FO
RenderX

Figure 5. Diurnal variation chart of posting volumes for various anxiety symptoms of the third stage.



Phase I: The Influenza Outbreak Stage (November 6, 2023, to November 26, 2023)

In terms of diurnal distribution, irritability and being easily fatigued peaked in the morning (9-10 AM), while difficulty concentrating or mind going blank showed a slight increase during nighttime, and the overall fluctuations in symptom expression were relatively modest throughout the outbreak stage.

Phase II: The Influenza Prevalent and Declining Stage (November 27, 2023, to February 29, 2024)

During this stage, the diurnal distribution exhibited a clearer pattern. Symptom expression was most concentrated in the morning (9-11 AM), with relative secondary peaks in the afternoon (4-5 PM) and evening (9-11 PM). Among them, irritability and being easily fatigued remained relatively prominent across different time periods, whereas changes in other symptoms were comparatively limited.

Phase III: The End of Influenza Season (March 1, 2024, to April 28, 2024)

After the influenza season ended, the diurnal distribution became more simplified, with symptom expression mainly concentrated in the morning (9-11 AM) and afternoon (4-5 PM), and fewer symptoms expressed in the evening. Irritability and being easily fatigued remained at relatively higher levels within the daily distribution, but the gap with other symptoms diminished.

The diurnal curves in this study illustrated relative distribution patterns within each stage, rather than absolute posting levels across stages; therefore, overall platform activity was not used as a baseline. Across stages, irritability and being easily fatigued consistently emerged as the most prominent symptoms. Morning was generally the period with the highest concentration of symptom expression, while secondary peaks appeared in the afternoon or evening in certain stages, a common feature throughout the influenza season.

Robustness Analysis: Examination of Circadian Patterns Based on Within-Sample Hourly Normalization

To further examine whether the observed circadian patterns were primarily influenced by hourly fluctuations in posting activity within the anxiety-related dataset, we conducted an additional robustness analysis using within-sample hourly normalization. Specifically, within each influenza stage and corresponding hour, the number of posts for a given anxiety symptom category was divided by the total number of anxiety-related posts identified during the same hour, thereby generating the relative expression proportion of each symptom within that hour rather than using raw posting counts. This approach helps control, to some extent, for differences in posting volume across hours within the dataset and was used to assess whether the observed circadian patterns remained stable after normalization. The normalized circadian variation plots are provided in Multimedia Appendix 5.

The normalized results were generally consistent with the original count-based analyses, suggesting that the observed circadian patterns were not solely driven by hourly fluctuations in posting activity within the anxiety-related dataset. Across all

XSL•FO
RenderX

3 influenza stages, irritability and being easily fatigued remained the predominant symptom categories, with higher concentrations of expression still observed during daytime hours. Overall, the major temporal structures identified in the original analyses remained generally stable after normalization.

At the same time, the normalization analyses also revealed several localized differences. Compared with the raw-count results, sleep disturbance exhibited a more pronounced relative proportion during late-night and early-morning hours, particularly during the influenza outbreak and decline stages, suggesting that sleep-related expressions may become relatively more concentrated when overall posting activity is lower. In addition, some pronounced morning peaks observed in the original analyses became attenuated after normalization and instead appeared as broader daytime plateau patterns. This suggests that certain morning peaks may have been partially influenced by routine platform activity cycles, although the overall circadian temporal structure remained largely unchanged.

During the epidemic and decline stages, the normalized results continued to show relatively high proportions of irritability during daytime and evening hours, while being easily fatigued remained consistently elevated throughout daytime periods. Meanwhile, difficulty concentrating or mind going blank demonstrated a relatively increased proportion during evening hours, whereas the remaining symptoms generally maintained relatively stable temporal distribution patterns.

Overall, the robustness analyses largely supported the stability of the original findings, indicating that the observed circadian variation patterns were not simply driven by differences in posting activity within the dataset. At the same time, the normalization analyses further revealed more subtle differences in the relative expression of specific symptoms across particular time periods.

However, because hourly posting-volume data for the overall Sina Weibo platform were unavailable, strict platform-level normalization could not be performed. Therefore, the potential confounding influence of broader platform activity cycles cannot be completely excluded. Accordingly, the findings of this study should be interpreted as descriptive temporal patterns of anxiety-related expressions rather than prevalence estimates at the platform-wide population level.

RQ3: The Evolution Trends of Various Anxiety Symptoms

The core objective of the ARIMA model is to extract the underlying correlation structure of the time series so that the residuals approximate white noise. Stationarity was assessed using the ADF test prior to model fitting. Residual autocorrelation analysis showed that the mean residuals of all symptom models were close to 0. Residual plots, autocorrelation function or partial autocorrelation function plots, and quantile-quantile plots (Multimedia Appendix 6) further confirmed that the residuals showed no systematic trends or outliers and were approximately normally distributed, thereby supporting the adequacy of the model fit. Meanwhile, Table 4 presents the predictive performance of the ARIMA time-series model, which was used to evaluate out-of-sample forecasting performance and model generalizability. Based on the chronological 80% training and 20% hold-out testing split, the out-of-sample prediction errors were within an acceptable range, suggesting that the ARIMA models demonstrated generally adequate short-term forecasting performance across symptom categories.

XSL•FO
RenderX

Table . Out-of-sample forecasting performance of ARIMA[a] models across anxiety symptom categories[b].

Symptom name	Model order(p,d,q)[c]	RMSE[d]	MAE[e]	MAPE[f]	MASE[g]	Ljung_Box, chi-square (*df*=10)	*P* value	White_noise
Irritability	ARIMA(2,0,0)(1,0,0)[7]	26.57	20.39	11.81	1.02	9.7	.47	Yes
Difficulty concentrating or mind going blank	ARIMA(0,1,2)(0,0,1)[7]	17.07	11.85	15.28	1.12	6.2	.80	Yes
Restlessness or feeling keyed up or on edge	ARIMA(1,1,1)(0,0,1)[7]	7.89	6.6	14.59	0.8	8,7	.56	Yes
Sleep disturbance	ARIMA(1,0,0)(0,0,2)[7]	28.38	17.54	21.79	1.38	4,4	.93	Yes
Being easily fatigued	ARIMA(2,0,2)(1,0,0)[7]	28.02	20.21	19.56	1.32	6,0	.81	Yes
Muscle tension	ARIMA(2,0,0)(2,0,0)[7]	14.5	10.14	17.86	0.95	7.1	.72	Yes

[a]ARIMA: autoregressive integrated moving average.

[b]RMSE, MAE, MAPE, and MASE were calculated exclusively on the unseen test set to evaluate out-of-sample forecasting performance and model generalizability.

[c]Models with *d*=0 included an intercept (nonzero mean).

[d]RMSE: root mean square error.

[e]MAE: mean absolute error.

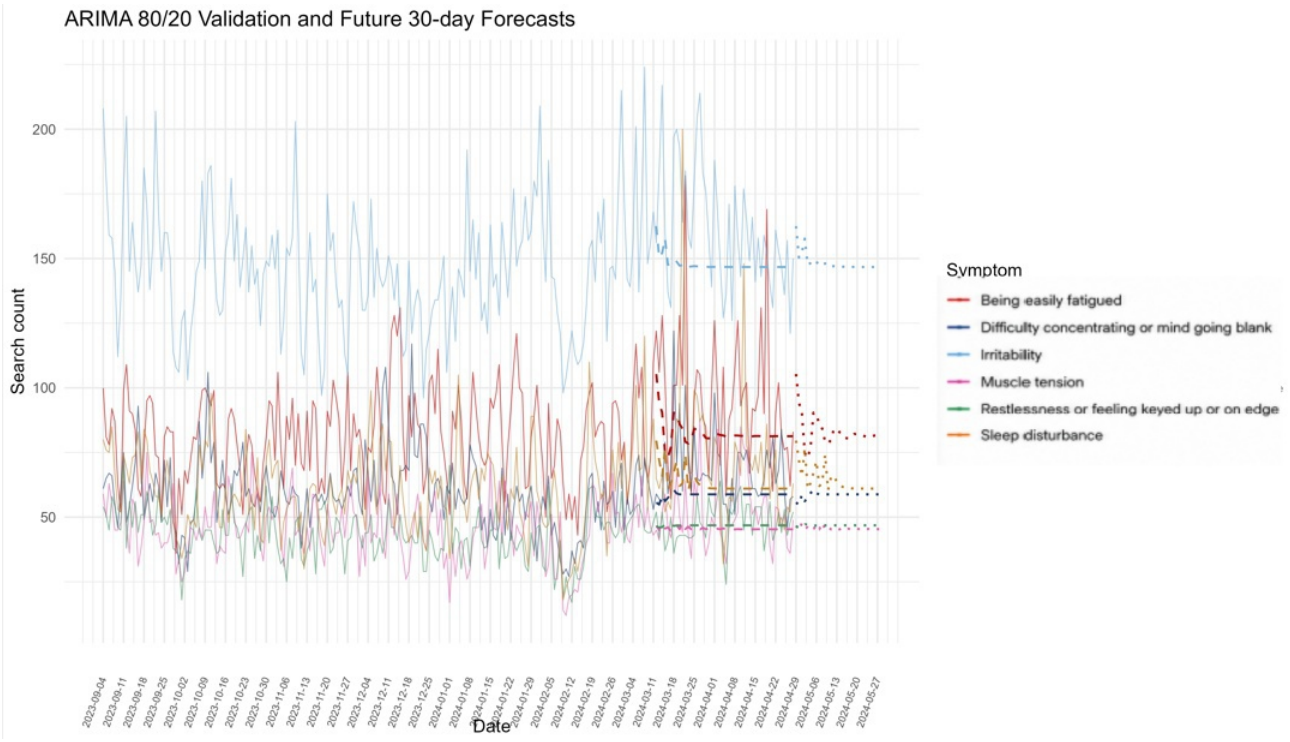
[f]MAPE: mean absolute percentage error.

[g]MASE: mean absolute scaled error.

As shown in Figure 6, discussions related to irritability and being easily fatigued were the most prevalent in the historical data, while other symptoms (such as sleep disturbance, difficulty concentrating or mind going blank, restlessness or feeling keyed up or on edge, and muscle tension) appeared at relatively low levels. Out-of-sample forecasting results based on the hold-out test set generally captured the major temporal patterns and fluctuations observed in the anxiety-related symptom series, thereby supporting the robustness of the short-term forecasts. The 30-day forecasting analysis further suggested that irritability would remain at its current level; being easily fatigued was generally stable but showed a slight increase; sleep disturbance and difficulty concentrating or mind going blank remained overall stable with only minor upward fluctuations; restlessness or feeling keyed up or on edge might show a slight decrease; and muscle tension was expected to remain relatively stable.

XSL•FO
RenderX

Figure 6. ARIMA forecasting of anxiety-related symptom expressions using an 80/20 train-test split. ARIMA: autoregressive integrated moving average.



Discussion

Principal Findings

This study identified dynamic temporal patterns of anxiety-related expressions during the 2023 - 2024 influenza season. Across analyses, anxiety-related expressions fluctuated alongside influenza activity and exhibited stage-dependent diurnal variation patterns. Supplementary ARIMAX and normalization analyses further suggested that several temporal associations and circadian structures remained relatively stable after accounting for lagged effects and within-sample posting fluctuations. In addition, ARIMA-based forecasting indicated generally stable short-term trajectories across symptom categories, with being easily fatigued remaining one of the most prominent symptoms over time.

Although keyword-based approaches may involve individual-level misclassification, the converging temporal patterns observed across correlation, time-series, and normalization analyses suggest that the extracted linguistic markers may capture meaningful population-level fluctuations in anxiety-related expressions rather than purely random noise. Nevertheless, these findings should be interpreted as descriptive temporal patterns of anxiety-related expressions within a public health context rather than evidence of causal psychological effects.

Theoretical Significance

Linguistic Markers on Social Media Linking Anxiety Symptoms With Influenza Incidence

This study explores the dynamic characteristics of social media in tracking public mental health. The results (step 1) suggest that anxiety-related expressions on social media exhibited dynamic temporal associations with influenza activity during the 2023 - 2024 influenza season. Initial Spearman correlation analyses showed significant negative correlations for irritability and restlessness or feeling keyed up or on edge, whereas other symptoms demonstrated nonsignificant trends. However, supplementary ARIMAX analyses further revealed that, after controlling for temporal autocorrelation and lagged effects, these negative associations were no longer stable. Instead, being easily fatigued and difficulty concentrating or mind going blank exhibited more consistent positive temporal associations with influenza activity, particularly under lagged conditions. This discrepancy suggests that simple correlation analyses may not fully capture the temporal complexity of anxiety-related expressions during influenza seasons, whereas time-series approaches are better able to identify delayed and dynamic anxiety-related expressions associated with influenza activity.

In addition, anxiety-related posts remained elevated until April 2024, when they began to decline as temperatures rose and the influenza season came to an end, suggesting that anxiety-related expressions during influenza seasons may not fluctuate in complete synchrony with influenza positivity rates and may also be influenced by daily routines, seasonal variations, and broader social adjustment processes. These results demonstrate that social media tracking not only reflects the immediate anxiety-related expressions to influenza but also captures the potential persistence of anxiety symptoms during the recovery phase.

Further examination of the temporal dynamics reveals local synchronization: during certain weeks, particularly around the peak influenza period, anxiety-related postings fluctuated in parallel with influenza activity. Between late November 2023 and the end of January 2024, the physical and psychological impacts of influenza were accompanied by an overall increase

XSL•FO
RenderX

in anxiety-related postings, mainly reflected in irritability and being easily fatigued. The lagged associations identified in the ARIMAX models further suggest that certain anxiety-related expressions may emerge or persist after increases in influenza activity, indicating delayed and sustained psychological or behavioral responses rather than purely concurrent reactions.

Overall, the findings suggest that mental health tracking on social media is not static but dynamically adjusts to changes in public health conditions. During influenza outbreaks, public attention to health issues tended to increase, as did the online expression of anxiety symptoms [3]. Through social media, individuals promptly shared health experiences, expressed anxiety, and interacted with others, forming a collective coping mechanism [3]. Such dynamic participation supports individual mental well-being and contributes to a broader emotional support network at the societal level.

Linguistic Markers of Anxiety Symptoms' Diurnal Patterns in Social Media

The research results (step 2) show that the diurnal expression patterns of anxiety symptoms exhibited dynamic changes across 3 phases of the influenza season—from the insignificant diurnal differences observed in the initial outbreak phase, to a concentration in the morning with smaller secondary peaks in the afternoon and evening during the prevalent and declining phase, and finally to a primary morning peak with a smaller secondary rise in the afternoon during the ending phase. This dynamic change indicates that social media tracking can sensitively capture subtle temporal differences in the expression of public anxiety symptoms. Supplementary within-sample hourly normalization analyses further demonstrated that the major diurnal structures remained generally stable after normalization, suggesting that the observed temporal patterns were unlikely to be explained entirely by hourly fluctuations in posting activity within the anxiety-related dataset.

As a real-time platform for the public to express emotions and health status, the linguistic marker nature of social media data enables us to precisely capture the different peak times of anxiety symptoms throughout the day [23,24]. For example, the morning peak of anxiety may be related to the stress of starting a new day, the afternoon secondary peak may reflect accumulated workload and social interaction–related tension, and the evening secondary peak may be associated with psychological fatigue and reduced self-regulation capacity after daily activities [33]. Although several peak intensities became attenuated after normalization, the overall morning and daytime concentration patterns persisted across influenza stages, further supporting the temporal stability of the observed diurnal structures. This dynamic change in diurnal rhythm shows that social media is a window reflecting individual biological rhythms and social activity patterns [34]. On the one hand, individual physiological rhythms (such as the sleep-wake cycle) may affect the frequency and timing of expressing anxiety on social media; on the other hand, the patterns of social activities (such as the timing of work, study, and social interactions) also influence individuals' attention to and enthusiasm for discussing health issues. This dynamic change in diurnal rhythm further enriches the connotation of social media in mental health tracking, revealing behavioral characteristics and psychological needs at different times of the day [35].

Differential Evolutionary Trends of Anxiety Symptoms as Linguistic Markers on Social Media

This study found that, in both historical data and out-of-sample forecasting analyses based on a chronological 80% training and 20% testing split, irritability and being easily fatigued consistently dominated the discussions, underscoring the central role of emotional swings and physical fatigue in public anxiety experiences. The fitted results closely matched the original data, indicating that the time-series approach reliably captured historical trends. Based on the 30-day forecast, the overall trajectories are expected to remain stable, with being easily fatigued showing a slight increase. Although this change is modest, its persistent prominence across the historical period suggests that physical fatigue exerts a considerable impact on public experiences of anxiety during influenza seasons.

Therefore, as a widely prevalent disease with substantial physiological consequences, influenza not only imposes direct physical burdens but is also closely linked to public mental health [36-38]. These findings indicate that a dynamic tracking approach combining out-of-sample validation and short-term forecasting can effectively characterize the temporal evolution of anxiety symptoms during influenza seasons and offer methodological support for understanding their psychological and behavioral mechanisms as well as optimizing public mental health responses.

Practical Implications

This study provides several practical implications for public health and mental health practice. First, the observed associations between influenza incidence and anxiety [37]-related discussions suggest that social media monitoring can serve as a descriptive observational resource of population mental health during public health events. During flu season, public mental health education and assistance reduce anxiety [3].

Second, the identified diurnal variations highlight the importance of time-sensitive strategies. Mental health campaigns and support services delivered during peak anxiety periods—particularly in the morning, with additional rises in the afternoon and evening—may enhance engagement and effectiveness [39].

Finally, the findings emphasize the need for sustained mental health monitoring during and after the influenza season to address delayed or lingering anxiety responses and promote long-term resilience [3,40].

Limitations and Future Directions

This study has several limitations.

First, the use of Sina Weibo data may introduce sampling bias [4,41], limiting the generalizability of the findings to non-Weibo users [7].

Second, the keyword-based approach may introduce misclassification bias, as selected terms (eg, fatigue or sleep disturbance) may appear in nonanxiety contexts [32]. In

XSL•FO
RenderX

addition, colloquial, metaphorical, or culturally specific expressions of anxiety (eg, slang terms) may not be fully captured. Future studies should incorporate context-sensitive AI methods [42], such as supervised classifiers or semantic embedding models, to improve measurement validity [43,44].

In addition, the keyword-based approach did not explicitly account for contextual linguistic phenomena such as negation, sarcasm, or irony. Therefore, posts containing anxiety-related keywords in nonanxious contexts may still have been included, which may affect classification accuracy. Accordingly, the findings should be interpreted as descriptive temporal patterns rather than direct measures of individual psychological states or clinical anxiety prevalence. Future studies should incorporate advanced context-sensitive natural language processing approaches, such as transformer-based models (eg, Bidirectional Encoder Representations from Transformers), to improve semantic understanding and classification accuracy.

Third, certain expressions such as "fatigue" may reflect multiple overlapping conditions or experiences, including work-related exhaustion, physical illness, sleep deprivation, or emotional distress, rather than anxiety specifically. This semantic ambiguity highlights the difficulty of inferring discrete psychological states from keyword-based social media data. Future studies should further incorporate contextual semantic analysis to better distinguish different sources and meanings of similar expressions.

Furthermore, given the large data volume, manual verification of each post was not feasible, and some irrelevant content may remain. Although some degree of misclassification may still exist, there is no clear reason to assume that these errors would systematically align with the temporal structures examined in this study (eg, influenza stages or diurnal cycles). Therefore, while potential bias cannot be completely excluded, the major temporal patterns observed are unlikely to be solely driven by preprocessing decisions.

In addition, this study did not control for potentially important contextual factors, such as holidays, academic schedules, media coverage, weather changes, or broader social events, which may also shape temporal fluctuations in anxiety-related expressions. Therefore, the findings should be interpreted as descriptive associations rather than evidence of causal inference.

Meanwhile, future research should address these issues by integrating multiplatform or survey data to reduce sampling bias, applying advanced natural language processing to capture informal expressions, and expanding the lexical framework beyond *DSM-IV* terms to reflect cultural nuances.

Finally, although supplementary within-sample hourly normalization analyses were conducted, the study was unable to perform full platform-level normalization because hourly posting volume data for the entire Weibo platform were unavailable. Therefore, residual confounding associated with broader platform activity patterns cannot be completely excluded. Future studies incorporating platform-wide posting volume as a baseline may help better distinguish true diurnal variations in anxiety-related expressions from general social media use rhythms.

Conclusions

This study used social media data during the influenza season to examine diurnal patterns, dynamic fluctuations, and associations with influenza activity in public anxiety symptoms. The findings revealed influenza-related shifts in symptom expressions and differences across symptom categories, as reflected on social media.

Overall, the ubiquity of social media use provides valuable linguistic markers of affective symptoms [45], highlighting its potential value as a descriptive observational resource for understanding population mental health dynamics during infectious disease seasons [4].

Acknowledgments

The authors used Chatgpt to help check for grammatical and lexical issues while writing the manuscript. No AI tools were used for data analysis, interpretation of results, or generation of scientific conclusions.

Funding

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

Data Availability

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

Authors' Contributions

YO was responsible for the research design and manuscript writing. YO collected the data and performed the analysis. GJB and PJS supervised the project, and GJB is the first supervisor for the project. GJB supervised and modified the research proposal and provided guidance and suggestions on data collection and data analysis during the pretesting and formal research stages. GJB and PJS advised on the manuscript outline, gave feedback, and made manuscript's revisions. GJB made meticulous revisions and proofread the manuscript from the initial draft to the final version.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary pilot analysis of synonymous anxiety-related expressions on Sina Weibo.
[ZIP File, 25800 KB - infodemiology_v6i1e88849_app1.zip]

Multimedia Appendix 2
Detailed data preprocessing steps.
[DOC File, 17 KB - infodemiology_v6i1e88849_app2.doc]

Multimedia Appendix 3
Positive rate of influenza A and percentage of influenza A among all positive cases.
[XLSX File, 10 KB - infodemiology_v6i1e88849_app3.xlsx]

Multimedia Appendix 4
Robustness analysis: autoregressive integrated moving average with exogenous regressors analysis of temporal associations.
[ZIP File, 3459 KB - infodemiology_v6i1e88849_app4.zip]

Multimedia Appendix 5
Robustness analysis: examination of circadian patterns based on within-sample hourly normalization.
[ZIP File, 3205 KB - infodemiology_v6i1e88849_app5.zip]

Multimedia Appendix 6
Details: residual plots, autocorrelation function or partial autocorrelation function plots, and quantile-quantile plots, and stationarity test of autoregressive integrated moving average.
[ZIP File, 1550 KB - infodemiology_v6i1e88849_app6.zip]

References

1. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
2. Oluwalade TI, Ahmadi H, Huo L, Sharpe R, Zhou SM. Comorbidities and emotions—unpacking the sentiments of pediatric patients with multiple long-term conditions through social media feedback: a large language model-driven study. J Affect Disord 2025 Nov 1;388:119752. [doi: 10.1016/j.jad.2025.119752] [Medline: 40550274]
3. Zheng P, Adams PC, Wang J. Shifting moods on Sina Weibo: the first 12 weeks of COVID-19 in Wuhan. New Media Soc 2024 Jan;26(1):346-367. [doi: 10.1177/14614448211058850]
4. Wang R, Ji Z, Chen Y, Wu H, Liang X. Identifying possible influence factors for depression through social media. J Affect Disord 2025 Sep 15;385:119385. [doi: 10.1016/j.jad.2025.119385] [Medline: 40374092]
5. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. ICWSM 2021;7(1):128-137. [doi: 10.1609/icwsm.v7i1.14432]
6. Jung H, Park HA, Song TM. Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. J Med Internet Res 2017 Jul 24;19(7):e259. [doi: 10.2196/jmir.7452] [Medline: 28739560]
7. Yu L, Jiang W, Ren Z, Xu S, Zhang L, Hu X. Detecting changes in attitudes toward depression on Chinese social media: a text analysis. J Affect Disord 2021 Feb 1;280(Pt A):354-363. [doi: 10.1016/j.jad.2020.11.040] [Medline: 33221722]
8. Zhu J, Li Z, Zhang X, Zhang Z, Hu B. Public attitudes toward anxiety disorder on Sina Weibo: content analysis. J Med Internet Res 2023 Apr 4;25:e45777. [doi: 10.2196/45777] [Medline: 37014691]
9. Hofer M, Hargittai E. Online social engagement, depression, and anxiety among older adults. New Media Soc 2024 Jan;26(1):113-130. [doi: 10.1177/14614448211054377]
10. Skaik R, Inkpen D. Using social media for mental health surveillance. ACM Comput Surv 2021 Nov 30;53(6):1-31. [doi: 10.1145/3422824]
11. Boursier V, Gioia F, Musetti A, Schimmenti A. Facing loneliness and anxiety during the COVID-19 isolation: the role of excessive social media use in a sample of Italian adults. Front Psychiatry 2020;11:586222. [doi: 10.3389/fpsyt.2020.586222] [Medline: 33363484]
12. Hung LP, Tadius EM. Journaling system with embedded machine learning text depression detection alert. Presented at: 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET); Sep 13-15, 2022; Kota Kinabalu, Malaysia p. 1-6. [doi: 10.1109/IICAIET55139.2022.9936789]

XSL•FO
RenderX

13. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176. [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
14. Wu S, Yao M, Deng C, Marsiglia FF, Duan W. Social isolation and anxiety disorder during the COVID-19 pandemic and lockdown in China. J Affect Disord 2021 Nov 1;294:10-16. [doi: 10.1016/j.jad.2021.06.067] [Medline: 34256180]
15. Liu J, Tai Z, Hu F. Prevalence and coping of depression and anxiety among college students during COVID-19 lockdowns in China. J Affect Disord 2024 Mar;348:305-313. [doi: 10.1016/j.jad.2023.12.061]
16. Feng L, Zhang T, Wang Q, et al. Impact of COVID-19 outbreaks and interventions on influenza in China and the United States. Nat Commun 2021 May 31;12(1):3249. [doi: 10.1038/s41467-021-23440-1] [Medline: 34059675]
17. Wheaton MG, Abramowitz JS, Berman NC, Fabricant LE, Olatunji BO. Psychological predictors of anxiety in response to the H1N1 (swine flu) pandemic. Cogn Ther Res 2012 Jun;36(3):210-218. [doi: 10.1007/s10608-011-9353-3]
18. Ou Y, de Bruijn GJ, Schulz PJ. Social media as an emotional barometer: bidirectional encoder representations from transformers-long short-term memory sentiment analysis on the evolution of public sentiments during influenza A on Sina Weibo. J Med Internet Res 2025 Sep 3;27:e68205. [doi: 10.2196/68205] [Medline: 40900625]
19. Wang G, Sabran K. Assessing depression and anxiety among young adults after epidemics and pandemics: a cross-sectional study in Anyang, China. Sci Rep 2024;14(1):2759. [doi: 10.1038/s41598-024-53292-w]
20. Bults M, Beaujean DJ, de Zwart O, et al. Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: results of three consecutive online surveys. BMC Public Health 2011 Jan 3;11:2. [doi: 10.1186/1471-2458-11-2] [Medline: 21199571]
21. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. Health Commun 2012;27(2):179-185. [doi: 10.1080/10410236.2011.571759] [Medline: 21827326]
22. Boiko DI, Skrypnikov AM, Shkodina AD, Hasan MM, Ashraf GM, Rahman MH. Circadian rhythm disorder and anxiety as mental health complications in post-COVID-19. Environ Sci Pollut Res Int 2022 Apr;29(19):28062-28069 Retracted inEnviron Sci Pollut Res Int. 2025;32(57):30970. [doi: 10.1007/s11356-025-37292-5] [Medline: 34988815]
23. Tian X, Yu G, He F. An analysis of sleep complaints on Sina Weibo. Comput Human Behav 2016 Sep;62:230-235. [doi: 10.1016/j.chb.2016.04.014]
24. Tian X, He F, Batterham P, Wang Z, Yu G. An analysis of anxiety-related postings on Sina Weibo. Int J Environ Res Public Health 2017 Jul 13;14(7):775. [doi: 10.3390/ijerph14070775] [Medline: 28703773]
25. Cox RC, Wright KP Jr, Axelsson J, Balter LJT. Diurnal variation in anxiety and activity is influenced by chronotype and probable anxiety-related disorder status. Psychiatry Res 2024 Aug;338:116006. [doi: 10.1016/j.psychres.2024.116006] [Medline: 38850890]
26. Zou H, Bao C, Wang X, et al. Depressive symptoms moderate diurnal variation in response-inhibition-related neural activity among patients with depression: an EEG study. J Affect Disord 2025 Nov 1;388:119617. [doi: 10.1016/j.jad.2025.119617] [Medline: 40484183]
27. Taylor BJ, Hasler BP. Chronotype and mental health: recent advances. Curr Psychiatry Rep 2018 Jul 23;20(8):59. [doi: 10.1007/s11920-018-0925-8] [Medline: 30039327]
28. Ten Thij M, Bathina K, Rutter LA, et al. Depression alters the circadian pattern of online activity. Sci Rep 2020 Oct 14;10(1):17272. [doi: 10.1038/s41598-020-74314-3] [Medline: 33057099]
29. Substance Abuse and Mental Health Services Administration. Impact of the DSM-IV to DSM-5 Changes on the National Survey on Drug Use and Health: Substance Abuse and Mental Health Services Administration; 2016. URL: https://www.ncbi.nlm.nih.gov/books/NBK519697/ [accessed 2025-09-03]
30. Weekly report. Chinese National Influenza Center. 2025. URL: https://ivdc.chinacdc.cn/cnic/en/Surveillance/WeeklyReport/ [accessed 2025-09-08]
31. Cao Y, Dai J, Wang Z, et al. Machine learning approaches for depression detection on social media: a systematic review of biases and methodological challenges. JBDS 2025;5(1):67-102. [doi: 10.35566/jbds/caoyc]
32. Ali G, Mohd-Yasin F. ARIMA modeling of noise of piezoelectric accelerometer. IEEE Access 2025;13:112121-112131. [doi: 10.1109/ACCESS.2025.3581427]
33. Sabet SM, Dautovich ND, Dzierzewski JM. The rhythm is gonna get you: social rhythms, sleep, depressive, and anxiety symptoms. J Affect Disord 2021 May 1;286:197-203. [doi: 10.1016/j.jad.2021.02.061] [Medline: 33735764]
34. Gleasure R, Saigot M, Kanat I. Let's talk about it in the morning: how circadian rhythms impact information sharing on social media. Affect Sci 2024 Dec;5(4):389-404. [doi: 10.1007/s42761-024-00254-0] [Medline: 39649455]
35. Gleasure R. Circadian rhythms and social media information-sharing. In: Davis FD, Riedl R, Brocke J, Léger PM, Randolph A, Fischer T, editors. Information Systems and Neuroscience: NeuroIS Retreat 2019 Lecture Notes in Information Systems and Organisation: Springer; 2020:1-11. [doi: 10.1007/978-3-030-28144-1_1]
36. Li Y, Luan S, Li Y, Hertwig R. Changing emotions in the COVID-19 pandemic: a four-wave longitudinal study in the United States and China. Soc Sci Med 2021 Sep;285:114222. [doi: 10.1016/j.socscimed.2021.114222] [Medline: 34418734]
37. Li J, Luo C, Liu L, et al. Depression, anxiety, and insomnia symptoms among Chinese college students: a network analysis across pandemic stages. J Affect Disord 2024 Jul;356:54-63. [doi: 10.1016/j.jad.2024.04.023]

XSL•FO
RenderX

38. Metzler H, Rimé B, Pellert M, Niederkrotenthaler T, Di Natale A, Garcia D. Collective emotions during the COVID-19 outbreak. Emotion 2023 Apr;23(3):844-858. [doi: 10.1037/emo0001111] [Medline: 35787108]
39. Ahmed A, Aziz S, Toro CT, et al. Machine learning models to detect anxiety and depression through social media: a scoping review. Comput Methods Programs Biomed Update 2022;2:100066. [doi: 10.1016/j.cmpbup.2022.100066] [Medline: 36105318]
40. Eichstaedt JC, Schwartz HA, Kern ML, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169. [doi: 10.1177/0956797614557867] [Medline: 25605707]
41. Zhao Y, Cheng S, Yu X, Xu H. Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. J Med Internet Res 2020 May 4;22(5):e18825. [doi: 10.2196/18825] [Medline: 32314976]
42. Li X, Qing L, Wei H, Wei J. Artificial intelligence and innovation ambidexterity in advanced manufacturing firms: a knowledge-management perspective. J Manuf Technol Manag 2026:1-26. [doi: 10.1108/JMTM-06-2025-0489]
43. Ou Y, Zhang P, Yu J, et al. The application of the BERTopic model in natural language processing: in-depth text topic modeling. Presented at: 2025 5th International Conference on Consumer Electronics and Computer Engineering (ICCECE); Feb 28 to Mar 2, 2025. [doi: 10.1109/ICCECE65250.2025.10984639]
44. Ou Y, Sun F, Tan C, Shi H, Zhang M, Gong C. Analyzing user behavior in online communities using data crawling and machine learning algorithms (erratum). Presented at: Fifth International Conference on Telecommunications, Optics, and Computer Science; Dec 5-7, 2024. [doi: 10.1117/12.3072604]
45. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. J Affect Disord 2017 Jan 1;207:163-166. [doi: 10.1016/j.jad.2016.08.040] [Medline: 27723539]

Abbreviations

ADF: augmented Dickey-Fuller
ARIMA: autoregressive integrated moving average
ARIMAX: autoregressive integrated moving average with exogenous regressors
***DSM-IV*:** *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition

Edited by T Mackey; submitted 14.Jan.2026; peer-reviewed by A Samraj, KKL Lau, Y Cao; revised version received 22.Jun.2026; accepted 23.Jun.2026; published 07.Aug.2026.

Please cite as:
Ou Y, Bruijn GJD, Schulz PJ
Temporal Dynamics of Influenza-Associated Anxiety Symptom Linguistic Markers on Weibo (2023-2024): Observational Study
JMIR Infodemiology 2026;6:e88849
URL: https://infodemiology.jmir.org/2026/1/e88849
doi:10.2196/88849

© Yifan Ou, Gert-Jan de Bruijn, Peter Johannes Schulz. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 7.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Self-Reported Tianeptine Experiences on Reddit: Natural Language Processing–Assisted Qualitative Study

Christopher J Counts[1], MD, MSc; Anthony V Spadaro[1], MD; Sahithi Lakamana[2], MS; Abeed Sarker[2], PhD; Rachel Wightman[3], MD; Jennifer Love[4], MD; Diane Calello[1], MD; Jeanmarie Perrone[5], MD

[1]Department of Emergency Medicine, Rutgers New Jersey Medical School, 140 Bergen St, Newark, NJ, United States

[2]Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, United States

[3]Department of Emergency Medicine, Brown University Health, Providence, RI, United States

[4]Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[5]Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Christopher J Counts, MD, MSc
Department of Emergency Medicine, Rutgers New Jersey Medical School, 140 Bergen St, Newark, NJ, United States

Abstract

This study, using natural language processing and manual thematic analysis of Reddit posts, revealed a rapid rise in discussions about tianeptine, with posts frequently reporting dependence, withdrawal, and coingestion with other unregulated substances, highlighting tianeptine as an emerging public health concern.

(JMIR Infodemiology 2026;6:e86683) doi:10.2196/86683

KEYWORDS

tianeptine; Reddit; social media; opioid use disorder; natural language processing; surveillance

Introduction

Tianeptine is an atypical antidepressant that modulates glutamate activity and, at high doses, agonizes mu opioid receptors. Although not approved by the US Food and Drug Administration, it is sold in unregulated marketplaces such as gas stations, convenience stores, smoke shops, and online. Reports suggest that individuals may use tianeptine recreationally or to manage pain, opioid use disorder (OUD), depression, or anxiety [1,2]. In countries where tianeptine is an approved and regulated antidepressant, its use has been linked to dependence and harm, leading several countries to classify it as a controlled substance. The population using tianeptine outside of formal medical settings remains poorly characterized. Social media platforms such as Reddit offer large, publicly available corpora of anonymous narratives that can support toxicovigilance when combined with natural language processing (NLP) methods [3,4].

This study used NLP approaches for data processing and manual coding for thematic analysis to characterize tianeptine-related discourse. The research questions were as follows: (1) How have posts mentioning tianeptine and related keywords changed over time? and (2) What themes characterize self-reported experiences with tianeptine?

Methods

Data Source and Identification of Tianeptine-Related Posts

We collected all Reddit data from 2005 - 2024 via Pushshift (2005 - 2023) and the Python Reddit API Wrapper (2024) We identified approximately 9000 drug-related subreddits to ensure discussion relevancy and reduce potential noise introduced by ambiguous/polysemous expressions. We searched for the keyword *tianeptine* in the collected data to identify relevant posts. We used a word2vec model [5] to identify expressions, including multiword expressions, that are semantically similar to *tianeptine*, and manually selected relevant ones. The final keyword set included *tianeptine*, *Neptune's Fix*, *Neptune's Elixir*, *Zaza*, *Zaza Red*, *Pegasus*, *Pegasus Red*, *Tianaa*, *TD Red*, *gas station heroin*, and *gas station dope*. Neptune's Fix Elixir, Zaza, Tianaa, Pegasus, and TD Red are common brands of tianeptine products. For potentially ambiguous terms (eg, "Pegasus"), we manually reviewed random samples to confirm relevance to tianeptine and excluded terms generating frequent false positives. Posts were grouped by year and keyword to describe trends over time. Counts for 2024 are reported descriptively but excluded from the trend figure.

Manually Coded Thematic Analysis

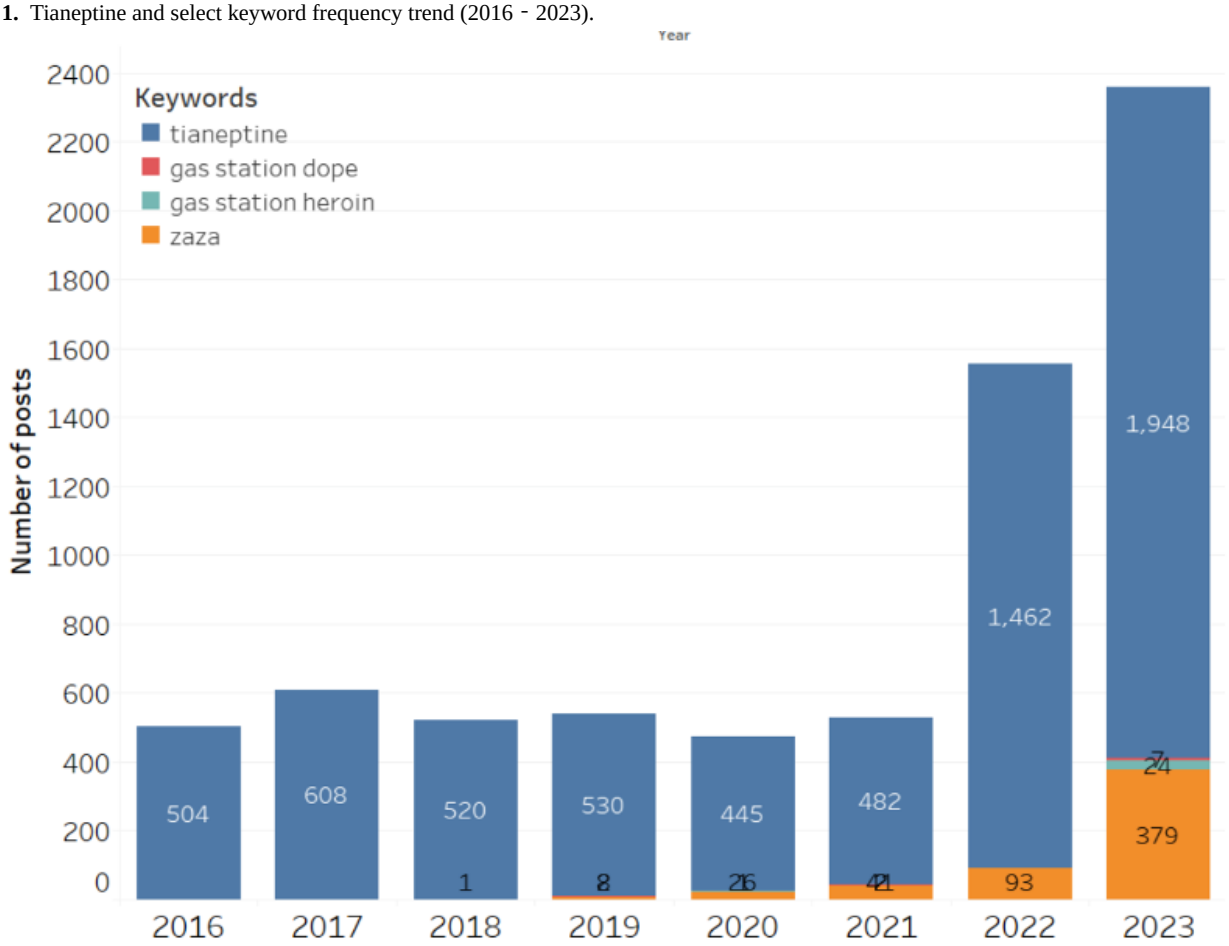
From 19,188 tianeptine-mentioning posts across 9707 subreddits, we drew a random sample of 200 posts for qualitative analysis. This analysis is described in detail in Multimedia

XSL•FO
RenderX

Appendix 1. Briefly, 3 authors with medical toxicology expertise conducted a preliminary open-coding exercise and noted recurrent topics, experiences, and patterns related to tianeptine use. The group then met to compare preliminary codes and organize them into broader conceptual categories. Through repeated discussion, the authors created a structured codebook specifying, for each theme, a brief descriptive label, an operational definition, criteria for inclusion, and one or more example summaries drawn from representative posts. The codebook was iteratively revised across multiple coding cycles until no additional themes were identified (see Multimedia Appendix 1 for the full codebook). After the final codebook was established, 2 reviewers independently coded all posts, with disagreements resolved through discussion and, when needed, adjudication by a third team member.

To assess interrater reliability, 2 coders independently applied the final 8-theme codebook to a random subset of 50 posts (25% of the qualitative sample). For each theme, we treated coding as a binary present/absent decision and calculated the Cohen κ based on a 2×2 contingency table of coder decisions.

Figure 1. Tianeptine and select keyword frequency trend (2016 - 2023).



Ethical Considerations

This study analyzed secondary data from only publicly available Reddit posts and did not involve interaction with individuals. Usernames and URLs were not included in the analysis. All examples are paraphrased to reduce reidentification risk. Additional methodological detail is provided in Multimedia Appendix 1.

Results

Frequency and Trends

We identified 19,188 posts from 9707 subreddits in 2005 - 2024. Approximately 26% (4912/19,188) were posted in the last 4 years. Annual counts increased sharply from 1555 (2022) to 2358 (2023; Figure 1). Posts referencing *Zaza* rose from 26 (2020) to 379 (2023). Comentioned substances included kratom (n=328), buprenorphine (Suboxone) (n=139), gabapentin (n=121), and pregabalin (n=75).

Manually Coded Themes

Among 200 sampled posts, 8 themes were identified (Table 1).

XSL•FO
RenderX

Table . Manually coded thematic analysis.

Theme	Theme frequency (n=200), n (%)	Paraphrased example
Withdrawal symptoms and management	68 (34)	A poster described experiencing cold sweats, diarrhea, confusion, panic, nausea, and headache within 24 hours of stopping tianeptine after a period of regular use. They attempted to manage symptoms with kratom.
Tianeptine dependence	56 (28)	A poster described an “insane rush” followed by the disappearance of all their problems and worries after their first time using tianeptine. After experiencing some hardships, they began using tianeptine daily and spending a significant amount of money on it.
Tianeptine use and opioid use disorder	16 (8)	A poster with a history of opioid use disorder reported discontinuing buprenorphine maintenance therapy and substituting tianeptine products. After experiencing withdrawal, they asked whether it would be safe to restart buprenorphine immediately.
Substances coingested	12 (6)	A poster described potentiating the effects of tianeptine by taking it with pregabalin, clonazepam, lorazepam, and dextromethorphan.
State-level bans	12 (6)	After a state implemented a tianeptine ban, a poster described visiting multiple retailers that had previously sold products “under the counter,” only to find all supplies depleted. They discussed ordering online from out-of-state vendors and considered traveling to a neighboring state.
Use for pain and depression	9 (4.5)	A poster with chronic gastrointestinal symptoms described using tianeptine when prescription medications and standard antidepressants provided insufficient relief, framing it as a way to “take the edge off” and improve daily functioning.
Adverse effects	7 (3.5)	A poster who had taken tianeptine regularly for one year inquired about tingling in the face and throat, nausea, and bodily sensations of heat and cold.
Suspected adulteration	1 (0.5)	A poster who had taken the same tianeptine brand for months reported suddenly experiencing severe effects from a new batch that resulted in hospitalization, raising concern that the product might be adulterated with a synthetic cannabinoid or other substance.

Withdrawal Symptoms and Management

Withdrawal symptoms and management strategies were most frequently mentioned (n=68, 34%). Posters described gastrointestinal symptoms, sweating, dysphoria, restlessness, and sleep disruption, comparing their experiences to opioid withdrawal (n=34, 17%). Many described self-directed tapering or switching to other substances (n=23, 11.5%).

Dependence

Tianeptine dependence was coded in 56 posts (28%). Posters recounted escalating doses, difficulty controlling use despite adverse consequences, and financial strain (n=44, 22%). Some described hospitalizations after ingesting gram-level doses of tianeptine (n=3, 1.5%).

OUD

Sixteen posts (8%) discussed tianeptine use in the context of OUD, including transitioning from prescription opioids or heroin to tianeptine, or using buprenorphine to manage tianeptine withdrawal. Kratom was the most frequently comentioned substance (n=38, 19%).

State-Level Bans, Adverse Effects, and Suspected Adulteration

Twelve posts (6%) addressed state-level bans, including abruptly losing access and turning to online vendors. Some described concern about phenibut withdrawal from coformulated products (n=6, 3%). Seven posts (3.5%) focused on adverse effects, and one described suspected adulteration.

XSL•FO
RenderX

Reliability Analysis

Interrater reliability for thematic classification was moderate overall. The mean Cohen κ across themes that occurred at least once in the reliability subset was 0.52, with a mean raw agreement of 89%. For withdrawal symptoms and management, agreement was 90% (κ=0.72); for dependence, 82% (κ=0.48); for OUD, 84% (κ=0.29); for coingestants, 80% (κ=0.52); for state-level bans, 100% (κ=1.00); for use for pain and depression, 92% (κ=0.62); and for suspected adulteration, 100% (κ=1.00). No posts in this subset were coded with the theme of adverse effects by either coder, so interrater reliability could not be meaningfully estimated for that theme.

Discussion

Reddit discussion of tianeptine increased markedly in 2022‐2023. Thematic analysis highlighted dependence, withdrawal management, and discussions of other substances. Reports of buprenorphine use to manage withdrawal align with case reports, though posters frequently described unproven strategies, underscoring the need for improved access to evidence-based addiction treatment.

Frequent comentions of kratom and gabapentinoids suggest overlapping populations engaged in polysubstance use. Posts describing abrupt withdrawal after state scheduling illustrate possible unintended consequences of rapidly changing drug markets, though these data are descriptive and cannot establish causality.

Several posts provide insight into the impact of state-level scheduling of tianeptine as a controlled substance. Individuals may acquire tianeptine from alternative markets or may experience abrupt withdrawal. One post from October 2023 described an experience with the tianeptine product Neptune's Fix in which the poster suspected adulteration of the product, describing an "alternative cannabinoid high." Samples of this product were later analytically confirmed to be adulterated with synthetic cannabinoids. This finding was reported in the New York Times on January 10, 2024, and in the literature on February 1, 2024 [6,7].

This study suggests that applying NLP for data processing alongside manually coded thematic analysis of Reddit posts may be useful for toxicosurveillance. Recent work using transformer models and large language models has shown that automated pipelines can classify and track opioid-related social media chatter at scale, extending earlier machine learning–based toxicovigilance approaches. Additionally, social media data can capture clinically relevant signals about mental health, reinforcing the value of online platforms for population-level surveillance of risk and vulnerability [8-10]. Our findings add to this literature by focusing on self-reported experiences with an emerging drug.

Limitations include Reddit's nonrepresentative demographics, self-reported data lacking clinical verification, coding of only 200 randomly selected posts, and the descriptive design precluding causal inference. Additionally, changes to the data available for 2024 limited analysis to posts through 2023, which may limit temporal generalizability. Nonetheless, the observed increase in tianeptine-related posts, coupled with frequent self-reported dependence and withdrawal, reinforces tianeptine as a drug of public health concern. Simple NLP-assisted approaches may help monitor evolving substance use patterns in near real time.

Funding

Research reported in this publication was supported in part by the National Institute on Drug Abuse of the National Institutes of Health (NIH) under award R01DA057599. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authors' Contributions

JP conceived and supervised the study. SL and AS performed the natural language processing–driven frequency analysis. CJC, AVS, SL, and JP conducted the thematic analysis. RW, JL, and DC assisted with the study and analyses. CJC led the writing, and all authors contributed substantially to revision.

Conflicts of Interest

SL and AS report receipt of grant funding from the National Institutes of Health. Otherwise, the authors declare no actual or potential conflicts of interest.

Multimedia Appendix 1
Methods for data collection and natural language processing and details of qualitative thematic analysis.
[DOCX File, 200 KB - infodemiology_v6i1e86683_app1.docx]

References

1. Rawal VY, Gallardo M, Henderson K, Hall OT, Klisovic N, Sikic-Klisovic E. Severe tianeptine withdrawal symptoms managed with medications for opioid use disorder: a case report. J Addict Dis 2025;43(1):98-103. [doi: 10.1080/10550887.2023.2290139] [Medline: 38095574]

XSL•FO
RenderX

2. Marraffa JM, Stork CM, Hoffman RS, Su MK. Poison control center experience with tianeptine: an unregulated pharmaceutical product with potential for abuse. Clin Toxicol 2018 Nov 2;56(11):1155-1158. [doi: 10.1080/15563650.2018.1476694]
3. Spadaro A, O'Connor K, Lakamana S, et al. Self-reported xylazine experiences: a mixed-methods study of Reddit subscribers. J Addict Med 2023;17(6):691-694. [doi: 10.1097/ADM.0000000000001216] [Medline: 37934533]
4. Smith KE, Rogers JM, Strickland JC, Epstein DH. When an obscurity becomes trend: social-media descriptions of tianeptine use and associated atypical drug use. Am J Drug Alcohol Abuse 2021 Jul 4;47(4):455-466. [doi: 10.1080/00952990.2021.1904408] [Medline: 33909525]
5. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. arXiv. Preprint posted online on Oct 16, 2013. [doi: 10.48550/arXiv.1310.4546]
6. 'Gas-station heroin' sold as dietary supplement alarms health officials. New York Times. 2024. URL: https://www.nytimes.com/2024/01/10/health/gas-station-heroin-tianeptine-addiction.html [accessed 2026-06-24]
7. Counts CJ, Spadaro AV, Cerbini TA, et al. Notes from the field: cluster of severe illness from Neptune's Fix tianeptine linked to synthetic cannabinoids — New Jersey, June–November 2023. MMWR Morb Mortal Wkly Rep 2024 Feb 1;73(4):89-90. [doi: 10.15585/mmwr.mm7304a5] [Medline: 38300852]
8. Fassi L, Ferguson AM, Przybylski AK, Ford TJ, Orben A. Social media use in adolescents with and without mental health conditions. Nat Hum Behav 2025 Jun;9(6):1283-1299. [doi: 10.1038/s41562-025-02134-4] [Medline: 40325198]
9. Sidorov G, Ahmad M, Basile P, Waqas M, Orji R, Batyrshin I. Monitoring opioid-related social media chatter using natural language processing and large language models: temporal analysis. JMIR Infodemiology 2025 Nov 4;5:e77279. [doi: 10.2196/77279] [Medline: 41187282]
10. Ahmad M, Batyrshin I, Sidorov G. Sentiment analysis using a large language model-based approach to detect opioids mixed with other substances via social media: method development and validation. JMIR Infodemiology 2025 Jun 19;5:e70525. [doi: 10.2196/70525] [Medline: 40536906]

Abbreviations:

NLP : natural language processing
OUD: opioid use disorder

Edited by T Mackey; submitted 28.Oct.2025; peer-reviewed by M Ahmad, SJ Kim; revised version received 29.Apr.2026; accepted 11.May.2026; published 07.Jul.2026.

Please cite as:
Counts CJ, Spadaro AV, Lakamana S, Sarker A, Wightman R, Love J, Calello D, Perrone J
Self-Reported Tianeptine Experiences on Reddit: Natural Language Processing–Assisted Qualitative Study
JMIR Infodemiology 2026;6:e86683
URL: https://infodemiology.jmir.org/2026/1/e86683
doi:10.2196/86683

© Christopher J Counts, Anthony V Spadaro, Sahithi Lakamana, Abeed Sarker, Rachel Wightman, Jennifer Love, Diane Calello, Jeanmarie Perrone. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 7.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mpox on Instagram: Content Analytic Study

Elizabeth E Havron*, BSPH; Kylee Chenault*, BSPH; Danny Valdez*, PhD; Rebecca F Houghton*, MPA; Eric R Walsh-Buhi*, MPH, PhD

Department of Applied Health Science, Indiana University School of Public Health-Bloomington, 1025 E. 7th Street, Room 116, Bloomington, IN, United States

*all authors contributed equally

Corresponding Author:
Eric R Walsh-Buhi, MPH, PhD
Department of Applied Health Science, Indiana University School of Public Health-Bloomington, 1025 E. 7th Street, Room 116, Bloomington, IN, United States

Abstract

Background: Mpox was declared a public health emergency of international concern in 2022. Instagram is widely used by age groups and communities disproportionately affected by mpox; yet platform-specific evidence on mpox information characteristics and engagement remains limited.

Objective: The aim of this study is to characterize sources, content, and engagement features of mpox-related Instagram posts, to describe prevention and treatment framing, and to compare the top 10% most-liked posts with the remaining corpus.

Methods: We retrieved English-language public Instagram posts via CrowdTangle containing “mpox” or “monkeypox” dated May 5, 2022, to January 17, 2023 (initial N=18,616). Using a pretested, deductive codebook adapted from prior Instagram health studies, 2 coders completed 2 pilot rounds; variables with low agreement were excluded. A randomized analytic sample of 1000 posts was coded for source type, content features, and prevention/treatment framing. Descriptive statistics were computed. For engagement contrasts, we compared the top 10% most-liked posts with the bottom 90% using tests of differences in independent proportions (mean differences [MD] with *P* values).

Results: Most posts originated from organizations (760/1000, 76%) versus individuals (240/1000, 24%). Organizational sources most commonly included businesses (436/760, 57.4%) and news/media outlets (401/760, 52.8%); government (174/760, 22.9%), nonprofits (131/760, 17.3%), and health care organizations (70/760, 9.2%) were less frequent. About one-third of posts cited a source (344/1000, 34.4%), most often the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC)/other federal entity. Posts predominantly used illustrated images/infographics (827/1000, 82.7%); photos appeared in 47.3% (473/1000) and videos in 12.4% (124/1000) of posts. Prevention content appeared in 38.4% (384/1000) of posts, most commonly vaccination (684/1000, 68.5% of prevention posts), followed by avoiding close contact (145/1000, 14.5%), avoiding contact with objects (83/1000, 8.3%), abstaining from sexual activity (76/1000, 7.6%), and condom use (13/1000, 1.3%); 28.9% (289/1000) of prevention posts noted barriers. Treatment mentions were uncommon (25/1000, 2.5% traditional biomedical; 2/1000, 0.2% alternative). Compared with the bottom 90%, the top 10% most-liked posts (1) were more likely to originate from public figures/celebrities among individuals (MD=−0.591; *P*<.001) and from businesses (MD=−0.299; *P*<.001) or news/media (MD=−0.350; *P*<.001) among organizations; (2) were less likely to be from government (*P*<.001) , nonprofit (*P*=.006), or health care organizations (*P*=.005); and (3) more often included nonmoving images (MD=−0.119; *P*=.024), visible lesion depictions (MD=−0.081; *P*=.035), prevalence mentions (MD=−0.180; *P*<.001), and citations (MD=−0.162; *P*=.001).

Conclusions: During the initial outbreak period, the highly engaged mpox content on Instagram skewed toward posts by public figures and news/business accounts and toward static, citation-bearing visuals that included prevalence context and occasionally lesion imagery. Public-health communicators seeking reach on Instagram should prioritize clear static infographics with explicit source citation and epidemiologic context and consider copublishing with trusted creators and news outlets, while addressing access barriers highlighted in prevention posts.

(JMIR Infodemiology 2026;6:e85379) doi:10.2196/85379

KEYWORDS

digital health; social media; mpox; Instagram; content analysis

XSL•FO
RenderX

Introduction

Mpox, a rare zoonotic disease caused by the "Monkeypox" virus, a species of the genus *Orthopoxvirus*, was declared a public health emergency of international concern on July 23, 2022 by the World Health Organization (WHO) [1,2]. Since the May 2022 outbreak, the disease has spread to 114 nonendemic countries. As of January 2024, the latest mpox outbreak has resulted in 99,518 cases globally and 32,063 cases in the United States [3]. A great majority of confirmed cases, about 82%, are people who identify as bisexual or who are men who have sex with men [4]. These data depict a disproportionate infection rate among men who have sex with men, which has accompanied concerns about even greater numbers of infections within this community and other priority populations.

Additionally, for many adults, social media represents a primary outlet for engaging with current events, including those pertaining to health concerns such as mpox [5]. Social media is used by almost 60% of the world and reaches a diverse population [6]. Instagram, in particular, offers users exposure to a variety of media, including photos, videos, and infographics, and has 3 billion monthly active users and over 500 million daily active users [7]. In 2025, 80% of Americans aged 18 to 29 years and 62% of Americans aged 30 to 49 years reported Instagram use as a social media platform [8]. In terms of race and ethnicity, 45% of White individuals, 52% of Black individuals, and 59% of Hispanic individuals report using Instagram [8]. Social media is used more by people who identify as lesbian, gay, bisexual, transgender, and queer/questioning (LGBTQ+) than by people who identify as heterosexual [9]. These data correlate with the demographics of reported mpox cases in the United States. According to the Centers for Disease Control and Prevention (CDC), the median age of people diagnosed with mpox was 35 years, and 94% of cases have been among men who have sex with men; 41%, 26%, and 28% of people diagnosed with mpox are White, Black, and Hispanic, respectively [10]. Furthermore, there is insufficient evidence on the use of Instagram as a modality for spreading information and awareness specifically regarding mpox. As such, a deep review of mpox-related Instagram content, as an example of a modern public health crisis unfolding online [11], is both timely and warranted, especially given that mpox knowledge and awareness in the United States are very low [12], and the populations with higher mpox infection prevalence are also among those with higher proportions utilizing Instagram.

The purpose of this study was to explore mpox-related content conveyed on Instagram and intentionally focus on the initial outbreak period (May 2022 to January 2023) to capture early-stage infodemic dynamics, which are critical for understanding how information ecosystems emerge during public health crises. Our study was guided by the following research questions:

- RQ1: What are the source and content characteristics of Instagram posts related to mpox?
- RQ2: What mpox causes and solution framing are conveyed in Instagram posts (ie, what is the content on Instagram related to mpox causes, treatment, and prevention)?
- RQ3: What are the content characteristics of the top 10% most-liked posts on mpox within the sample?

Findings from this work stand to inform the general digital ecosystem pertaining to mpox on Instagram. Furthermore, the contextualization of these findings may similarly suggest the possibility for future digital health interventions choosing to leverage social media outlets to distribute accurate, scientifically vetted information.

Methods

Data

Data for this study were collected from Instagram, a Metaverse photo-sharing social media platform. To source Instagram posts, our team utilized CrowdTangle Search, a public insights and third-party web-scraping tool with access to Metaverse data. We leveraged CrowdTangle to filter through all publicly accessible English-language posts that contained one of 2 queries (similar to a library search): (1) "mpox," the updated CDC-supported name for the virus, or (2) "monkeypox," the former name of the virus. Data were collected for posts dated from May 5, 2022, through January 17, 2023. English-language posts were identified based on caption language. This study used publicly available social media data. No private or identifiable user data were collected or reported. All data were analyzed in aggregate, further described below, and no individual accounts are identifiable in this study. A total of 18,616 posts met these inclusion criteria and were downloaded for use in 2 iterative pilot coding rounds, and for the final randomized data sample, as specified below. As of August 14, 2024, CrowdTangle has been shut down and replaced by the Meta Content Library, which has more limited access.

Analyses

Our study employed a manual qualitative deductive coding procedure to review the data. "Deductive coding" refers to a typology of qualitative research where a researcher begins with a theoretical framework, sets of hypotheses, or previously established coding schemas. For our study, we relied on a pretested codebook that had been applied to previous Instagram-related research studies, including pre-exposure prophylaxis for HIV prevention and skin cancer [13,14]. We adapted this codebook to reflect discourse pertaining to mpox, which was further stratified along several key dimensions, including source characteristics; mpox risk factors; prevention methods; treatment methods, benefits, and barriers; stigma; and citation attributions. These dimensions reflected our coding schema. Each post included for analysis was compared against our rubric. Prior to analyzing the final dataset, Cohen κ statistics were analyzed for interrater reliability among our human coders using RStudio (Posit, PBC, formerly known as RStudio, Inc). The full codebook with associated Cohen κ values can be found in Multimedia Appendix 1. For the final analysis, IBM SPSS 26 was used to analyze descriptive statistics. For RQ3, specifically, tests for statistical differences (ie, Wald H0 differences in proportion) between independent sample proportions were run between the top 10% most-liked posts in the sample and those in the bottom 90% of liked posts. The "top 10% most-liked posts" were defined as the subset of posts in

XSL•FO
RenderX

our analytic sample (N=1000) with the highest number of likes. Specifically, posts were ranked by total like count, and the top decile (n=100 posts) was classified as "most-liked," while the remaining 90% (n=900 posts) served as the comparison group. This approach allowed us to examine differences in content characteristics between highly visible posts and the broader corpus. Wald tests were used due to a large sample size and suitability for comparing independent proportions. Sparse categories were interpreted cautiously. Mean differences (MDs) and associated *P* values are reported.

Data Coding

Two human coders examined profile source and message content characteristics. The 2 coders and a moderator conducted 2 iterative pilot rounds using the most chronologically recent posts in the downloaded CrowdTangle dataset, where both coders coded the same set of posts and were assessed for interrater reliability. After the second round, among 72 variables and 63 out of 70 coded posts that met inclusion criteria, Cohen κ statistics ranged from −0.17 to 1.0, with a mean κ of 0.70 and median κ of 0.83, indicating substantial to perfect agreement across the majority of variables. Variables with none to slight agreement after the second round of pilot coding included posts' cues to action, portrayal of mpox risk factors, diagnosis- and treatment-related content, and the extent to which posts conveyed the seriousness of mpox. Variables with low interrater reliability (eg, cues to action, risk factors, and seriousness) were excluded to preserve analytic validity. Prevention-related variables were retained and serve as proxies for risk-related content.

Sample

Following the 2 pilot rounds, a total of 1200 Instagram posts were randomized and split between the 2 independent human coders with the intention of achieving a final dataset of 1000 for analyses. A sample of 1000 posts was selected to ensure adequate representation across content features, and this sample size is consistent with prior Instagram-based content analyses using manual (human) coding approaches [13,14]. Coder 1 coded 499 posts and Coder 2 coded 501 posts that ultimately met the inclusion criteria. We excluded posts that did not mention mpox or were not written in English. Coded variables included, but were not limited to, source characteristics (ie, organization type [eg, business and nonprofit] and individual type [eg, physician and journalist]) as well as mention of mpox causes, prevention methods, and treatment. The determination of the type of individual or organization behind the profile was based on self-described information in the account biography and/or post content (eg, credentials provided such as MD, journalist, or organizational affiliation), consistent with prior Instagram content analyses [13,14]. Photographic content was coded based on visible features within the image (eg, mpox lesion depiction, presence of individuals, and contextual clues), following predefined codebook criteria. Hashtags were extracted from post captions and analyzed descriptively. Frequencies of hashtag usage were computed, and the top 10 most common hashtags were identified. The full codebook can be found in Multimedia Appendix 1.

Ethical Considerations

This study utilized publicly available Instagram data and did not involve direct interaction or intervention with human participants. The study was reviewed by the Institutional Review Board of Indiana University-Bloomington, affiliated with the last author and was determined not to constitute human subjects research.

Results

RQ1: What Are the Source and Content Characteristics of Instagram Posts Related to Mpox?

Sources

Roughly three-quarters (n=760, 76%) of the posts in this sample originated from organizational Instagram accounts, while the remaining quarter (n=240, 24%) originated from the accounts of individuals. Among organizations, a majority were classified as businesses (436/760, 57.4%) and/or news outlets (401/760, 52.8%). Posts originating from state, federal, or foreign equivalent comprised 22.9% (174/760) of organizational posts, followed by government entities (174/760, 22.9%), nonprofit organizations (131/760, 17.3%), health care organizations (70/760, 9.2%), and school districts (19/760, 2.5%). Among posts by individual accounts, almost half were made by public figures (107/240, 44.6%), followed by those who identified as parents (37/240, 15.4%), journalists (15/240, 6.3%), and physicians (13/240, 5.4%). Health workers, health educators, teachers, children, and epidemiologists each represented less than 3% (7/240) of the sample, as specified in Table 1.

XSL•FO
RenderX

Table . Source characteristics of Instagram posts regarding mpox[a].

Source characteristics	All, n (%)	Most liked, n (%)	Less liked, n (%)	Mean difference (SE)	*P* value
Of posts made by an individual (N=240 for all; n=29 for top 10% most liked; n=211 for less liked), the bio/profile or post content mentioned...					
Public figure/celebrity with a minimum of 20,000 followers or having a verified account	107 (44.6)	28 (96.6)	79 (37.4)	−0.591 (0.048)	<.001
Being a parent (eg, mother or father)	37 (15.4)	1 (3.4)	36 (17.1)	0.136 (0.043)	.06
Being a journalist or member of the press	15 (6.3)	1 (3.4)	14 (6.6)	0.032 (0.007)	.60
Being a physician (eg, medical doctor, MD[b], DO[c], resident)	13 (5.4)	1 (3.4)	12 (5.7)	0.022 (0.037)	.62
Being a health worker	6 (2.5)	1 (3.4)	5 (2.4)	−0.011(0.035)	.73
Being a health educator	3 (1.3)	1 (0)	3 (1.4)	0.014 (0.008)	.52
Being a teacher, educator, or other school official	3 (1.3)	1 (3.4)	2 (0.9)	−0.025 (0.035)	.26
Being a child (eg, son or daughter)	2 (0.8)	0 (0)	2 (0.9)	0.009 (0.007)	.60
Being an epidemiologist	2 (0.8)	0 (0)	2 (0.9)	0.009 (0.007)	.60
Of post made by an organization (N=760 for all posts; n=71 for top 10% most liked; n=689 for less liked), the bio/profile/affiliated website mentioned...					
Business (eg, company, franchise, store, product, or service)	436 (57.4)	60 (84.5)	376 (54.6)	−0.299 (0.047)	<.001
News or media organization	401 (52.8)	60 (84.5)	341 (49.5)	−0.350 (0.047)	<.001
City, state, or federal government	174 (22.9)	2 (2.8)	172 (25)	0.222 (0.026)	<.001
Nonprofit/advocacy group	131 (17.3)	4 (5.6)	127 (18.5)	0.129 (0.031)	.006
Health care organization	70 (9.2)	0 (0)	70 (10.2)	0.102 (0.012)	.005
School or school district	19 (2.5)	2 (2.8)	17 (2.5)	−0.003 (0.021)	.86

[a](1) Source characteristics do not add up to 100% due to information being unavailable on the individual poster's Instagram biography. (2) Source characteristics exceed 100% because an organization may be included under multiple categories (eg, a business and a health care organization).

[b]MD: Doctor of Medicine.

[c]DO: Doctor of Osteopathic Medicine.

Cited Content

Among the full sample, about one-third (344/1000, 34.4%) of posts directly cited a source for their content. Of those posts, citations of the WHO were most prevalent, with over one-third of posts with citations referencing the WHO (120/344, 34.9%). The CDC, other federal entities, or foreign equivalents were the next most referenced (94/344, 27.3%). Almost one-quarter (80/344, 23.3%) of the sampled posts cited a medical doctor, followed by government officials (77/344, 22.5%), state or local health departments (60/344, 17.4%), and members of the research community (Table 2). Only a handful of posts cited a personal account from an individual who mentioned either firsthand or secondhand experience with mpox.

XSL•FO
RenderX

Table . Source attribution in mpox-related Instagram post citations.

Source attribution: The post cites information from...	All (N=344), n (%)	Most liked (n=49), n (%)	Less liked (n=295), n (%)	Mean difference (SE)	*P* value
Centers for Disease Control or other federal level or foreign equivalent	94 (27.3)	17 (38.8)	77 (26.1)	−0.086 (0.073)	.21
A medical doctor (physician)	80 (23.3)	11 (22.4)	69 (23.4)	0.009 (0.065)	.89
Government officials (eg, senators, governors, and representatives) or from political organizations	77 (22.5)	12 (22.4)	65 (22.1)	−0.024 (0.066)	.71
A state or local health department	60 (17.5)	5 (10.2)	55 (18.6)	0.084 (0.049)	.15
World Health Organization	120 (34.9)	19 (38.8)	101 (34.2)	−0.045 (0.075)	.54
A member of the research community (eg, researchers and scientists)	28 (8.2)	4 (8.2)	24 (8.1)	0.000 (0.042)	.99
Personal account of an individual who mentioned a firsthand experience with monkeypox (N=1000, full sample)	3 (0.3)	0 (0)	3 (0.3)	0.003 (0.002)	.57
Personal account of an individual who mentioned a secondhand experience with monkeypox (N=1000, full sample)	2 (0.2)	0 (0)	2 (0.2)	0.002 (0.002)	.64

Content Type

The majority of all posts (827/1000, 82.7%) related to mpox provided an image with an illustration (eg, infographic or text on the image, including data visualizations). Almost half (473/1000, 47.3%) displayed a photograph. Videos were depicted in 12.4% (124/1000) of posts overall (Table 3).

XSL•FO
RenderX

Table . Definitions, examples, and frequencies of textual content characteristics/variables regarding mpox-related messages.

Message characteristic	Definition	Example post description or quoted text	All posts (N=1000), n (%)	Top 10% most liked (n=100), n (%)	Bottom 90% less liked (n=900), n (%)
Profile type: individual	Profile represents a single person, not an organization or cause.	• Profile of Senator Bill Frist, according to the profile biography	240 (24)	29 (29)	211 (23.4)
Profile type: organization	Profile represents an organization, not a single person.	• Official profile of the San Bernardino County Public Health	760 (76)	71 (71)	689 (76.6)
Specified location	Location is specified anywhere in the post, either geotagged or referenced in text.	• San Bernardino County Public Health logo appears within image	622 (62.3)	45 (45)	577 (57.7)
Content type: photo	Post includes some sort of photo (nonmoving photo or image and snapshot).	—[a]	473 (47.3)	58 (58)	415 (46.1)
Content type: video	Post includes some sort of video (eg, boomerangs, GIFs[b], and IGTV[c]).	—	124 (12.4)	10 (10)	114 (12.7)
Content type: image with text	Post includes an illustration, text on image, and/or charts, etc (not specifically photo or video).	—	827 (82.7)	87 (87)	740 (82.2)
Race hashtag included	Post includes race-associated tags (eg, #Black and #Latino) or any mention of race in the text of the post.	• “#BlackQueer #BlackQueerJoy”	32 (3.2)	1 (1)	31 (3.4)
Gender hashtag included	Post includes gender-associated tags (eg, “man,” “boy,” “woman,” “girl,” “trans,” and “transgirl"). If yes, copy and paste text.	• “#empoweringmen #empoweringwomen”	15 (1.5)	1 (1)	14 (1.6)
Sexuality hashtag included	Post includes sexuality tags (eg, ”gay,” “homosexual,” “msm” [men who have sex with men], and “LGBTQ+”[d]). If yes, copy and paste text.	• “#LGBTQcommunity #QueerCouple #BlackQueerJoy #BlackQueer #GayCouple #GayLove #GayPride #Queerp-Pride#LGBTQPride #LGBTQLove”	42 (4.2)	4 (4)	38 (4.2)
Mpox visual depictions	Image displays mpox on a specific body part (ie, face, chest, arms, hands, legs, and genitalia).	—	157 (15.7)	22 (22)	134 (14.9)
Monkey	The image or video displays a picture of a monkey or monkeys.	—	41 (4.1)	6 (6)	35 (3.9)

XSL•FO
RenderX

Message characteristic	Definition	Example post description or quoted text	All posts (N=1000), n (%)	Top 10% most liked (n=100), n (%)	Bottom 90% less liked (n=900), n (%)
Mpox prevention methods	The post mentions mpox prevention methods?	• "Avoid close, skin-to-skin contact with infected people. Don't share cups, utensils, bedding, towels, clothing, upholstered seating, or other textiles with an infected person. Wash your hands or use an alcohol-based hand sanitizer. Get the smallpox vaccine if you are at high risk for infection"	384 (38.4)	36 (36)	348 (38.7)
Prevention: condoms (n=384)	Post promotes the use of condoms as a form of mpox prevention.	• "Because of current uncertainties about transmission, wear a condom for 12 weeks even after you have fully recovered"	5 (1.3)	1 (2.7)	4 (1.1)
Prevention: vaccine promotion (n=384)	Post promotes getting the vaccine as a form of mpox prevention.	• "Get the smallpox vaccine if you are at high risk for infection"	263 (68.5)	21 (56.8)	242 (69.5)
Prevention: avoiding close contact (n=384)	Post promotes avoiding close contact with infected persons as a form of mpox prevention.	• "Avoid close, skin-to-skin contact with infected people"	56 (14.6)	6 (16.2)	50 (14.4)
Prevention: abstinence (n=384)	Post promotes abstaining from kissing, hugging, cuddling, and any other sexual activity as a form of mpox prevention.	• "Stop or limit gay orgy attendance"	29 (7.6)	3 (8.1)	26 (7.5)
Prevention: objects (n=384)	Avoiding contact with objects (including bedding, sex toys, towels, and clothing) or sharing food/drink/other that an infected person has touched/used.	• "Don't share cups, utensils, bedding, towels, clothing, upholstered seating, or other textiles with an infected person"	32 (8.3)	4 (10.8)	28 (8)
Benefits of prevention (n=384)	The post mentions the benefits of a prevention method (eg, condoms also protect against sexually transmitted infections, pregnancy, etc).		28 (7.3)	1 (2.7)	27 (7.1)

XSL•FO
RenderX

Message characteristic	Definition	Example post description or quoted text	All posts (N=1000), n (%)	Top 10% most liked (n=100), n (%)	Bottom 90% less liked (n=900), n (%)
		• "Intradermal administration of the JYNNEOS vaccine allows providers to administer up to five separate doses from a single vial. This method will help us protect more people from getting #monkeypox with our existing supply of vaccines"			
Barriers to prevention (n=384)	The post mentions barriers to the prevention of mpox (ie, lack of knowledge on prevention methods, vaccines are unavailable).	• "ACAM2000 has a larger stockpile but more restrictions for administration, whereas JYNNEOS has more limited supply but can be administered to more people."	111 (28.9)	13 (35.1)	98 (28.2)
Alternative treatment methods	Mentions alternative mpox treatment (ie, herbal remedies, homemade remedies, and foods) either in text or reflected in the image.	• "they have mentioned making sure that you stay hydrated" • "make sure your body is clean and safe" • "use soap and water or antibiotic ointment"	2 (0.2)	0 (0)	2 (0.2)
Traditional biomedical treatment	Mentions traditional biomedical mpox treatment (ie, antiviral treatments, such as tecovirimat [TPOXX], etc) either in the text or reflected in the image.	• "The antiviral medication tecovirimat (TPOXX) is used to treat patients at risk for severe diseases who have painful lesions"	25 (2.5)	3 (3)	21 (2.3)
Benefits of treatment (n=25)	Mentions the benefits of a treatment (ie, smallpox antiviral drugs may be used to treat monkeypox)?	• "While there are no specific treatments for monkeypox infections, antiviral drugs licensed for smallpox use are effective and can be used against monkeypox. These might be advised for people more likely to get very ill, including those with weak immune systems."	8 (32)	1 (25)	7 (33.3)
			7 (28)	0 (0)	7 (33.3)

XSL•FO
RenderX

Message characteristic	Definition	Example post description or quoted text	All posts (N=1000), n (%)	Top 10% most liked (n=100), n (%)	Bottom 90% less liked (n=900), n (%)
Barriers to treatment (n=25)	Mentions barriers to treatment of mpox (ie, antiviral meds not accessible).	• "While there are no specific treatments for monkeypox infections, antiviral drugs licensed for smallpox use are effective and can be used against monkeypox. These might be advised for people more likely to get very ill, including those with weak immune systems."			
Prevalence	Mentions the prevalence of mpox (eg, "1 in 100 people are infected" or "50 cases have been found in Indiana").	• "There are now over 31,000 cases of Monkeypox reported globally, with over 9400 cases reported in the United States."	271 (27.1)	44 (44)	227 (25.2)
Stigma	The post is negative toward or stigmatizing of some priority population (eg, gay and bisexual men).	• "Don't have sex with your gay friend's dog."	43 (4.3)	6 (6)	37 (4.1)
Sources cited	The post cites a source for mpox information? (eg, CDC[e], WHO[f], a doctor, peer-review journal, etc. It is okay to indicate "yes," if the post originator is citing itself, as long as it is a clear citation.)	• "Listen as Dr Makary shares the reality of the situation."	344 (34.4)	49 (49)	295 (32.8)

[a]Not applicable.

[b]GIFs: graphics interchange formats.

[c]IGTV: Instagram TV.

[d]LGBTQ+: lesbian, gay, bisexual, transgender, and queer/questioning.

[e]CDC: Centers for Disease Control and Prevention.

[f]WHO: World Health Organization.

Location

A total of 622 posts included a geographical tag of some kind (Table 3). Of the 622 posts, 41.8% (n=260) geotagged a location outside of the United States. Among US-based posts, the top 10 most commonly represented states were California (n=66), New York (n=60), Michigan (n=30), New Jersey (n=24), Texas (n=20), Illinois (n=19), Georgia (n=17), North Carolina (n=12), Florida (n=10), and Massachusetts (n=9).

Hashtag Analysis

A total of 589 posts contained at least one hashtag. The most prevalent hashtag found across posts was #monkeypox, appearing roughly 8 times more frequently than the next most prevalent hashtags (ie, #health and #monkeypoxvirus). Other health-related terms used ranged from "#virus," "#who," and "#health" to other conditions like "#covid19" and sexuality-related tags such as "#lgbtq." Less than half (414/1000; 41.4%) of posts mentioned mpox by name, while the subsequent hashtags were referenced on much smaller portions of the overall sample. The top 10 most used hashtags are displayed in Table 4.

XSL•FO
RenderX

Table . Ten most common hashtags found in mpox-related Instagram posts (n=589 containing hashtags).

Hashtag	Count and prevalence, n (%)
#monkeypox	398 (67.6)
#health	48 (8.1)
#monkeypoxvirus	47 (8)
#covid19	40 (6.8)
#who	38 (6.5)
#news	37 (6.3)
#virus	34 (5.8)
#publichealth	30 (5.1)
#coronavirus	28 (4.8)
#lgbtq	26 (4.4)

RQ2: What Mpox Causes and Solution Framing Are Conveyed in Instagram Posts (ie, What Is the Content on Instagram Related to Mpox Causes, Treatment, and Prevention)?

Regarding solution framing, just over one-third (384/1000, 38.4%) of posts promoted at least one prevention method, most commonly the mpox vaccine (263/384, 68.5%), followed by avoiding close contact (56/384, 14.5%), avoiding contact with objects (32/384, 8.3%), and abstaining from sexual activity (29/384, 7.6%) and condom use (5/384, 1.3%). A sizable number of prevention-related posts highlighted barriers to prevention methods (111/384, 28.9%).

A very small percentage of the sample mentioned mpox treatments (25/1000, 2.5% mentioned traditional biomedical treatments, such as antiviral medication; 2/1000, 0.2% mentioned alternative treatment methods). Of those posts that were coded as having depicted traditional biomedical treatments for mpox (n=25), about one-third mentioned the benefits of treating it (eg, “antiviral drugs licensed for smallpox use are effective and can be used against monkeypox”). Among these same posts, 28% (7/25) mentioned barriers to receiving treatment (eg, “WHO does not recommend vaccinating against mpox” and “there is currently a limited supply of JYNNEOS, also known as Immune or Imvanex”).

Due to poor interrater reliability regarding depictions of mpox causes on Instagram, this element of the research question was not directly addressed. However, the prevention variables do capture aspects of risk, such as close exposure to infected persons (eg, skin-to-skin and exchange of bodily fluids) and objects used by infected persons.

RQ3: What Are the Content Characteristics of the Top 10% Most-Liked Posts on Mpox Within the Sample?

Statistically significant differences were identified between the top 10% most-liked posts and the rest of the posts in the sample. One of the most notable and expected differences is that the top 10% most-liked posts primarily originated from public figures or celebrities (MD=−0.591; P<.001) if they originated from an individual’s account, compared to the rest of the posts. Among organizational accounts, the top 10% most-liked posts were more likely to be coded as a business (MD=−0.299; P<.001) or news/media organization (MD=−0.350; P<.001). Among the bottom 90% liked posts, posts were more likely to originate from official governmental accounts (MD=0.222; P<.001), nonprofit organizations (MD=0.129; P=.006), and health care organizations (MD=0.102; P=.005).

A statistically significant higher proportion of the top 10% most-liked posts included a nonmoving photo/image in their posts, as opposed to the rest of the posts (MD=−0.119; P=.024), included visual depictions of mpox on body parts (MD=−0.081; P=.035), mentioned prevalence of mpox (MD=−0.18; P<.001), and cited sources (MD=−0.162; P=.001).

Discussion

Main Context

This study aimed to evaluate the content and source characteristics of mpox-related posts on Instagram. Broadly, our results highlight the quantity and diversity of health-related content that can be found online, including the range of content types, sources, and message framing within mpox-related Instagram posts. We contextualize our findings in relation to the existing academic literature below.

Mpox Prevention and Treatment Posts Highlight the Utility of Social Media for Health Communication

Social media maintains an unfortunate reputation as a harbinger of misinformation, social anxiety, and political divide [15]. Numerous studies have highlighted that increased polarization in the United States, and the rise of misinformation can largely be attributed to excessive social media use and poor content moderation [16]. However, other studies have highlighted that, in a health communication context, social media can potentially transcend numerous cultural, structural, and systemic barriers to care to promote knowledge and improve health outcomes [17]. This observation has been previously made in computational and qualitative studies pertaining to addiction and recovery [18], pre-exposure prophylaxis for HIV prevention [19], reproductive decision-making, and vaccine communication [20]. Collectively, these studies have observed that helpful content can offset poorly vetted or low-quality information.

XSL•FO
RenderX

However, certain criteria must be met, including evidence of a post originating from a verified account, culturally adaptive messaging, and the presence of online, real-time moderators to mitigate the risk of misinformation.

Our data largely contain evidence of accurate and insightful health promotion along the aforementioned criteria. For instance, within the top 10% of posts, much of our sample was coded as originating from a business entity or news organization. These posts were interpreted as part of information dissemination efforts during the early stages of the global mpox outbreak. In addition, we observed a substantive minority of posts related to treatment and prevention methods, most of which aligned with the standards set by the CDC and other similar entities. Importantly, our sample did not include hateful, vitriolic, or deliberately misinformed content about mpox. While this does not negate the potential existence of such content on Instagram, our findings suggest that, among highly engaged content, the information appears to be generally useful for individuals seeking answers about mpox risk, infection, spread, and treatment. Future research should consider expanding the scope to a larger sample of posts, which could help in identifying more generalizable findings or conducting a deeper qualitative review for information quality.

Broadly, our findings align with those of previous research illustrating the potential benefits of social media for cost-effective health communication and information sharing. For example, Jiang et al [21] highlighted that, despite known risks of social media and information quality, social media should play a continued and growing role in health communication for health promotion. As Valdez and Patterson [22] contended, the applicability of social media for health promotion will depend on the role of content moderation. On Reddit, for example, posts pertaining to health are highly moderated by trained professionals, which result in constructive and helpful messaging pertaining to myriad health behaviors, including those pertaining to sexual and reproductive health. Although it is unclear the extent to which Instagram is moderated by such professionals, our results at least illustrate that the platform, and likely the Metaverse more broadly, may be creating such avenues for informed health content. However, we acknowledge that caution must be undertaken given documented research citing Instagram's photo-sharing capability as a venue for body image concerns.

Comparison With Prior Work

Empirical work examining mpox on Instagram remains scarce. One study analyzing posts under #monkeypox documented early themes and misinformation risks, noting that stigmatizing and inaccurate narratives coexisted with public health messaging on the platform [22]. Compared with that work, our analysis uses a substantially larger randomized sample, a validated multidomain codebook adapted from prior Instagram health studies, and a direct comparison of the top 10% most liked posts to the remaining corpus, enabling inference about engagement-linked content characteristics (eg, source types, image features, prevalence mentions, citations).

Findings from adjacent platforms converge on several patterns that our study also observes. On TikTok, for example, multiple analyses reported variable information quality and prominent engagement with visually compelling content; videos created by physicians or authorities were relatively limited; and misinformation or conspiracy content appeared early in the outbreak, especially in trending feeds [23,24]. On X, formerly known as Twitter, studies have documented high volumes of mpox discourse with substantial misinformation and stigma toward LGBTQ+ communities, alongside network dynamics where news outlets and a small number of influencers shaped attention [11,25]. The analyses of Reddit have emphasized community problem-solving and information-seeking among gay and bisexual- and men who have sex with men–focused subreddits, often moderated to reduce misinformation and toxicity, with themes evolving as the outbreak and public health declarations unfolded [26]. Cross-platform work comparing Reddit and X found topic and psycholinguistic differences that map onto platform cultures—more community support and practical guidance on Reddit versus faster, more event-driven discourse on X [27]. Broader social media studies during the 2022 mpox emergency also traced toxicity dynamics and the need for targeted mitigation strategies, reinforcing the public health value of verified source cues and active moderation [28].

Against that backdrop, our Instagram results add 3 contributions: information regarding source composition, content features associated with engagement, and tone and quality signals. First, we show that highly engaged Instagram posts disproportionately originate from public figures/celebrities and news/business accounts, while government, nonprofit, and health care organizations populate the long tail—mirroring influence concentrations on X but documented here for Instagram with a larger coded sample. Second, nonmoving images (vs videos), visible depictions of lesions, explicit prevalence mentions, and the presence of citations all characterize the top-liked posts; prior TikTok and X studies did not simultaneously test this combination of features on Instagram. Third, unlike some cross-platform studies that reported on misinformation and stigma, our sample of highly engaged Instagram posts contained comparatively little hateful or deliberately misleading content, suggesting that—at least during the window we studied—Instagram's most amplified mpox content skewed toward broadly informative material. This aligns with municipal and public health agency experiences that consistent, branded messaging can perform effectively on real-time platforms when paired with clear visuals and timely updates [29].

Limitations, Strengths, and Future Research Recommendations

This study is not devoid of limitations. First, we were only able to collect information from publicly available Instagram accounts. Second, this study intentionally focused on the initial outbreak period (May 2022 to January 2023). Accordingly, an area for future research might incorporate more recent data. Third, our inclusion criteria may have limited the generalizability of our results. For example, we only included Instagram content written in English. Fourth, we acknowledge that likes are influenced by account size and reach. While engagement rates were not available via CrowdTangle, we interpret the findings as reflecting visibility rather than

XSL•FO
RenderX

normalized engagement. Finally, given multiple statistical tests, findings should be interpreted cautiously as exploratory.

Despite these limitations, this study has several strengths and implications. First, this research remains the largest of its kind, to our knowledge, examining mpox on Instagram, one of the most used social media platforms in the United States. Only one other research study has examined mpox on Instagram, and that research focused on a smaller number of posts and was limited to assessing misinformation [30]. To date, a few other studies have examined mpox content via other social media platforms, including Reddit [31] and X [11,32]. Given that mpox-related posts by public figures and celebrities were among the most liked in this sample, medical experts could partner with celebrities and influencers to lead awareness campaigns on mpox.

Conclusions

Instagram has functioned as a consequential channel in the mpox information ecosystem. In a large, randomized sample of posts spanning the initial 2022 to 2023 outbreak period, we found that organizational accounts—especially news/media and businesses—and public figures disproportionately produced the most-liked content; visual stills, lesion depictions, prevalence statistics, and explicit citations were each associated with higher engagement; and prevention and treatment content—when present—generally aligned with authoritative guidance.

Taken together, these findings imply that public health communicators can improve Instagram reach and relevance by packaging guidance as clear static visuals (eg, infographics and single-image posts) that include succinct prevalence/context cues, citing sources (eg, WHO) prominently in the caption or image, and partnering with trusted creators and news accounts to copublish updates and frequently asked questions. Given cross-platform evidence of misinformation, stigma, and toxicity—particularly on TikTok and X—strategies that foreground verified provenance, rapid myth-busting, and community-informed moderation remain essential [23].

Finally, our engagement-contrast design suggests a practical playbook for future outbreaks: monitor which content attributes travel on each platform and then rapidly iterate message formats and partnerships accordingly. Future work should extend this analysis to reels and video formats, test message experiments (A/B) with creator collaborations, and link post-level exposures to downstream behaviors (eg, vaccine information-seeking and clinic clicks). As emergent mpox developments continue to surface—and new emergencies arise—Instagram can be leveraged as a scalable, equity-minded conduit for timely, accurate risk communication when campaigns adopt these evidence-based practices.

Acknowledgments

ChatGPT-5.5 was used in an editorial capacity to improve the readability of certain manuscript sections. At no point was artificial intelligence used for idea generation, direct text writing, inferencing, or data analysis procedures.

Funding

The authors declared no financial support or grants were received from any public, commercial, or not-for-profit entities for the research summarized in this article.. Publication costs are supported by Indiana University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook.
[DOCX File, 65 KB - infodemiology_v6i1e85379_app1.docx]

References

1. WHO recommends new name for monkeypox disease. World Health Organization. URL: https://www.who.int/news/item/28-11-2022-who-recommends-new-name-for-monkeypox-disease?ACSearchID=USCDCNPH162DM9882&ACSearchLabel=Mpox%20name%20change%20-%20mpox&ACSearchName=USCDCNPH162DM9882 [accessed 2026-03-29]
2. Mpox (monkeypox). World Health Organization. URL: https://www.who.int/europe/emergencies/situations/monkeypox [accessed 2026-03-29]
3. Ongoing 2022 global outbreak cases and data. Centers for Disease Control and Prevention (CDC). 2022. URL: https://archive.cdc.gov/www_cdc_gov/poxvirus/mpox/response/2022/index.html [accessed 2026-03-29]
4. Sharma A, Prasad H, Kaeley N, Bondalapati A, Edara L, Kumar YA. Monkeypox epidemiology, clinical presentation, and transmission: a systematic review. Int J Emerg Med 2023 Mar 17;16(1):20. [doi: 10.1186/s12245-023-00491-3] [Medline: 36932335]
5. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]

XSL•FO
RenderX

6. Malik A, Khan ML, Quan-Haase A. Public health agencies outreach through Instagram during the COVID-19 pandemic: crisis and emergency risk communication perspective. Int J Disaster Risk Reduct 2021 Jul;61:102346. [doi: 10.1016/j.ijdrr.2021.102346] [Medline: 36337987]
7. Kumar N. How Many People Use Instagram in 2026 [Active Users Stats]. DemandSage. URL: https://www.demandsage.com/instagram-statistics/ [accessed 2026-03-29]
8. Social media fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2025-11-20]
9. Seidenberg AB, Jo CL, Ribisl KM, et al. A national study of social media, television, radio, and internet usage of adults by sexual orientation and smoking status: implications for campaign design. Int J Environ Res Public Health 2017 Apr 21;14(4):450. [doi: 10.3390/ijerph14040450] [Medline: 28430161]
10. Philpott D, Hughes CM, Alroy KA, et al. Epidemiologic and Clinical Characteristics of Monkeypox Cases - United States, May 17-July 22, 2022. MMWR Morb Mortal Wkly Rep 2022 Aug 12;71(32):1018-1022. [doi: 10.15585/mmwr.mm7132e3] [Medline: 35951487]
11. Edinger A, Valdez D, Walsh-Buhi E, et al. Misinformation and public health messaging in the early stages of the mpox outbreak: mapping the Twitter narrative with deep learning. J Med Internet Res 2023 Jun 6;25:e43841. [doi: 10.2196/43841] [Medline: 37163694]
12. Walsh-Buhi ER, Walsh-Buhi ML, Houghton RF. Mpox knowledge in the U.S.: results from a nationally representative survey. J Infect Public Health 2024 Feb;17(2):359-361. [doi: 10.1016/j.jiph.2023.12.009] [Medline: 38198968]
13. Walsh-Buhi E, Houghton RF, Lange C, Hockensmith R, Ferrand J, Martinez L. Pre-exposure prophylaxis (PrEP) information on Instagram: content analysis. JMIR Public Health Surveill 2021 Jul 27;7(7):e23876. [doi: 10.2196/23876] [Medline: 34061759]
14. Gomaa B, Houghton RF, Crocker N, Walsh-Buhi ER. Skin cancer narratives on Instagram: content analysis. JMIR Infodemiology 2022;2(1):e34940. [doi: 10.2196/34940] [Medline: 37113805]
15. Walther JB. Social media and online hate. Curr Opin Psychol 2022 Jun;45:101298. [doi: 10.1016/j.copsyc.2021.12.010] [Medline: 35158213]
16. Tucker J, Guess A, Barbera P, et al. Social media, political polarization, and political disinformation: a review of the scientific literature. SSRN. 2018 Mar 19. [doi: 10.2139/ssrn.3144139]
17. Valdez D, Patterson M, Prochnow T. The importance of interdisciplinary frameworks in social media mining: an exploratory approach between Computational Informatics and Social Network Analysis (SNA). Health Behav Res 2021 Aug 19;4(2). [doi: 10.4148/2572-1836.1098]
18. Russell AM, Valdez D, Chiang SC, et al. Using natural language processing to explore “Dry January” posts on Twitter: longitudinal infodemiology study. J Med Internet Res 2022 Nov 18;24(11):e40160. [doi: 10.2196/40160] [Medline: 36343184]
19. Edinger A, Valdez D, Walsh-Buhi E, Bollen J. Deep learning for topical trend discovery in online discourse about pre-exposure prophylaxis (PrEP). AIDS Behav 2023 Feb;27(2):443-453. [doi: 10.1007/s10461-022-03779-2] [Medline: 35916950]
20. Valdez D, Soto-Vásquez AD, Montenegro MS. Geospatial vaccine misinformation risk on social media: online insights from an English/Spanish natural language processing (NLP) analysis of vaccine-related tweets. Soc Sci Med 2023 Dec;339:116365. [doi: 10.1016/j.socscimed.2023.116365] [Medline: 37984184]
21. Jiang S, Tay J, Ngien A, Basnyat I. Social media health promotion and audience engagement: the roles of information dissemination, organization-audience interaction, and action confidence building. Health Commun 2024 Apr;39(1):4-14. [doi: 10.1080/10410236.2022.2150809] [Medline: 36426852]
22. Valdez D, Patterson MS. Computational analyses identify addiction help-seeking behaviors on the social networking website Reddit: insights into online social interactions and addiction support communities. PLOS Digit Health 2022 Nov;1(11):e0000143. [doi: 10.1371/journal.pdig.0000143] [Medline: 36812569]
23. Shi A, El Haddad J, Cai P, et al. Mpox (monkeypox) information on TikTok: analysis of quality and audience engagement. BMJ Glob Health 2023 Mar;8(3):e011138. [doi: 10.1136/bmjgh-2022-011138] [Medline: 36918216]
24. Ji-Xu A, Htet KZ, Leslie KS. Monkeypox content on TikTok: cross-sectional analysis. J Med Internet Res 2023 Jan 17;25:e44697. [doi: 10.2196/44697] [Medline: 36649057]
25. Dsouza VS, Rajkhowa P, Mallya BR, et al. A sentiment and content analysis of tweets on monkeypox stigma among the LGBTQ+ community: a cue to risk communication plan. Dialogues Health 2023 Dec;2:100095. [doi: 10.1016/j.dialog.2022.100095] [Medline: 36573228]
26. Hong C. Mpox on Reddit: a thematic analysis of online posts on mpox on a social media platform among key populations. J Urban Health 2023 Dec;100(6):1264-1273. [doi: 10.1007/s11524-023-00773-4] [Medline: 37580545]
27. Cascalheira CJ, Corro K, Hong C, et al. An analysis of mpox communication on Reddit vs Twitter during the 2022 mpox outbreak. Sex Res Soc Policy 2025 Sep;22(3):1536-1551. [doi: 10.1007/s13178-024-01058-4]
28. Fan L, Li L, Hemphill L. Toxicity on social media during the 2022 mpox public health emergency: quantitative study of topical and network dynamics. J Med Internet Res 2024 Dec 12;26:e52997. [doi: 10.2196/52997] [Medline: 39666969]

XSL•FO
RenderX

29. Boyce MR, Gordon M, Piltch-Loeb R, Katz R. Chicago public health department social media communications on Twitter during the COVID-19 pandemic and the mpox epidemic: cross-sectional content analysis. J Med Internet Res 2025 Jul 18;27:e68200. [doi: 10.2196/68200] [Medline: 40680161]
30. Dalton ME, Duffy R, Quinn E, et al. A qualitative review of social media sharing and the 2022 monkeypox outbreak: did early labelling help to curb misinformation or fuel the fire? Sex Health 2024 Feb;21(1):SH23158. [doi: 10.1071/SH23158] [Medline: 38295805]
31. Anoop VS, Sreelakshmi S. Public discourse and sentiment during mpox outbreak: an analysis using natural language processing. Public Health 2023 May;218:114-120. [doi: 10.1016/j.puhe.2023.02.018] [Medline: 37019026]
32. Thakur N, Duggal YN, Liu Z. Analyzing public reactions, perceptions, and attitudes during the mpox outbreak: findings from topic modeling of tweets. Computers 2023;12(10):191. [doi: 10.3390/computers12100191]

Abbreviations

CDC: Centers for Disease Control and Prevention
LGBTQ+: lesbian, gay, bisexual, transgender, and queer/questioning
MDs: mean differences
WHO: World Health Organization

Edited by T Purnat; submitted 06.Oct.2025; peer-reviewed by A Jain, MA Zoubi; revised version received 13.Apr.2026; accepted 06.May.2026; published 30.Jun.2026.

Please cite as:
Havron EE, Chenault K, Valdez D, Houghton RF, Walsh-Buhi ER
Mpox on Instagram: Content Analytic Study
JMIR Infodemiology 2026;6:e85379
URL: https://infodemiology.jmir.org/2026/1/e85379
doi:10.2196/85379

© Elizabeth E Havron, Kylee Chenault, Danny Valdez, Rebecca F Houghton, Eric R Walsh-Buhi. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 30.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Tetanus Health Information on YouTube and TikTok: Cross-Sectional Analysis of Quality, Reliability, and Public Health Implications

Rui Huang*, BM; Yajing Shen*, MMed; Guangqing Zou, MMed; Jun Zhou, MMed; Xiaolu Deng, MSc; Qian Li, BM; Xiaoxiong Chen, MMed

Department of Emergency Medicine, The First Affiliated Hospital (Southwest Hospital) of Army Medical University, No. 30, Gaotanyan Zheng Street, Shapingba District, Chongqing, China

*these authors contributed equally

Corresponding Author:
Xiaoxiong Chen, MMed
Department of Emergency Medicine, The First Affiliated Hospital (Southwest Hospital) of Army Medical University, No. 30, Gaotanyan Zheng Street, Shapingba District, Chongqing, China

Abstract

Background: Tetanus is a severe but vaccine-preventable neurological disease that remains a public health concern, especially in resource-limited settings. As social media becomes an important source of health information, concerns persist regarding the quality and reliability of tetanus-related content online.

Objective: This study aimed to evaluate the quality, reliability, and thematic characteristics of tetanus-related videos on YouTube and TikTok and to examine the relationship between engagement metrics and information quality.

Methods: A cross-sectional study was conducted using tetanus-related videos retrieved from YouTube and TikTok on August 1, 2025. The top 100 eligible videos from each platform were included (n=200). Video quality was assessed using the Global Quality Scale, whereas reliability and transparency were evaluated using the modified DISCERN tool and the *Journal of the American Medical Association* benchmark criteria. A thematic content analysis based on predefined coding categories was also performed. Video characteristics, source types, and engagement metrics were also collected. Spearman correlation analysis was used to examine associations between engagement indicators and quality scores.

Results: YouTube videos showed significantly higher quality and reliability than TikTok videos, with higher median Global Quality Scale, modified DISCERN, and *Journal of the American Medical Association* scores (all $P<.001$). Compared with TikTok, YouTube videos more frequently discussed symptoms (92% vs 81%, $P=.02$), prevention (95% vs 78%, $P<.001$), treatment (88% vs 70%, $P=.002$), and wound management (77% vs 38%, $P<.001$). Lower-quality videos commonly contained incomplete prevention information, vague symptom descriptions, and limited source attribution. Videos produced by official medical organizations and professional health care creators generally achieved higher quality scores. Although engagement indicators were strongly correlated with each other, their associations with informational quality were relatively limited.

Conclusions: YouTube provided more comprehensive and reliable tetanus-related information than TikTok, although content quality on both platforms remained inconsistent. Greater involvement from health care professionals and clearer evidence-based communication may help improve the quality of health information shared on social media platforms.

(JMIR Infodemiology 2026;6:e85397) doi:10.2196/85397

KEYWORDS

content analysis; health communication; infodemiology; reliability; social media; tetanus; TikTok; video quality; YouTube

Introduction

Tetanus is a severe neurological disease caused by *Clostridium tetani*, a spore-forming bacterium commonly found in soil and the gastrointestinal tracts of animals. It produces tetanus neurotoxin, a potent toxin that disrupts inhibitory neurotransmission and leads to muscle spasms and autonomic dysfunction [1]. Although effective vaccines and improved clinical management are available, tetanus continues to cause substantial morbidity and mortality worldwide [2]. The burden is particularly high in settings with limited health care resources, where access to vaccination and timely treatment may be restricted [3]. Recent reports from the World Health Organization indicate that 21,830 cases were documented globally in 2023, although the actual number is likely higher due to underreporting [4]. It is estimated that tetanus still accounts for tens of thousands of deaths each year [5]. These patterns suggest that, despite being preventable, tetanus remains

XSL•FO
RenderX

an ongoing public health challenge, highlighting the need for better awareness and more effective health communication.

With the widespread use of the internet, social media platforms such as YouTube and TikTok have become important channels through which the public accesses health information [6]. Approximately 80% of internet users seek health information online. YouTube, a widely used video-sharing platform that lacks formal peer review mechanisms for uploaded content, was accessed by 73% of adults and 91% of individuals aged 15 - 29 in the United States from 2019 to 2020 [7]. During and after the COVID-19 pandemic, YouTube experienced substantial growth in the consumption of health-related content, with reports indicating a 75% increase in views of health and news videos compared with the prepandemic period [8]. Although these platforms enable rapid dissemination and broad accessibility of health information, concerns remain regarding the accuracy, reliability, and completeness of online content. In addition, online engagement has increasingly been linked to research visibility through alternative metrics (Altmetrics), suggesting that digital dissemination may also influence the broader impact and public reach of health information [9,10].

The widespread use of video-based health information also raises concerns regarding health literacy. Major organizations, including the National Institutes of Health, the US Department of Health and Human Services, and the American Medical Association, recommend that patient education materials be written at or below a sixth-grade reading level [11]. However, many online resources exceed this level, potentially limiting accessibility and comprehension.

Infodemiology provides a useful framework to evaluate the distribution and quality of online health information [12]. Although prior studies have assessed social media content across various health topics, no study has systematically evaluated tetanus-related videos on major platforms. By integrating content quality assessment with engagement and dissemination metrics, this study contributes to the core objectives of infodemiology, offering insights into how health information is structured, spread, and interpreted within contemporary digital environments.

Therefore, this study aimed to assess the quality, reliability, and content characteristics of tetanus-related videos on YouTube and TikTok. Specifically, we aimed to (1) evaluate video quality and reliability using the Global Quality Scale (GQS) [13], the *Journal of the American Medical Association* (*JAMA*) benchmark criteria [14], and the modified DISCERN (mDISCERN; originally derived from the DISCERN instrument: Quality Criteria for Consumer Health Information on Treatment Choices) [15]; (2) analyze thematic content characteristics of the videos; (3) compare engagement metrics and their association with information quality across platforms; and (4) identify gaps to inform improved digital health communication strategies. The instruments mentioned in (1) have been validated in previous studies [16,17].

Methods

Ethical Considerations

This study did not involve human participants, clinical data, biological specimens, or laboratory animals. All data were obtained from publicly accessible video content on YouTube and TikTok. No identifiable personal information was collected. Therefore, institutional review board approval was not required.

Study Design

This cross-sectional study examined the quality, reliability, and dissemination characteristics of tetanus-related videos on YouTube and TikTok. The aim was to evaluate how these platforms present health information and identify potential gaps in online health communication.

Search Strategy and Data Collection

Videos were identified on August 1, 2025, using the keyword “tetanus” on YouTube and TikTok. Data collection was conducted on a single day to provide a cross-sectional snapshot of available content and reduce potential variation caused by changes in platform algorithms over time.

To reduce personalization bias, all searches were performed using newly created accounts with no prior viewing history or interactions. Browser cache and search history were cleared before data collection to further minimize algorithmic influence on search results. Videos were retrieved using the default platform ranking, and the top 100 most relevant videos from each platform were included (n=200). All selected videos were recorded and archived to ensure reproducibility.

Eligibility Criteria

Eligibility criteria were defined to ensure that the included videos were relevant, comparable, and aligned with the study objective of evaluating publicly accessible tetanus-related health information.

Videos were eligible for inclusion if they met the following criteria: (1) addressed tetanus-related education, prevention, or management; (2) were presented in English; (3) had a duration between 20 seconds and 20 minutes; and (4) had been available online for at least 10 days prior to data collection. Videos were excluded if they (1) were commercial advertisements or promotional materials, to avoid a noneducational bias; (2) were duplicates or reposted content, to prevent redundancy; (3) were not directly related to tetanus health information, to maintain thematic consistency. These criteria ensured that the final sample reflected meaningful, nonredundant, and publicly relevant educational content.

Video Categorization

Each video was categorized by its source to capture differences in credibility and educational approach. Categories included (1) medical official accounts (eg, hospitals, public health agencies), (2) nonmedical official accounts (eg, news media, science organizations), (3) professional medical creators (eg, doctors, nurses, medical educators), (4) nonmedical professional creators (eg, science communicators, educators), and (5) the general public (eg, personal experiences, lay users).

XSL•FO
RenderX

Evaluation Instruments

Videos were independently evaluated by two emergency physicians with expertise in tetanus prevention. Prior to formal assessment, both reviewers underwent standardized training to ensure consistent application of the scoring criteria.

Three validated instruments were used: (1) the mDISCERN tool, which assesses the reliability and quality of health information; (2) the GQS, which evaluates the overall educational quality and usefulness of the content; and (3) the *JAMA* benchmark criteria, which assess transparency based on authorship, attribution, disclosure, and currency.

All videos were scored strictly in accordance with the original published criteria for each instrument [13-15]. Detailed scoring frameworks have been extensively described in previous studies and were not repeated here to avoid redundancy.

Each video was independently rated by both reviewers, and the final score was calculated as the mean of the two assessments. Discrepancies were resolved through discussion until consensus was reached.

Thematic Content Analysis

In addition to quantitative quality assessment, a thematic content analysis was conducted to examine the informational characteristics of the videos. The coding framework was informed by prior infodemiology literature, commonly evaluated dimensions of online health information quality, and clinically relevant aspects of tetanus prevention and management. The predefined thematic categories included [18,19]: (1) disease overview, (2) symptoms, (3) prevention (eg, vaccination), (4) treatment, (5) wound management, and (6) misinformation or incomplete information.

Misinformation was defined as content that was inconsistent with established medical evidence or lacked essential information required for safe and accurate understanding. Two reviewers independently coded all videos according to this framework. Each thematic category was recorded as present or absent for each video. Disagreements were resolved through discussion and consensus. Intercoder reliability for thematic classification was assessed using Cohen κ statistics. In addition to structured coding, reviewers noted recurring features observed in lower-quality videos, such as incomplete or oversimplified explanations, potentially misleading statements, limited or absent supporting evidence, and lack of clear source attribution. For interpretive purposes, videos were generally considered to be of lower quality when they received relatively low scores on the applied evaluation tools (eg, GQS ≤2 and/or mDISCERN ≤1) and/or when such issues were identified during qualitative review. This approach enabled us to reflect not only measurable differences in scoring but also practical limitations in how health information was presented to viewers. The relationship between account type and thematic content was also descriptively examined to explore whether different sources emphasized distinct types of health information.

Data Extraction

For each video, we collected descriptive and engagement-related variables, including number of likes, comments, shares, favorites, views, followers, upload date, video duration, and account verification status. The geographical origin of the account was also recorded to explore potential regional differences in content distribution.

Statistical Analysis

All statistical analyses were performed using GraphPad Prism (version 10.1). Data normality was assessed prior to analysis. Categorical variables were expressed as frequencies and percentages, while nonnormally distributed continuous variables were presented as medians and interquartile ranges (IQR). Between-group comparisons of continuous variables were conducted using the Mann-Whitney *U* test for nonnormally distributed data and the independent-samples *t* test for normally distributed data, as appropriate. Categorical variables were compared using the χ^2 test or Fisher exact test, as appropriate. Spearman correlation coefficients were calculated to examine relationships between engagement metrics and professional quality scores. All statistical tests were 2-tailed, and a *P* value <.05 was considered statistically significant.

Results

Video Characteristics

A total of 200 videos met the inclusion criteria, comprising 100 from YouTube and 100 from TikTok. Overall, YouTube videos were significantly longer in duration compared with TikTok videos (*P*<.001). Basic video characteristics and engagement metrics are summarized in Table 1.

XSL•FO
RenderX

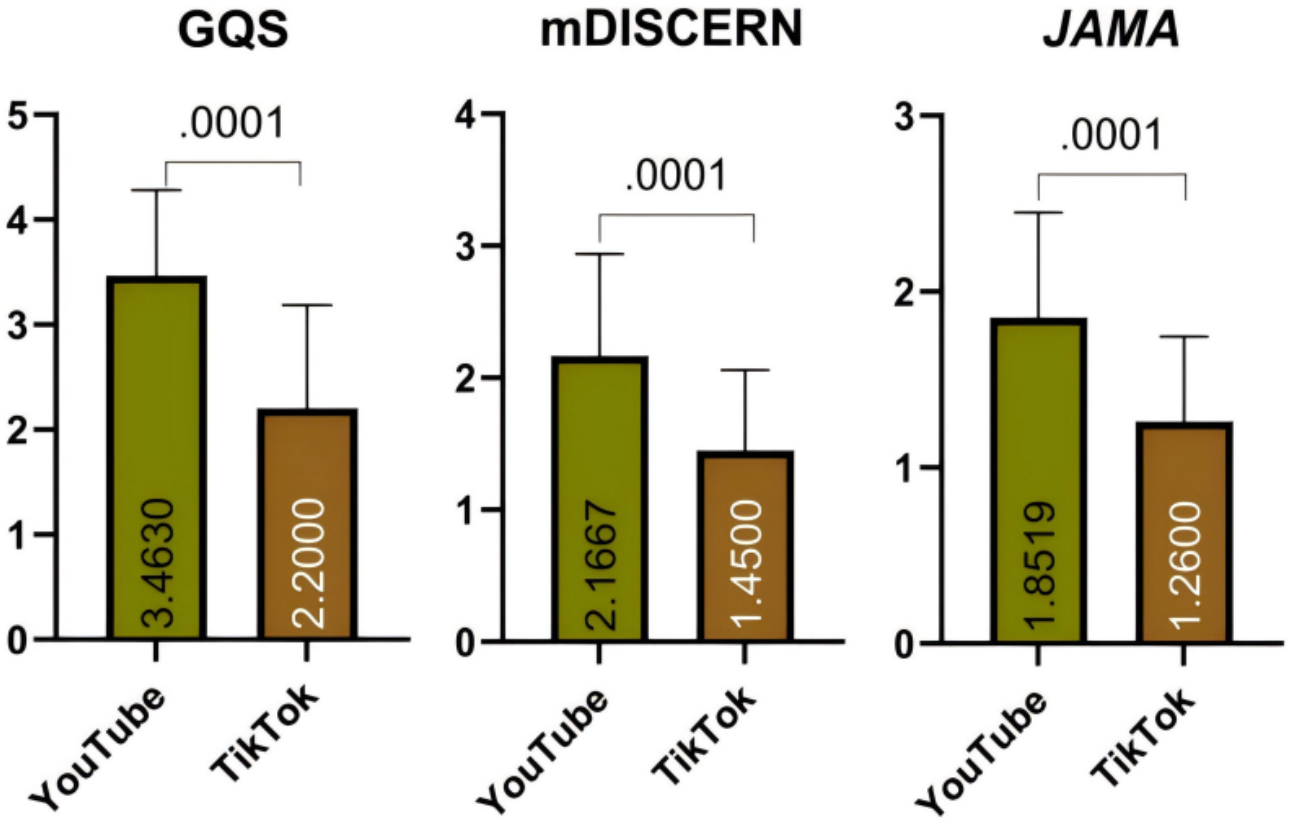
Table . Basic characteristics of the videos.

Information	TikTok	YouTube	*P* value
Likes (IQR)	642 (89.25-4538)	262.5 (27.25-1493)	.98
Comments (IQR)	44 (8.25-291)	9.5 (1.00-52.75)	.19
Favorites (IQR)	87.5 (14.25-503.8)	Not applicable	.004
Views (IQR)	75,350 (10475-619,275)	15,693 (2584-72,514)	.49
Followers (IQR)	11,500 (1314-93,870)	38,950 (3153-343,500)	.23
Duration (s) (IQR)	38.5 (25-74.75)	78.5 (45.25-346)	<.001
Days since posting (IQR)	403 (163.5-692)	483 (258.8-867)	.09
Global Quality Scale score (IQR)	2 (1-3）	3 (3-4）	<.001
Modified DISCERN score (IQR)	1 (1-2）	2 (2-3）	<.001
Journal of the American Medical Association score (IQR)	1 (1-1）	2 (1-2）	<.001

Quality and Reliability Scores

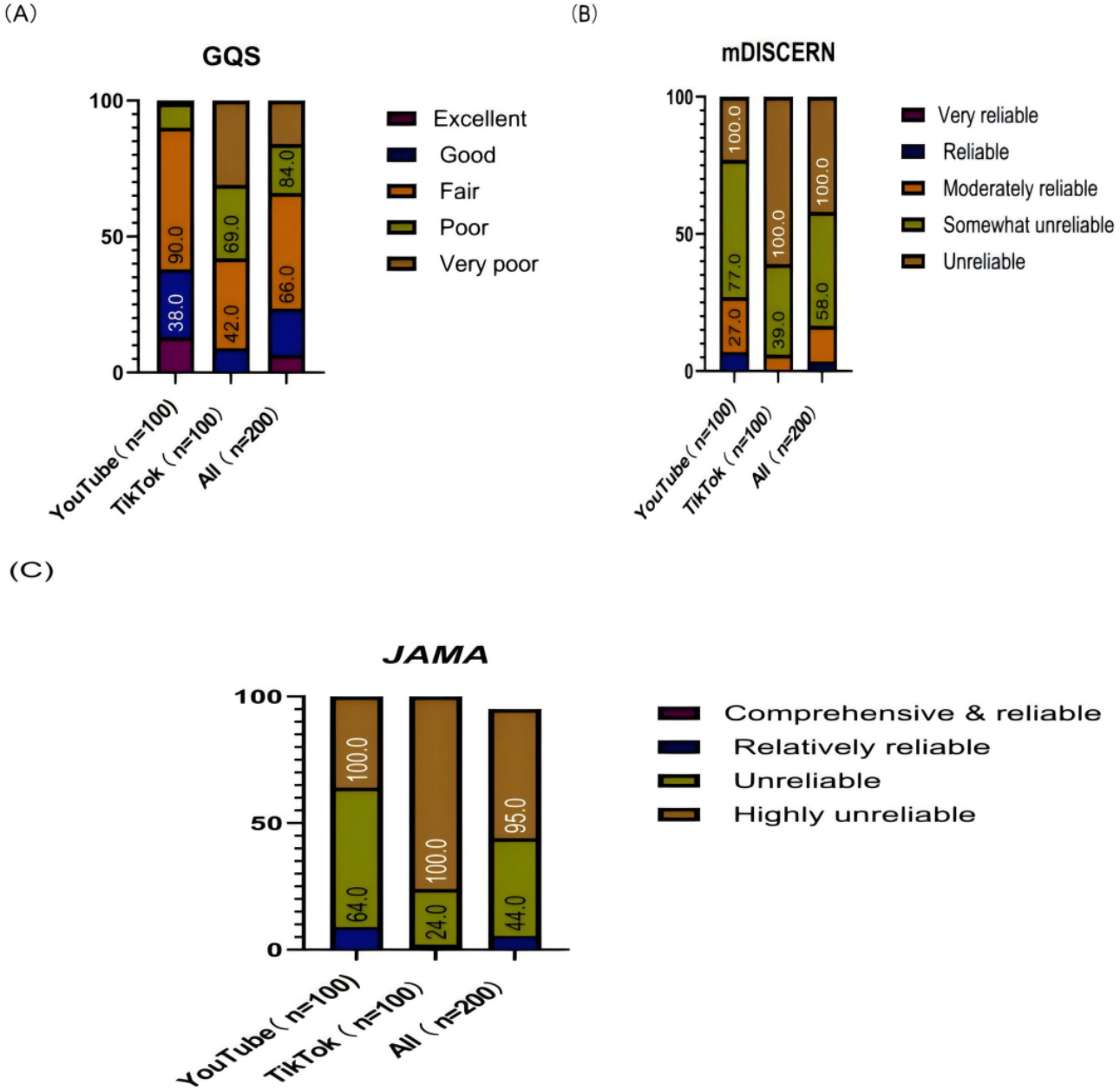
Across all three evaluation tools, YouTube videos demonstrated significantly higher quality and reliability than those on TikTok (Figure 1). The median GQS score was 3 (IQR 2‐4) for YouTube and 2 (IQR 1‐3) for TikTok (*P*<.001). Similarly, the median mDISCERN score was 2 (IQR 1‐3) for YouTube and 1 (IQR 0‐2) for TikTok (*P*<.001). *JAMA* benchmark scores followed the same pattern, with median scores of 2 for YouTube and 1 for TikTok (*P*<.001). Overall, while YouTube content showed better educational quality, videos from both platforms were generally of moderate to low reliability (Figure 2).

Figure 1. Comparison of video quality and reliability scores between YouTube and TikTok. Box plots show median values and interquartile ranges for Global Quality Scale (GQS), modified DISCERN (mDISCERN), and *Journal of the American Medical Association* (*JAMA*) benchmark scores. YouTube videos demonstrated significantly higher scores across all three instruments (all *P*<.001).



XSL•FO
RenderX

Figure 2. Overall distribution of video quality and reliability levels across platforms. Although YouTube videos showed higher scores than TikTok videos, the majority of videos on both platforms were of moderate to low quality and reliability. GQS: Global Quality Scale; *JAMA*: *Journal of the American Medical Association*; mDISCERN: modified DISCERN.



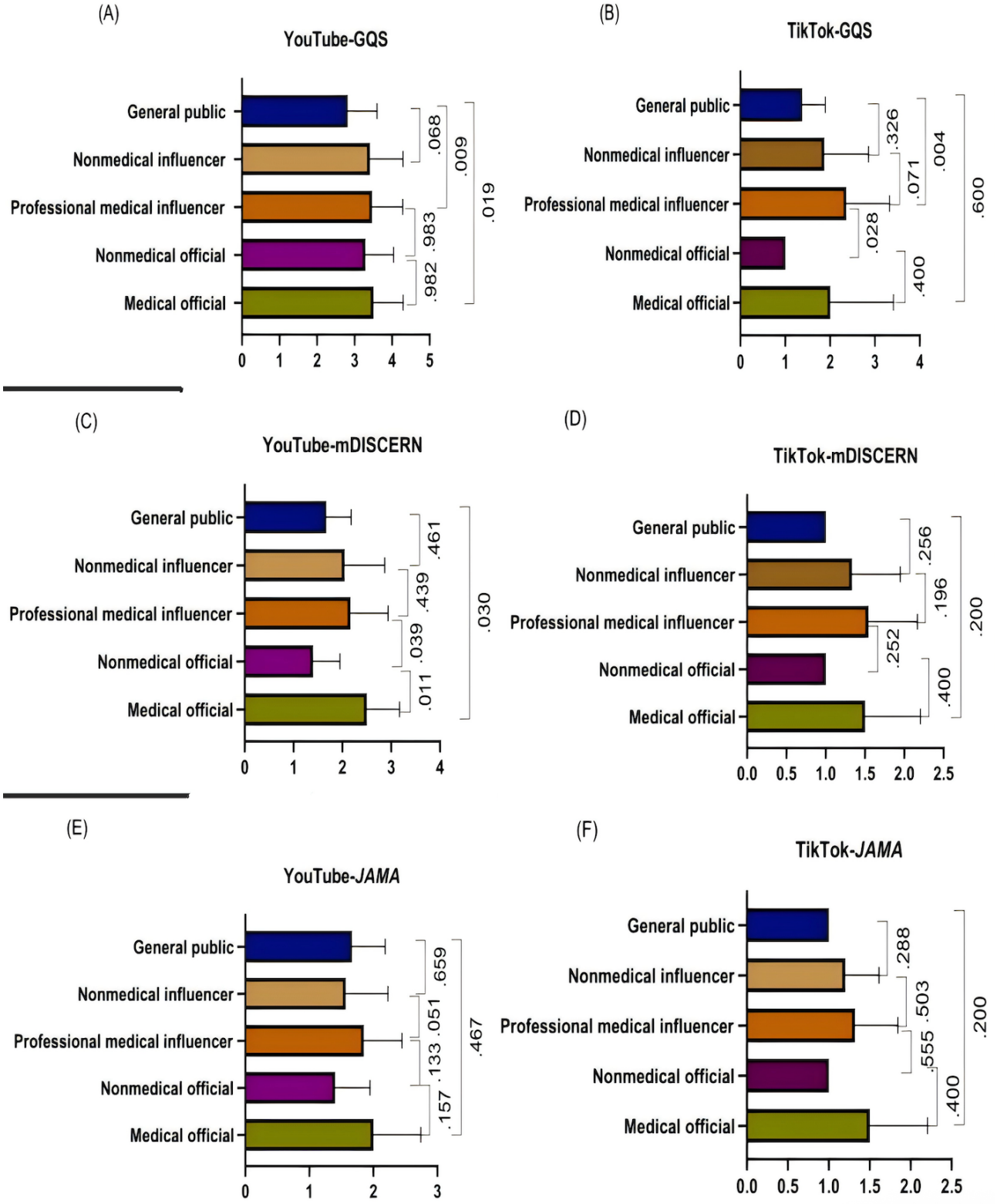
Quality and Reliability by Source

The quality and reliability of videos varied considerably based on their source (Figure 3). On YouTube, videos from official medical organizations (n=12, 12%) achieved the highest median scores across all instruments, followed by content produced by professional medical creators (n=54, 54%). In contrast, TikTok content was predominantly produced by professional medical creators (n=72, 72%), who also achieved the highest quality scores on that platform. There was no significant difference in *JAMA* scores between platforms (P>.05). These findings highlight the role of professional oversight and expertise in producing trustworthy educational content. Differences in thematic emphasis were also observed across account types. Videos uploaded by official medical organizations and health care professionals more commonly focused on prevention, vaccination, wound management, and treatment-related information. These videos were generally more structured and were more likely to include evidence-based recommendations or references to clinical care.

In contrast, videos posted by nonmedical creators more often emphasized brief symptom descriptions, personal experiences, or general awareness messages, with less detailed discussion of prevention or management. Simplified explanations and omission of key clinical information were more frequently observed in these videos. On TikTok in particular, short-form videos from noninstitutional accounts tended to prioritize concise delivery and audience engagement over informational depth.

XSL•FO
RenderX

Figure 3. Quality and reliability scores stratified by source of video upload. Videos from official medical organizations and professional medical creators demonstrated higher Global Quality Scale (GQS), modified DISCERN (mDISCERN), and *Journal of the American Medical Association* (*JAMA*) benchmark scores compared with other sources. Distribution patterns differed between YouTube and TikTok.



Thematic Content Characteristics of Tetanus-Related Videos

Intercoder agreement for thematic classification was substantial to almost perfect across the predefined thematic categories. Cohen κ values were 0.703 for disease overview, 0.826 for symptoms, 0.831 for prevention, 0.887 for treatment, 0.908 for wound management, and 0.858 for misinformation or incomplete information.

Thematic content analysis revealed significant differences in the distribution of informational components between platforms (Table 2). Videos on YouTube more frequently addressed

XSL•FO
RenderX

symptoms (n=92, 92% vs n=81, 81%, *P*=.02), prevention (n=95, 95% vs n=78, 78%, *P*<.001), treatment (n=88, 88% vs n=70, 70%, *P*=.002), and wound management (n=77, 77% vs n=38, 38%, *P*<.001) compared to TikTok videos. In contrast, there were no statistically significant differences between platforms in the inclusion of general disease overview (n=95, 95% vs n=93, 93%, *P*=.55) or the presence of misinformation or incomplete information (n=9, 9% vs n=12, 12%, *P*=.50).

Table . Thematic content characteristics of tetanus-related videos on YouTube and TikTok (N=200).

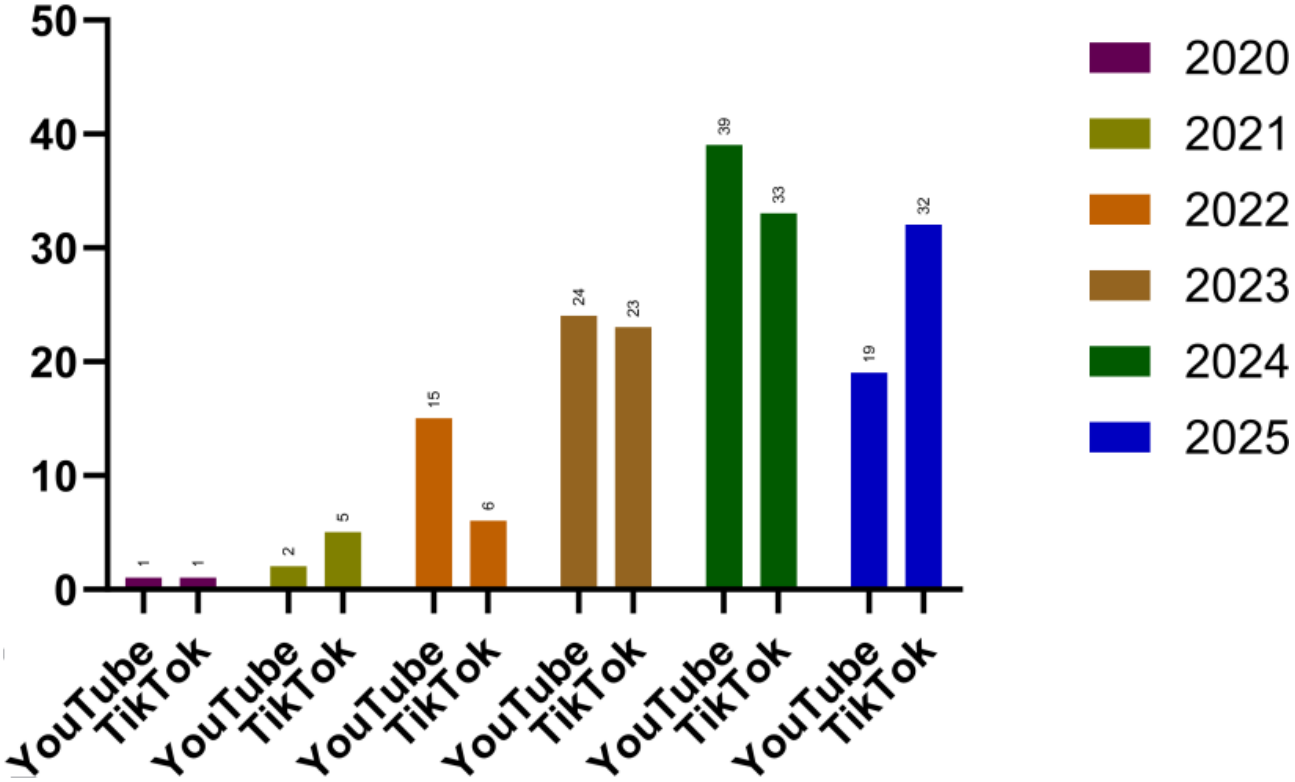
Characteristic	YouTube (n=100), n (%)	TikTok (n=100), n (%)	*P* value
Disease overview	95 (95)	93 (93)	.55
Symptoms	92 (92)	81 (81)	.02
Prevention	95 (95)	78 (78)	<.001
Treatment	88 (88)	70 (70)	.002
Wound management	77 (77)	38 (38)	<.001
Misinformation or incomplete information	9 (9)	12 (12)	.50

To complement these findings, qualitative review identified several recurring patterns in lower-quality videos. Some videos presented overly simplified or incomplete explanations of tetanus, particularly regarding symptoms and prevention. For instance, wound care was sometimes described as sufficient on its own, without reference to vaccination as the primary preventive measure. In other cases, symptom descriptions were vague or imprecise, such as attributing muscle stiffness to general fatigue without explaining its neurological basis. In addition, missing source attribution and the absence of clear, actionable guidance were common. These features were observed on both platforms and were more frequently seen in videos with lower-quality scores.

Publication Timeline and Account Verification

Videos analyzed were uploaded between 2020 and 2025, with the majority posted in 2024 (YouTube: n=39, 39%; TikTok: n=33, 33%) (Figure 4). Account verification was more common on YouTube (n=20, 20%) than on TikTok (n=3, 3%), suggesting differences in platform moderation and credibility indicators.

Figure 4. Distribution of tetanus-related videos by year (2020–2025) on YouTube and TikTok. Most videos were uploaded in 2024 on both platforms.



Geographical Distribution of Video Sources

The geographic distribution of video sources varied across platforms. On YouTube, most videos originated from India (n=40, 40%) and the United States (n=31, 31%), followed by Indonesia (n=4, 4%) and Pakistan (n=4, 4%). On TikTok, the distribution was more dispersed, with India (n=19, 19%) and the United States (n=15, 15%) representing the largest shares. When excluding videos with unspecified origins, India, the

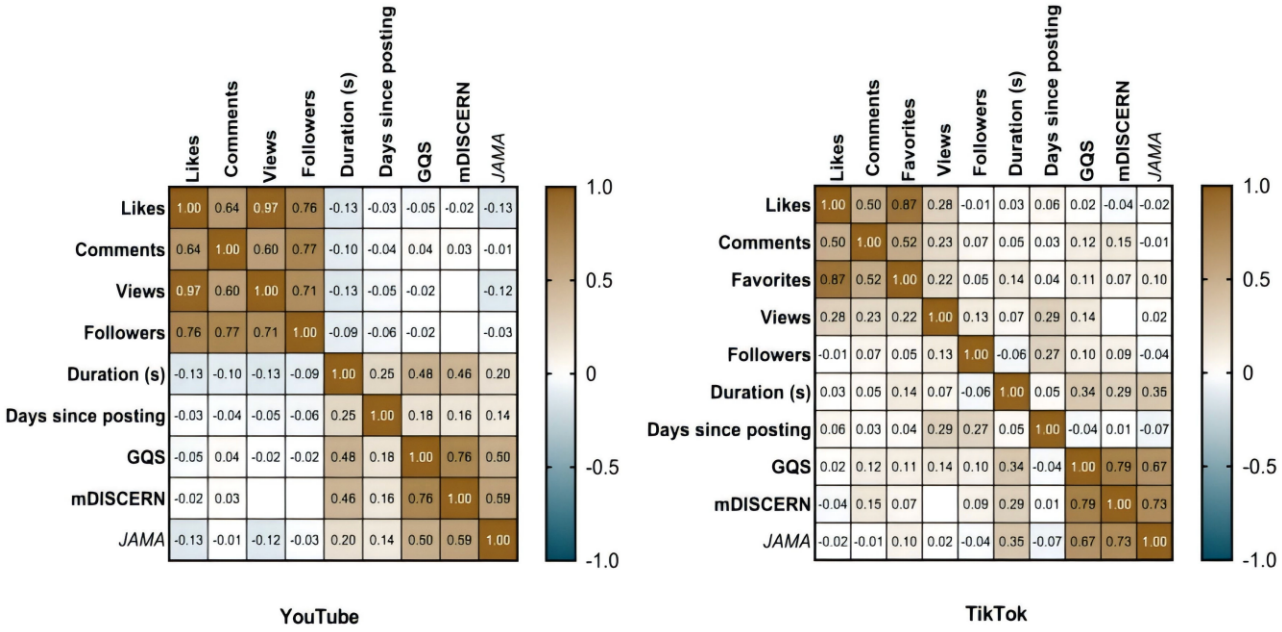
XSL•FO
RenderX

United States, and Indonesia accounted for the majority of content on both platforms.

Engagement Metrics and Correlation Analysis

Engagement metrics showed strong interrelationships, particularly between likes and views on YouTube (r=0.97, P<.001) and between likes and favorites on TikTok (r=0.87). However, the correlation between engagement indicators and quality scores was moderate, suggesting that highly interactive content does not always align with higher educational value. For example, on YouTube, GQS scores correlated moderately with mDISCERN (r=0.76) and *JAMA* (r=0.50) scores, while on TikTok, GQS correlated strongly with mDISCERN (r=0.79) and *JAMA* (r=0.67) scores. Notably, video duration showed only a weak positive correlation with quality scores (YouTube: r=0.48; TikTok: r=0.34), indicating that while longer videos may provide more comprehensive information, length alone does not guarantee higher educational quality (Figure 5).

Figure 5. Correlations between engagement metrics and quality scores. Scatter plots show the relationships between engagement indicators (likes, views, favorites) and GQS, mDISCERN, and *JAMA* scores. Engagement metrics were strongly correlated with each other, whereas their correlations with quality scores were weaker. GQS: Global Quality Scale; *JAMA*: *Journal of the American Medical Association*; mDISCERN: modified DISCERN.



Discussion

Principal Findings

From an infodemiology perspective, this study provides a systematic evaluation of how health information is generated, disseminated, and consumed on social media platforms, highlighting key challenges in digital health communication. This study evaluated tetanus-related videos on YouTube and TikTok in terms of quality, reliability, and content characteristics. Across all evaluation tools, YouTube videos had higher scores than TikTok videos. However, the overall quality on both platforms remained limited, with most videos falling within the low-to-moderate range.

Platform Differences in Quality and Video Characteristics

Differences between platforms may be partly related to video format. In this study, YouTube videos were generally longer, which allows more space to explain clinical concepts. In contrast, TikTok videos were shorter and more condensed, which may restrict the amount of information presented.

Thematic Content Differences

Thematic analysis showed clear differences in content coverage. YouTube videos more frequently included key clinical topics such as symptoms, prevention, treatment, and wound management. These elements are essential for tetanus prevention and early management. TikTok videos less often covered these components, particularly wound management, where the gap between platforms was most pronounced. Despite these differences, the proportion of videos containing misinformation or incomplete information did not differ significantly, suggesting that accuracy issues exist on both platforms.

Implications of Low-Quality Content

Inaccurate or vague descriptions of symptoms may further reduce the educational value of these videos, especially for conditions such as tetanus where early recognition is important. In addition, the lack of source attribution and absence of clear recommendations suggest limited transparency and may undermine the credibility of the information presented.

Taken together, these findings indicate that the challenge extends beyond overt misinformation to include the omission of essential clinical details. Similar patterns have been reported in previous evaluations of online health information, where incomplete or poorly contextualized content reduced overall reliability [20]. This may partly reflect the constraints of short-form content formats and the tendency to prioritize simplicity and engagement over informational depth [21].

XSL•FO
RenderX

Improving the clarity, accuracy, and completeness of key health messages should therefore be a priority for enhancing the effectiveness of social media–based health communication [22].

Role of Content Source

Videos produced by health care professionals and medical organizations tended to have higher quality scores. This pattern has been reported in previous studies and reflects the importance of professional involvement [23]. At the same time, variation within this group indicates that expertise alone does not guarantee high-quality communication, especially when content is adapted for social media formats. Thematic differences across account types may also help explain the variability in video quality observed in this study. Videos produced by health care professionals or institutional accounts more frequently focused on prevention, vaccination, and wound management, whereas nonmedical creators tended to emphasize brief symptom descriptions or general awareness content [24].

This pattern suggests that professional background influences not only overall information quality but also the depth and focus of health communication. On short-form platforms in particular, content creators may prioritize concise delivery and audience engagement, which can result in the omission of clinically important details.

Engagement and Information Quality

Engagement indicators such as likes and views were strongly correlated with each other but only moderately associated with quality scores. This suggests that widely viewed content is not necessarily more reliable. The discrepancy between popularity and informational value may contribute to the spread of incomplete or oversimplified health information.

Comparison With Previous Studies

These findings are consistent with prior research on online health information [25,26]. Studies evaluating YouTube content across different medical topics have reported substantial variability in quality and frequent gaps in completeness and reliability [27,28]. Systematic reviews have similarly shown that health-related videos often lack key clinical information despite high visibility [20]. Comparative studies have also suggested that short-form video platforms may further limit informational depth [29,30]. The present study extends these observations to tetanus-related content, particularly in areas such as prevention and wound management. These findings contribute directly to the growing field of infodemiology by identifying gaps in the quality and dissemination of online health information, with implications for improving digital health communication strategies. Similar concerns regarding the readability and quality of online health information have also been reported in non-English settings, including Turkish-language patient education materials [31], suggesting that variability in accessibility and reliability may represent a broader international challenge in digital health communication.

In recent years, artificial intelligence (AI)–assisted information systems and conversational chatbots have increasingly been used as sources of health information. Previous studies have shown that AI-generated medical content may improve accessibility and user engagement, although concerns remain regarding readability, reliability, and clinical accuracy [32]. Compared with AI-generated responses, social media videos may provide more intuitive and visually accessible information, but they can also contain oversimplified or incomplete medical explanations. These issues suggest that maintaining the quality and consistency of online health information remains challenging across different digital platforms.

Public Health Implications

Tetanus is a preventable disease, and accurate information on vaccination and wound care is essential [33]. Social media platforms provide broad access to such information [34,35], but the variability in quality observed in this study indicates that current content may not fully support effective prevention.

Improving content quality may require greater involvement from health care professionals, clearer credibility indicators, and platform-level strategies to promote reliable information [36]. Adapting content to platform characteristics while maintaining accuracy is also important.

Strengths

This study evaluated tetanus-related videos on both YouTube and TikTok using multiple validated assessment tools, including GQS, mDISCERN, and the *JAMA* benchmark criteria. By combining quality evaluation with thematic analysis and engagement metrics, the study provides a broader view of how tetanus-related health information is presented and disseminated on social media platforms. In addition, the inclusion of two widely used video platforms improves the practical relevance of the findings within current digital health communication settings.

Limitations

Several limitations should be acknowledged. First, only English-language videos were included, which may limit the generalizability of the findings to other linguistic and cultural contexts. Second, data collection was performed at a single time point, whereas social media content and platform algorithms may change over time. Although newly created accounts and search-history controls were used to reduce personalization bias, algorithmic influence on search results could not be completely eliminated. Finally, because of the cross-sectional design, the study was unable to evaluate causal relationships between video characteristics, user engagement, and information quality.

Conclusions

YouTube generally provided more comprehensive and reliable tetanus-related information than TikTok, although the overall quality of videos on both platforms remained variable. Videos produced by health care professionals or institutional accounts were more likely to include information on prevention, vaccination, treatment, and wound management. Highly viewed videos were not always associated with better informational quality.

Many videos contained incomplete explanations or lacked important clinical details, which may limit their educational value for the public. Given that tetanus is largely preventable, improving the clarity, accuracy, and transparency of online

XSL•FO
RenderX

health information remains important. Greater participation from qualified health care professionals and improved content oversight may help improve the quality of health communication on social media platforms.

Although AI-assisted tools may improve access to health information, they should currently be regarded as supplementary resources rather than replacements for professional medical advice. Care should be taken when interpreting online medical information that may be incomplete, oversimplified, or insufficiently evidence-based.

Acknowledgments

The authors would like to thank all reviewers and editors for their valuable comments and constructive suggestions, which greatly helped improve the quality and clarity of this manuscript. We also acknowledge the public availability of video content on YouTube and TikTok that made this infodemiology study possible. Finally, we thank the independent reviewers involved in the video evaluation process for their careful assessments and consensus discussions.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The videos included in this study were collected and stored in the favorites lists of newly created research accounts on YouTube and TikTok to reduce personalization bias during data collection. The extracted raw dataset used for analysis has been provided as a supplementary file in Multimedia Appendix 1. Additional data are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: RH, XC.
Data curation: YS, RH.
Formal analysis: RH, YS, JZ.
Funding acquisition: XC.
Investigation: RH, YS, XD, QL.
Methodology: RH, GZ, JZ.
Project administration: XC.
Supervision: JZ, XC.
Writing – original draft: RH, YS.
Writing – review & editing: RH, YS, GZ, JZ, XD, QL, XC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Raw data.
[XLSX File, 22 KB - infodemiology_v6i1e85397_app1.xlsx]

References

1. Cook TM, Protheroe RT, Handel JM. Tetanus: a review of the literature. Br J Anaesth 2001 Sep;87(3):477-487. [doi: 10.1093/bja/87.3.477] [Medline: 11517134]
2. Thwaites CL, Thanh TT, Ny NTH, et al. Seroprotection against tetanus in southern Vietnam. Vaccine (Auckl) 2023 Mar 24;41(13):2208-2213. [doi: 10.1016/j.vaccine.2023.02.036] [Medline: 36849339]
3. Sudarshan R, Sayo AR, Renner DR, et al. Tetanus: recognition and management. Lancet Infect Dis 2025 Nov;25(11):e645-e657. [doi: 10.1016/S1473-3099(25)00292-0] [Medline: 40543524]
4. Tetanus reported cases and incidence. World Health Organization. URL: https://immunizationdata.who.int/global/wiise-detail-page/tetanusreported-cases-and-incidence [accessed 2024-11-19]
5. Kyu HH, Mumford JE, Stanaway JD, et al. Mortality from tetanus between 1990 and 2015: findings from the global burden of disease study 2015. BMC Public Health 2017 Feb 8;17(1):179. [doi: 10.1186/s12889-017-4111-4] [Medline: 28178973]
6. Basch CH, Basch CE, Hillyer GC, Reeves R. YouTube Videos related to skin cancer: a missed opportunity for cancer prevention and control. JMIR Cancer 2015 Mar 2;1(1):e1. [doi: 10.2196/cancer.4204] [Medline: 28410167]

XSL•FO
RenderX

7. Choudhary OA, Dhawan J, Sohrabipour S, et al. Assessment of online YouTube videos as a source of information and instruction for pulmonary rehabilitation. Respir Med 2025 Mar;238:107968. [doi: 10.1016/j.rmed.2025.107968] [Medline: 39884588]
8. Ozduran E, Büyükçoban S. A content analysis of the reliability and quality of YouTube videos as a source of information on health-related post-COVID pain. PeerJ 2022;10:e14089. [doi: 10.7717/peerj.14089] [Medline: 36193427]
9. Nißen M, Harperink S, Joshi P, et al. Leveraging influencers to reach and engage vulnerable individuals with a digital health intervention: quasi-experimental field study. J Med Internet Res 2025 Sep 5;27:e67174. [doi: 10.2196/67174] [Medline: 40911352]
10. Bagcier F, Yurdakul OV, Ozduran E. Top 100 cited articles on ankylosing spondylitis. Reumatismo 2021 Jan 18;72(4):218-227. [doi: 10.4081/reumatismo.2020.1325] [Medline: 33677949]
11. Ozduran E, Hanci V, Erkin Y. Evaluating the readability, quality and reliability of online patient education materials on chronic low back pain. Natl Med J India 2024;37(3):124-130. [doi: 10.25259/NMJI_327_2022] [Medline: 39399994]
12. Wang M, Yao N, Wang J, Chen W, Ouyang Y, Xie C. Bilibili, TikTok, and YouTube as sources of information on gastric cancer: assessment and analysis of the content and quality. BMC Public Health 2024 Jan 2;24(1):57. [doi: 10.1186/s12889-023-17323-x] [Medline: 38166928]
13. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007 Sep;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
14. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
15. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111. [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
16. Libretti A, Vitale SG, Saponara S, et al. Hysteroscopy in the new media: quality and reliability analysis of hysteroscopy procedures on YouTube. Arch Gynecol Obstet 2023 Nov;308(5):1515-1524. [doi: 10.1007/s00404-023-07172-9] [Medline: 37568070]
17. Zhou Y, Yang L, Luo L, Cao L, Qiu J. Quality assessment of videos about dengue fever on Douyin: cross-sectional study. JMIR Infodemiology 2025 Sep 26;5:e76474. [doi: 10.2196/76474] [Medline: 41004281]
18. Szmuda T, Syed MT, Singh A, Ali S, Özdemir C, Słoniewski P. YouTube as a source of patient information for coronavirus disease (COVID-19): a content-quality and audience engagement analysis. Rev Med Virol 2020 Sep;30(5):e2132. [doi: 10.1002/rmv.2132] [Medline: 32537771]
19. Loeb S, Sengupta S, Butaney M, et al. Dissemination of misinformative and biased information about prostate cancer on YouTube. Eur Urol 2019 Apr;75(4):564-567. [doi: 10.1016/j.eururo.2018.10.056] [Medline: 30502104]
20. Birch EM, Leziak K, Jackson J, Dahl E, Niznik CM, Yee LM. Content quality of YouTube videos about gestational diabetes: systematic evaluation. JMIR Diabetes 2022 Apr 7;7(2):e30156. [doi: 10.2196/30156] [Medline: 35389355]
21. İlhan S, Evran T. The quality and reliability of YouTube and TikTok videos on epidural blood patch: a cross-sectional analysis. Medicine (Baltimore) 2025 Aug 1;104(31):e43628. [doi: 10.1097/MD.0000000000043628] [Medline: 40760581]
22. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552. [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
23. Stimpson JP, Park S, Adhikari EH, Nelson DB, Ortega AN. Perceived health misinformation on social media and public trust in health care. Med Care 2025 Sep 1;63(9):686-693. [doi: 10.1097/MLR.0000000000002180] [Medline: 40793916]
24. Meyer M, Shoukfeh R, Turk D, Potts G. Misinformation about Monkeypox on TikTok. Skinmed 2023;21(4):248-251. [Medline: 37771014]
25. Toprak T, Tokat E. A quality analysis of nocturnal enuresis videos on YouTube. J Pediatr Urol 2021 Aug;17(4):449. [doi: 10.1016/j.jpurol.2021.03.014] [Medline: 33824069]
26. Travis LM, Jahn J, Dean KK, Dodds SD. Evaluating the educational quality and accuracy of carpal tunnel-and cubital tunnel-related videos on TikTok. Hand (N Y) 2025 Sep 7:15589447251366458. [doi: 10.1177/15589447251366458] [Medline: 40914853]
27. Yagiz B, Coskun BN, Halil EY, Dalkilic E, Pehlivan Y. The efficacy and reliability of English YouTube videos as a source of information for pregnant rheumatoid arthritis patients. Clin Rheumatol 2023 Dec;42(12):3311-3320. [doi: 10.1007/s10067-023-06780-9] [Medline: 37814092]
28. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ 2022 May 19;22(1):382. [doi: 10.1186/s12909-022-03446-z] [Medline: 35590410]
29. Alvarado Garcia A, Rahman L, Gregoski MJ, Andrade I. Comparison of patient testimonials of orthognathic surgery on YouTube and TikTok. Angle Orthod 2025 Mar 1;95(2):199-204. [doi: 10.2319/071724-571.1] [Medline: 39880020]
30. Özkent MS, Kılınç MT. Female urinary incontinence on TikTok and YouTube: is online video content sufficient? Int Urogynecol J 2023 Nov;34(11):2775-2781. [doi: 10.1007/s00192-023-05607-0] [Medline: 37470796]

XSL•FO
RenderX

31. Sargın F, Alçı M, Kaygusuz Aydemir B, Şahin S, Çaksen H, Güven AS. Evaluation of the readability and quality of online Turkish materials in the field of pediatric neurology. Turk Arch Pediatr 2025 Dec 29;61(2):147-151. [doi: 10.5152/TurkArchPediatr.2025.25337] [Medline: 42044431]
32. Kara M, Ozduran E, Kara MM, Özbek İ, Hancı V. Evaluating the readability, quality, and reliability of responses generated by ChatGPT, Gemini, and Perplexity on the most commonly asked questions about ankylosing spondylitis. PLoS One 2025;20(6):e0326351. [doi: 10.1371/journal.pone.0326351] [Medline: 40531978]
33. Lamberto Y, Vargas C, Sanchez Cunto M, Saúl P, Chediack V, Cunto E. [Tetanus: an immunopreventable disease] [Article in Spanish]. Medicina (B Aires) 2023;83(5):841-845. [Medline: 37870348]
34. Zambrano Rodriguez VC, Bellur S, Christensen JL. Health information seeking via social media: a study of college students using the MAIN model. J Am Coll Health 2026 Mar;74(3):634-647. [doi: 10.1080/07448481.2025.2542411] [Medline: 40902010]
35. Menakaya NC, Unuigbe EI. Health-related social media misinformation: implications for the burden of noncommunicable diseases in sub-Saharan Africa. West Afr J Med 2025 Apr 30;42(4):321-329. [Medline: 40986974]
36. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16(1):442. [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]

Abbreviations

AI: artificial intelligence
GQS: Global Quality Scale
***JAMA*:** *Journal of the American Medical Association*
mDISCERN: modified DISCERN
WHO: World Health Organization

Edited by T Mackey; submitted 07.Oct.2025; peer-reviewed by A Murcko, E Ozduran; revised version received 26.May.2026; accepted 26.May.2026; published 24.Jun.2026.

Please cite as:
Huang R, Shen Y, Zou G, Zhou J, Deng X, Li Q, Chen X
Tetanus Health Information on YouTube and TikTok: Cross-Sectional Analysis of Quality, Reliability, and Public Health Implications
JMIR Infodemiology 2026;6:e85397
URL: https://infodemiology.jmir.org/2026/1/e85397
doi:10.2196/85397

© Rui Huang, Yajing Shen, Guangqing Zou, Jun Zhou, Xiaolu Deng, Qian Li, Xiaoxiong Chen. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 24.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Understanding Vaccine Hesitancy in Louisiana Through Social Media Listening and Community Feedback: Cross-Sectional Study

Angie Viviana Sanchez[1,2], MD, MPH; Sadie Smith[1], MPH; Sahithya Sakhamuri[1], MPH; Julia Ardoin[1], BA; Henry Chu[3,4], PhD

[1]Louisiana Center for Health Innovation (LCHI), University of Louisiana at Lafayette, 635 Cajundome Blvd, Lafayette, LA, United States

[2]Brigham And Women's Hospital, 75 Francis Street, Boston, MA, United States

[3]School of Computing and Informatics, University of Louisiana, Lafayette, LA, United States

[4]Informatics Research Institute (IRI), University of Louisiana, Lafayette, LA, United States

Corresponding Author:
Angie Viviana Sanchez, MD, MPH
Louisiana Center for Health Innovation (LCHI), University of Louisiana at Lafayette, 635 Cajundome Blvd, Lafayette, LA, United States

Abstract

Background: The rise of social media has significantly impacted public health programs, with platforms such as YouTube, Facebook, X (formerly known as Twitter), Instagram, and, more recently, TikTok being used to promote health information, raise awareness about disease outbreaks, and support disease prevention programs. However, the diverse and often unverified nature of the content on social media can make it challenging to discern accurate information, contributing to user uncertainty, which may in turn contribute to low vaccination rates in some regions. This is especially true in Louisiana as its COVID-19 vaccination rates were among the lowest in the country in 2022. Therefore, understanding public sentiment on social media and developing targeted campaigns to counter unverified information is essential for advancing public health campaigns.

Objective: The goal was to gain insights into the underlying factors that contribute to Louisiana's low vaccination rates for routine immunizations by (1) performing social media listening to develop an infodemic management plan and (2) promoting accurate information via a social media campaign.

Methods: Social media listening was conducted using Meltwater, a media monitoring and social media listening platform, supplemented by Google Alerts and Google News to identify if vaccine-related stories or sentiments were attracting unusual attention. Additionally, a social media campaign aimed at educating Louisiana residents about disease manifestation, symptoms, vaccines available for disease prevention, and potential side effects was developed. Posts were published 2 to 3 times a week and boosted for 7 days.

Results: From November 13, 2023, to June 11, 2024, social media listening identified at least 15 unique, noteworthy stories that signified sentiment spikes. These conversations were predominantly related to vaccine hesitancy, with users expressing opposition to vaccines or reluctance to engaging with vaccine-related information. Sentiment spikes included themes related to mistrust of vaccines and concerns about their safety and efficacy. The social media campaign received 69,600 impressions, reached 43,429 users, and received 652 reactions and likes, 62 shares, and 105 comments. Most of the audience was female, with higher engagement from older users on Facebook and younger users on Instagram. Finally, posts related to hepatitis B, rotavirus, and measles, mumps, and rubella vaccines received the most attention.

Conclusions: Social media has become a key tool for digital health, helping to implement disease prevention programs and promoting advances in medicine. However, unverified information remains a major reason for the aversion to vaccination despite the dissemination of information from reputable public health organizations, health professionals, hospitals, and medical centers. To address this, information that is accessible, understandable, and culturally competent must be circulated to mitigate disinformation and improve attitudes toward vaccination. More research is needed to evaluate the effectiveness of social media campaigns in reducing vaccine hesitancy and improving willingness to adopt public health recommendations to increase vaccination rates.

(JMIR Infodemiology 2026;6:e76827) doi:10.2196/76827

KEYWORDS

vaccines; vaccine sentiments; vaccine hesitancy; social media listening; public health campaigns; Louisiana

XSL•FO
RenderX

Introduction

Since its popularization in the early 2000s, social media has evolved into a multifaceted tool for multiple uses: socializing with friends, family, and the public; professional networking; and accessing information in conjunction with traditional news media. Presently, social media has penetrated all generations as it offers a variety of options, including engaging with videos, entertainment, news, opinion forums, and more. Platforms such as YouTube, Facebook, X (formerly known as Twitter), Instagram, and, recently, TikTok have been used not only as spaces for connection and community but also as a means of disseminating information [1-3].

The emergence of new social media platforms and applications has substantially impacted public health strategies. Health care providers, public health professionals, and other organizations began using social media for health education campaigns to issue timely information about disease outbreaks and inform target audiences about varying health topics [4,5]. Furthermore, with advancements in medical technology, these platforms have become an increasingly fundamental part of medical research. Researchers now use social media to gain insights into public attitudes, such as trust in medical personnel, acceptance of new treatments or prevention mechanisms (eg, vaccines), and hesitancy toward adopting public health recommendations [5,6].

While skepticism and doubt toward vaccinations have been present and influential throughout history [7-9], the emergence of social media has amplified these challenges [3]. As an easily accessible, user-driven information channel, social media enables a rapid flow of content that can be overwhelming, confusing, and anxiety inducing [2,6]. The abundance of accurate, false, and contradictory information being shared creates a breeding ground for misinformation [1,3,10]. For example, during the COVID-19 pandemic, the rapid spread of unverified information on social media generated high levels of confusion, anxiety, and mainstream rejection of vaccines and prevention guidelines disseminated by health personnel and public health organizations [11-13]. Although distrust is not a new barrier in public health, the rapid spread of inaccurate information on social media has become a contributing factor to low vaccination rates in some regions [14-17]. For instance, in 2022, only 56% of Louisiana's population was vaccinated against COVID-19, ranking it as the state with the fourth lowest vaccination rate in the United States [18,19]. This gap is especially evident when comparing Louisiana to other states in the United States. In 2023, states such as Rhode Island, Massachusetts, the District of Columbia, and Hawaii each reported that 80% or more of their population were fully vaccinated against COVID-19, with over 38% having received a booster dose. In contrast, less than 1% of Louisiana's population had received a booster dose the same year [19].

Vaccination gaps extend past COVID-19 and are present in routine childhood and influenza immunizations. These disparities often vary by state. For example, during the 2022 to 2023 influenza season, only 41.1% of Louisiana's residents were vaccinated, falling below the national average of 46.9%, whereas states such as Massachusetts reached 62.4% [20]. Similarly, uptake of the measles, mumps, and rubella (MMR) vaccine, which is part of routine childhood vaccinations, declined by almost 30% among children with Medicaid in Louisiana in 2020 compared to previous years, perhaps as a consequence of the COVID-19 pandemic [21]. Additionally, racial gaps in childhood vaccination rates persisted at the state level from 2019 to 2021. In Louisiana, African American children were almost 12% less likely to have received the full schedule of recommended vaccinations compared to White children [22]. Additionally, 76.1% of privately insured children aged 35 months received all the recommended vaccines in 2020 to 2021 compared to 57.8% of children with Medicaid. In 2021, Louisiana ranked fifth lowest in childhood vaccination coverage with a rate of 62% compared to states such as Iowa, Massachusetts, Connecticut, Vermont, and North Dakota, all of which reported vaccination rates exceeding 80% [22].

With guidance from the World Health Organization's (WHO) Infodemic Management training and the social media listening and monitoring tools from the US Department of Health and Human Services Centers for Disease Control and Prevention (CDC), this study aimed to gain insights into the factors that contribute to Louisiana's low vaccination rates [23,24]. Using these resources, a social media listening plan was developed to analyze circulating conversations about vaccines. Using the insights from social media listening, a social media educational campaign was developed. We hypothesized that this campaign could help reduce erroneous information regarding vaccines and promote vaccine confidence. This approach aimed to empower individuals to make informed decisions without fear or false claims of adverse events. This paper describes the search strategies used to perform social media listening, topics covered in the health and vaccination educational campaign, and the resulting levels of user engagement and acceptance.

Methods

Overview

This was a qualitative, cross-sectional study assessing vaccine hesitancy through social media listening and a social media campaign. First, social media listening took place from November 13, 2023, to June 11, 2024. After 5 months, the insights gathered were used to develop a social media educational campaign that was implemented from April 22, 2024, to June 28, 2024.

Ethical Considerations

This study did not undergo institutional review board approval as it used publicly available social media data for the analysis, which were obtained from Facebook, Instagram, X, and Meltwater under their terms of service. Names or other identifiable information were not collected. Instead, the study used age and gender cluster data provided by Meta Analytics for Facebook and Instagram.

Social Media Listening

To understand the online conversation regarding vaccines, Meltwater—a media monitoring and social listening platform—was used to perform social media listening, occasionally supplementing search results using Google Alerts

XSL•FO
RenderX

and Google News depending on the capabilities of social media listening tools. In total, 12 saved searches grouped by topic and 1 saved search limited to Louisiana-based social media users were created. Saved searches were built to accurately capture vaccine sentiment. Each search was constructed using a series of keywords separated by Boolean operators to narrow down the focus of the results. For example, to explore discussions about vaccines related to opinions on the credibility of the CDC and the Food and Drug Administration, the following Meltwater search was built: "('vax' OR 'vaccine*') AND ('Center for Disease Control' OR 'CDC' OR 'FDA' OR 'Food and Drug Administration') AND ('lie*' OR 'credibility' OR 'trust' OR 'conspiracy' OR 'conspiracies' OR 'approval process' OR 'approval transparency' OR 'approval credibility' OR 'Fauci' OR 'World Health Organization' OR 'National Institute of Health')." The search was limited to return results that were in English and from US-based sources. These searches were developed and grouped by the following topics: CDC and Food and Drug Administration credibility, childhood vaccines, COVID-19 in the United States, influenza vaccine in the United States, Louisiana Department of Health, public health agencies (eg, CDC and WHO), respiratory syncytial virus (RSV) vaccine, rural health care, vaccine alternatives, vaccine brands, and vaccine hesitancy: general United States and vaccine hesitancy: Louisiana. All Meltwater saved searches and keywords can be found in Multimedia Appendix 1.

Initial queries were built iteratively with guidance from Meltwater advisors, who provided the best practices for optimizing social media listening searches on their platform. Team members then refined these queries to better align with the study's goal of capturing themes and opinions that offered insights into the general public's feelings and attitudes toward vaccines. Team members attended meetings with Meltwater to review search terms and occasionally edited them to improve the sensitivity and specificity of the searches. Team members sifted through search results daily using Meltwater's tagging feature to save noteworthy results and met biweekly to review saved posts and discuss discrepancies. Sentiment classification was conducted using Meltwater's automated sentiment analysis feature, which categorizes content as positive, neutral, or negative based on algorithms and manually. Team members would first allow Meltwater to apply its automated sentiment analysis and would then reclassify sentiments based on perceived tone, language, and overall message content. Every 2 weeks, a team member built a dashboard in Meltwater using analytics tools that combined the various search results into consumable insights, including top keywords, total mentions, total mentions trend, and sentiment breakdown. These dashboards highlighted instances when vaccine-related stories or sentiments were receiving higher-than-average levels of online engagement.

Social Media Campaign

From social media listening, one of the biggest issues identified that contributed to vaccine hesitancy was a general sense of misunderstanding vaccines and their role in preventing disease. To remedy this information gap, a vaccine education campaign was created. To conduct the campaign, social media accounts were created on Facebook, Instagram, and X under the name @ECHOatLCHI ("ECHO" stands for Empowering Communities for Healthy Outcomes and "LCHI" stands for Louisiana Center for Health Innovation), with the goal of providing general education about vaccines to Louisiana residents. To keep the audience engaged without overloading social media accounts, informative posts were published on these accounts 2 to 3 times a week and boosted for 7 days on Facebook and Instagram. Approximately US $15 were spent per post on each account. Social media interactions and impressions were tracked over the duration of the social media campaign. One member of the research team systematically reviewed all user comments on Facebook and Instagram posts and classified them into 2 sentiment categories (positive and negative) based on content and tone.

The included posts addressed fundamental questions such as "What is a vaccine?" "How do vaccines work?" and "Why are vaccines necessary?" (Figure S1 in Multimedia Appendix 2). From this, the content shifted to providing education about target vaccines: the influenza vaccine (Figure S2 in Multimedia Appendix 2), routine childhood vaccines (Figure S3 in Multimedia Appendix 2), and the RSV vaccine (Figure S4 in Multimedia Appendix 2). To align with current campaigns led by the Louisiana Department of Health and regional medical directors, information related to human papillomavirus (HPV) was included at the end of this campaign. Each post covered information related to disease manifestation, symptoms, available vaccines for prevention, and vaccine side effects. The target audience for the boosted posts included users in Louisiana aged 18 years and older. From this campaign, we intended to record the number of impressions we received for each post and also the basic demographic information of the target audience. Demographics such as gender and age range were captured using Meta Analytics for Facebook and Instagram. Unfortunately, X did not provide this information, limiting the analysis to Facebook and Instagram.

Results

Social Media Listening

Every 2 weeks, 2 to 3 noteworthy news stories or sentiments that attracted increased social media attention were highlighted. During the course of social media listening, up to 15 unique, noteworthy stories that signified a sentiment spike were identified (Table 1). On the basis of the analysis, most social media conversations about vaccines involved users who were either opposed to vaccines or uninterested in engaging with vaccine-related content. Generally, vaccine conversations often revolved around certain major, underlying themes, including mistrust of the government; pharmaceutical companies; the CDC, the WHO, and other public health organizations; mistrust of vaccine safety and efficacy; and health freedom (belief in bodily autonomy over mandatory vaccination).

XSL•FO
RenderX

Table . Stories or sentiments that were attracting increased social media attention.

Sentiment or story	Reporting time frame
RSV[a] vaccine awareness and lack of awareness	November 13 to 27, 2023
Texas Attorney General Ken Paxton announced plans to sue Pfizer	November 28, 2024, to December 11, 2024
A public health worker in New Zealand "proved" excess death caused by COVID-19 vaccines	November 28, 2024, to December 11, 2024
Rise in respiratory illness (COVID-19 and influenza) cases	December 25, 2023, to January 8, 2024
Florida Surgeon General Dr Joseph Ladapo's comments against mRNA[b] COVID-19 vaccine	January 9 to 22, 2024
Clusters of measles outbreaks generated conversations about childhood vaccines	January 9 to 22, 2024; February 20, 2024, to March 4, 2024; March 5 to 18, 2024; and March 19, 2024, to April 1, 2024
Mistrust toward the CDC[c], the WHO[d], and pharmaceutical companies[e]	January 23, 2024, to February 5, 2024; April 16 to 29, 2024; and May 13 to 27, 2024
Reports of COVID-19 vaccine side effects and injury[e]	February 6 to 19, 2024; March 5 to 18, 2024; April 2 to 15, 2024; and May 14 to 27, 2024
New COVID-19 isolation and vaccine guidance from the CDC	February 20, 2024, to March 4, 2024
Inaccurate information related to blood donation and COVID-19 vaccination status	February 20, 2024, to March 4, 2024
Louisiana legislature surrounding vaccines and Louisiana school vaccine opt-out rates	March 19, 2024, to April 1, 2024
H5N1 avian influenza	April 2 to 15, 2024; April 16 to 29, 2024; April 30, 2024, to May 13, 2024; and May 14 to 27, 2024
AstraZeneca withdrew its COVID-19 vaccine from the market	April 30, 2024, to May 13, 2024
Dr Anthony Fauci testified in Congress in COVID-19–related hearing	May 28, 2024, to June 10, 2024
Moderna developing combination influenza and COVID-19 vaccine	May 28, 2024, to June 10, 2024

[a]RSV: respiratory syncytial virus.

[b]mRNA: messenger RNA.

[c]CDC: Centers for Disease Control and Prevention.

[d]WHO: World Health Organization.

[e]Consistent, overarching theme in vaccine hesitancy conversations.

In addition to the sentiment spikes, we highlighted "information gaps" (Table 2), which were areas where social media users lacked accessible, credible information from trustworthy sources. In the absence of credible information, users often create and adopt unsubstantiated claims as an answer. Information gaps can be problematic as they can grow rapidly and fuel misunderstanding if not properly addressed by credible sources.

XSL•FO
RenderX

Table . Information gaps on social media among users without accessible, credible information from legitimate sources.

Information gap	Reporting time frame
RSV[a] vaccine: eligibility, availability, and how it works	November 13 to 27, 2023, and December 25, 2023, to January 8, 2024
Novavax: differences from other COVID-19 vaccines, effectiveness, and availability	November 28, 2023, to December 11, 2023
Vaccine cost and insurance coverage	November 28, 2023, to December 11, 2023
Vaccine "shedding" and other general questions about vaccine technology	January 9 to 22, 2024, and April 30, 2024, to May 13, 2024
mRNA[b] technology and safety	January 23, 2024, to February 5, 2024; February 6 to 19, 2024; and May 28, 2024, to June 10, 2024
MMR[c] vaccination and necessity	February 20, 2024, to March 4, 2024; March 5 to 18, 2024; and March 19, 2024, to April 1, 2024
New CDC[d] guidance on COVID-19	February 20, 2024, to March 4, 2024
H5N1: should I be concerned? How can I prevent it?	April 2 to 15, 2024; April 16 to 29, 2024; April 30, 2024, to May 13, 2024; and May 14 to 27, 2024
General necessity of vaccines vs just trusting one's immune system	May 28, 2024, to June 10, 2024

[a]RSV: respiratory syncytial virus.

[b]mRNA: messenger RNA.

[c]MMR: measles, mumps, and rubella.

[d]CDC: Centers for Disease Control and Prevention.

Social Media Campaign

Overall, the social media posts received 69,600 total impressions (calculated by the number of views that all posts received across all platforms; Table 3); reached 43,429 users (the number of users who saw any post); and received 652 reactions and likes, 62 shares, and 105 comments. Posts related to general vaccine information on Facebook reached a higher proportion of older users (aged 55 years and older), whereas posts on Instagram reached younger audiences (Table 3). Reach metrics were limited to Facebook and Instagram as X does not provide this metric for organic posts.

XSL•FO
RenderX

Table . Total impressions and reach on social media across different platforms, along with the most predominant age group reached, the corresponding number of users reached for each platform, and reactions.

Post ID and social media platform	Users reached, n	Impressions, n	Most prominent age range reached (y)	Users reached in this age range, n/N (%)	Comments, n	Shares, n	Total reactions, n	Breakdown of reactions
General 1								
Facebook	1609	2601	≥55	545/1609 (33.87)	0	2	3	3 likes
Instagram	828	1978	25-34	348/828 (42.03)	—[a]	—	—	—
X	—	19	—	—	—	—	—	—
General 2								
Facebook	455	821	≥55	227/455 (49.89)	12 (11 negative or spam)	5	12	5 likes, 3 laughs, 3 angry reactions, and 1 sad reaction
Instagram	925	1856	18-24	511/925 (55.24)	—	—	—	—
X	—	21	—	—	—	—	—	—
General 3								
Facebook	672	1192	18-24	370/672 (55.06)	0	2	2	2 likes
Instagram	695	1077	18-24	337/695 (48.49)	—	—	—	—
X	—	21	—	—	—	—	—	—
General 4								
Facebook	1496	2349	≥55	595/1496 (39.77)	0	3	4	3 likes and 1 angry reaction
Instagram	826	1976	25-34	330/826 (39.95)	—	—	—	—
X	—	23	—	—	—	—	—	—
General 5								
Facebook	789	1106	≥55	582/789 (73.76)	16 (all negative or spam)	8	62	47 likes, 10 laughs, 2 loves, 2 angry reaction, and 1 care reaction
Instagram	805	1307	18-24	266/805 (33.04)	—	—	—	—
X	—	31	—	—	—	—	—	—
General 6								
Facebook	633	864	≥55	243/633 (38.39)	1 (all negative or spam)	1	48	45 likes, 2 loves, and 1 laugh
Instagram	827	1752	25-34	353/827 (42.68)	—	—	—	—
X	—	18	—	—	—	—	—	—
General 7								

XSL•FO
RenderX

Post ID and social media platform	Users reached, n	Impressions, n	Most prominent age range reached (y)	Users reached in this age range, n/N (%)	Comments, n	Shares, n	Total reactions, n	Breakdown of reactions
Facebook	626	912	≥55	425/626 (67.89)	8 (all negative or spam)	1	58	33 likes, 17 laughs, 6 angry reactions, 1 care reaction, and 1 sad reaction
Instagram	1006	1882	18-24	533/1006 (52.98)	—	—	—	—
X	—	15	—	—	—	—	—	—
General 8								
Facebook	589	806	≥55	251/589 (42.61)	10 (all negative or spam)	0	56	45 likes, 9 laughs, and 2 angry reactions
Instagram	1085	1945	18-24	445/1085 (41.01)	—	—	—	—
X	—	16	—	—	—	—	—	—
General 9								
Facebook	397	442	≥55	240/397 (60.45)	0	0	53	49 likes, 3 laughs, and 1 care reaction
Instagram	1141	1858	18-24	485/1141 (42.51)	—	—	—	—
X	—	12	—	—	—	—	—	—
General 10								
Facebook	321	391	≥55	189/321 (58.88)	6 (all negative or spam)	0	56	54 likes, 1 love, and 1 angry reaction
Instagram	1247	2101	18-24	555/1247 (44.51)	—	—	—	—
X	—	18	—	—	—	—	—	—
Influenza 1								
Facebook	2076	2849	25-34	958/2076 (46.15)	1 (all negative or spam)	7	14	11 likes, 2 loves, and 1 “wow” reaction
Instagram	711	1465	25-34	259/711 (36.43)	—	—	—	—
X	—	17	—	—	—	—	—	—
Flu 2								
Facebook	468	657	≥55	151/468 (32.26)	3 (2 negative or spam)	2	58	54 likes, 3 loves, and 1 care reaction
Instagram	723	1254	18-24	235/723 (32.50)	—	—	—	—
X	—	16	—	—	—	—	—	—
Childhood 1								
Facebook	3262	4589	25-34	2158/3262 (66.16)	3 (all negative or spam)	4	2	2 likes
Instagram	867	1436	25-34	289/867 (33.33)	—	—	—	—

XSL•FO
RenderX

Post ID and social media platform	Users reached, n	Impressions, n	Most prominent age range reached (y)	Users reached in this age range, n/N (%)	Comments, n	Shares, n	Total reactions, n	Breakdown of reactions
X	—	32	—	—	—	—	—	—
Childhood 2								
Facebook	1561	1944	≥55	935/1561 (59.90)	1 (negative or spam)	0	44	41 likes, 2 loves, and 1 care reaction
Instagram	998	1337	25-34	352/998 (35.27)	—	—	—	—
X	—	22	—	—	—	—	—	—
Childhood 3								
Facebook	1983	2737	25-34	1147/1983 (57.84)	5 (all negative or spam)	5	7	7 likes
Instagram	796	1674	25-34	336/796 (42.21)	—	—	—	—
X	—	23	—	—	—	—	—	—
Childhood 4								
Facebook	1120	1551	≥55	743/1120 (66.34)	20 (19 negative or spam)	0	70	65 likes, 3 loves, 1 laugh, and 1 angry reaction
Instagram	955	1597	18-24	399/955 (41.78)	—	—	—	—
X	—	21	—	—	—	—	—	—
Childhood 5								
Facebook	1376	1886	25-34	721/1376 (52.40)	4 (3 negative or spam)	6	14	12 likes and 2 loves
Instagram	869	1541	25-34	332/869 (38.20)	—	—	—	—
X	—	29	—	—	—	—	—	—
Childhood 6								
Facebook	2250	2894	≥55	684/2250 (30.40)	11 (all negative or spam)	13	37	23 likes, 7 angry reactions, 4 laughs, 2 sad reactions, and 1 love
Instagram	1001	1545	25-34	353/1001 (35.26)	—	—	—	—
X	—	23	—	—	—	—	—	—
RSV[b] 1								
Facebook	1193	1814	25-34	586/1193 (49.12)	0	2	43	39 likes, 3 loves, and 1 care reaction
Instagram	882	1759	25-34	359/882 (40.70)	—	—	—	—
X	—	21	—	—	—	—	—	—
RSV 2								
Facebook	296	361	18-24	112/296 (37.84)	1 (all negative or spam)	0	3	3 likes
Instagram	748	1657	25-34	338/748 (45.19)	—	—	—	—

XSL•FO
RenderX

Post ID and social media platform	Users reached, n	Impressions, n	Most prominent age range reached (y)	Users reached in this age range, n/N (%)	Comments, n	Shares, n	Total reactions, n	Breakdown of reactions
X	—	13	—	—	—	—	—	—
HPV[c] 1								
Facebook	1784	2530	25-34	690/1784 (38.68)	3 (all negative or spam)	1	6	6 likes
Instagram	538	873	25-34	185/538 (34.39)	—	—	—	—
X	—	23	—	—	—	—	—	—
Total	43,429	69,600	—	—	105	62	652	549 likes, 48 laughs, 23 angry reactions, 21 loves, 6 care reactions, 4 sad reactions, and 1 “wow” reaction

[a]Not available.

[b]RSV: respiratory syncytial virus.

[c]HPV: human papillomavirus.

The most prominent and highly reached audience segment consisted of women aged 25 to 34 years, representing 28.14% (12,221/43,429) of the total reached audience (Table 4). Across Facebook and Instagram, posts achieved a higher reach and received a higher number of impressions and greater engagement among women than among men (Table 5). Facebook posts received significantly higher engagement than the other platforms. Posts focused on routine childhood vaccines outperformed posts related to general vaccine information and posts about influenza, RSV, and HPV in terms of engagement. The post related to hepatitis B and rotavirus vaccines (Childhood 1) achieved the highest reach and received the highest number of impressions, reaching 13.95% (6057/43,429) of the users and receiving 5.93% (4129/69,600) of the impressions across all platforms (Facebook, Instagram, and X; Table 3). The post related to the MMR vaccine (Childhood 4) garnered the most reactions and likes, receiving 20 comments, 65 likes, 3 love reactions, 1 “laugh” reaction, and 1 “angry” reaction across all platforms. Finally, the post with the most shares was about the childhood vaccine schedule (Childhood 6), receiving 13 shares across all platforms (Table 3).

XSL•FO
RenderX

Table . Demographic composition of the users reached on social media Empowering Communities for Healthy Outcomes accounts (n=43,429).

Age range (y) and gender	Users reached, n (%)
18-24	
Female	6168 (14.20)
Male	3234 (7.45)
Unknown	112 (0.26)
25-34	
Female	12,221 (28.14)
Male	2682 (6.18)
Unknown	76 (0.17)
35-44	
Female	6147 (14.15)
Male	1209 (2.78)
Unknown	53 (0.12)
45-54	
Female	2914 (6.71)
Male	718 (1.65)
Unknown	17 (0.04)
55-64	
Female	3007 (6.92)
Male	778 (1.79)
Unknown	13 (0.03)
≥65	
Female	3015 (6.94)
Male	1035 (2.38)
Unknown	30 (0.07)

Table . Total reach and impressions across Facebook and Instagram by gender based on the total reach and impressions across both platforms.

Gender	Users reached (n=43,429), n (%)	Impressions (Facebook, Instagram, and X) (n=69,166), n (%)
Female	33,472 (77.07)	51,550 (74.53)
Male	9656 (22.23)	17,187 (24.85)
Unknown	301 (0.69)	429 (0.62)

Of 105 comments, 84 (80%) expressed mistrust, negative attitudes, and hesitancy toward vaccines, and 17 (16.19%) were spam (Table 3). Examples included comments such as “Flu prevention comes from vitamins, juices, and sunshine,” “Complications including death come from vaccinations,” and “Many of us were skeptical about vaccinations before COVID-19, but now we know we can no longer trust any vax or any organization or doctor that pushes them.” To minimize the risk of influencing other users, all negative comments were hidden.

Discussion

Principal Findings

With the advancement of communication technologies, the medicine and public health fields have recognized the need to adopt and evolve the ways in which health messages are delivered to communities. Digital health has become vital in communicating health threats to the public; educating them about disease prevention; and promoting advances in medicine, such as new treatments [10]. Despite their potential for disseminating valuable information, social media networks have also widened gaps in access to health. This is due to an increase in skepticism, distrust, and inaccurate information caused by the overwhelming flow of true, false, and contradictory

XSL•FO
RenderX

information, exacerbating historical distrust in the government and health care providers.

During social media listening in this study, most social media conversations related to vaccines included narratives related to vaccine opposition or lack of interest in engaging with vaccine-related content. Additionally, mistrust of the government, pharmaceutical companies, the CDC, the WHO, and other public health organizations; mistrust of vaccine safety and efficacy; and health freedom (belief in bodily autonomy over mandatory vaccination) were common themes identified. This indicates that there is tremendous vaccine hesitancy among social media users, which can be dangerous and counterproductive in improving vaccine uptake, especially in vulnerable populations. Regarding information gaps, the findings indicated that social media users created and adopted unsubstantiated claims. If not swiftly addressed by credible sources, information gaps can expand and drive widespread misunderstanding. These findings can be used to inform the development of tailored vaccine messaging in Louisiana.

The demographic differences in platform engagement in this study have important implications for designing targeted public health interventions. Specifically, Facebook attracted an older demographic, whereas Instagram users tended to be between the ages of 18 and 34 years. In the breakdown of demographics of users engaged with the different vaccines posts, individuals were mostly between the ages of 25 and 34 years, with Facebook attracting a slightly older demographic for the childhood vaccination posts. This aligns with existing literature on age-related patterns in social media use and suggests that public health campaigns aiming to address vaccine hesitancy should tailor their messaging strategies accordingly [25-27]. For instance, interventions targeting older adults may benefit from prioritizing Facebook as the primary source of dissemination, whereas campaigns focused on younger populations may achieve more engagement and impact using Instagram. Understanding these platform-specific audience profiles is necessary for optimizing the delivery and effectiveness of health communication efforts, especially in the context of evolving digital media landscapes.

Furthermore, the findings highlight that women aged 25 to 34 years represented the most prominent and widely reached group across the total reached audience. Similarly, posts across Facebook and Instagram consistently achieved higher reach, a higher number of impressions, and greater engagement among women compared to their male counterparts. This is also supported by the literature, which indicates that women are more likely to use social media compared to men and more likely to express their opinions on various topics through these platforms [28-30]. Among all social media platforms, posts on Facebook received the highest engagement. Posts focusing on routine childhood vaccines demonstrated higher engagement than those addressing general vaccine information, influenza, RSV, and HPV. Among these, the post concerning the hepatitis B and rotavirus vaccines achieved the highest reach and number of impressions. The MMR vaccine post received the greatest number of reactions and likes, whereas the post detailing the childhood vaccine schedule was shared the most. Overall, most of the comments received reflected mistrust, negative sentiments, and vaccine hesitancy. Scientific literature indicates that analyses of social media content frequently reveal predominantly negative sentiment regarding vaccines [31-35].

Unlike other platforms, X has been widely used in research studies due to its relatively open access to audience and content data. In contrast, popular platforms such as YouTube, Facebook, Instagram, and TikTok have stricter privacy policies. These policies enforce more stringent data sharing regulations, which increase limitations to analyzing opinions and generating universally effective strategies [10,17,36,37]. While CrowdTangle—an insight tool that gathers and analyzes social media content [38,39]—was previously used to import user activity data from Facebook, Meta began limiting its access and eventually ended all data collection activities in mid-2024 [40]. This has led to even greater limitations in gathering information from social media and understanding people's perspectives on health prevention programs and treatments.

Despite these limitations, the research findings outlined that online vaccine sentiments are dependent on several factors, including what social media platform is more frequently used (different environments of Facebook vs Instagram vs X vs online forums), if the user stays within their algorithmically created echo chamber (users are more likely to post about a topic if their peers are also engaged), and whether vaccines are currently part of mainstream news coverage (users may not actively think about vaccines or feel encouraged to post about them until they are prompted to do so by current news stories). Past medical abuses and inconsistencies in the delivery of information have caused mistrust in many communities, especially those in underserved conditions [41-43]. Therefore, in addition to focusing on emergency response, education, and disease prevention, public health campaigns need to rebuild trust within the population. It is essential to recognize that different populations require tailored approaches depending on demographic characteristics, language, culture and religious beliefs, and needs [44-48]. Ensuring that information is accessible and understandable by everyone is crucial in public health campaigns that focus on reducing vaccine hesitancy.

Future Directions

To maximize the reach of messages, campaigns should target a variety of media sources and social media platforms, which may engage a more diverse population. According to these findings, Facebook should be prioritized. Additionally, collaborating with interdisciplinary teams such as marketing professionals, biostatisticians, educators, and researchers can further enhance the effectiveness of campaigns. Furthermore, there is a need for more research studies to evaluate the impact of social media campaigns in reducing hesitancy and increasing the adoption of public health recommendations. This will lead to the development of more community-based research approaches, which may offer additional perspectives on strategies that improve health and digital literacy by using a more cultural and linguistic approach, helping reduce barriers to trust and improve the credibility and performance of public health campaigns on social media. These studies should incorporate a more robust evaluation design, such as a pretest-posttest assessment or the use of a comparison group,

XSL•FO
RenderX

which would strengthen the ability to assess the campaign's impact.

Limitations

There are several limitations to the social media listening research conducted on the Meltwater platform. The absence of a baseline comparison meant that it was challenging to determine shifts in volume or sentiment over time. Future studies should incorporate a comprehensive evaluation framework as it would enable a more complete understanding of how educational content may contribute to shifts in public attitudes and understanding regarding vaccination. In addition, Meltwater could not obtain access to private social media groups such as those on Facebook, which could have omitted valuable discussion. Strict data collection restrictions on major platforms such as Facebook, Instagram, and YouTube limited the dataset mostly to X, news sites, blogs, and open forums. This constraint may introduce sampling bias and reduce the generalizability of the findings. Expanding data sources to include additional social media platforms could provide a more comprehensive understanding of vaccine hesitancy in the region and better align interventions with the audiences that need them the most. Finally, the short follow-up period after the start of social media listening may not have registered longer-term trends or shifting public sentiment as the attitudes toward vaccination can change with respect to emerging information and changing guidelines. These limitations point to the need to supplement social media listening with other research methods to have a complete view of knowledge and behavioral intentions related to vaccination.

The social media campaign and its analysis faced limitations as well. Relying on metrics of engagement such as likes, shares, comments, and impressions does not necessarily equate to the acquisition of knowledge and change in attitude or behavior regarding vaccination. Furthermore, the absence of a baseline or control group makes it difficult to attribute observed sentiment or engagement change to the campaign. Future studies should assess impact by incorporating a pretest-posttest assessment or a comparison group. Moreover, the brief 9-week campaign duration may not have been enough time to create permanent shifts in the attitudes or actions of people, so it is vital to implement a longer evaluation period to determine the true effect. Furthermore, there is a potential bias introduced by hiding negative comments, which may have skewed sentiment toward more positive or neutral responses. Future studies should aim to include all relevant user comments, including negative ones, to provide a more balanced and accurate assessment of public sentiment and engagement. Finally, the manual classification of social media comments into positive and negative sentiment by a single team member introduces the potential for researcher bias. Future studies should consider using multiple independent coders and calculating intercoder reliability or using a validated sentiment analysis tool.

Funding

This study was funded by the Louisiana Department of Health under grant 2000761057. The authors are fully responsible for all the work presented in this manuscript. This work does not represent the official views of the Louisiana Department of Health.

Data Availability

The data that support the results of this study are available upon request.

Authors' Contributions

AVS contributed to project execution, supervision, analysis, writing, review, and editing (The work performed for this manuscript was conducted while the author was affiliated with the University of Louisiana at Lafayette). S Smith contributed to the conceptualization of social media listening, project development, and writing. S Sakhamuri contributed to project execution, analysis, and writing. JA contributed to the conceptualization of social media listening, project development, analysis, and writing—original draft. HC contributed to methodology, conceptualization, supervision, funding acquisition, review, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Meltwater: saved searches and keywords.
[DOCX File, 23 KB - infodemiology_v6i1e76827_app1.docx]

Multimedia Appendix 2
Social media campaign examples.
[DOCX File, 1423 KB - infodemiology_v6i1e76827_app2.docx]

References

1. Broniatowski DA, Jamison AM, Johnson NF, et al. Facebook pages, the "Disneyland" measles outbreak, and promotion of vaccine refusal as a civil right, 2009-2019. Am J Public Health 2020 Oct;110(S3):S312-S318. [doi: 10.2105/AJPH.2020.305869] [Medline: 33001718]

XSL•FO
RenderX

2. Maris R, Dorner Z, Hess S, Tucker S. Social media and the evolution of vaccine preferences during the COVID-19 pandemic: discrete choice experiment. JMIR Infodemiology 2025 May 28;5:e66081. [doi: 10.2196/66081] [Medline: 40435386]
3. Adebesin F, Smuts H, Mawela T, Maramba G, Hattingh M. The role of social media in health misinformation and disinformation during the COVID-19 pandemic: bibliometric analysis. JMIR Infodemiology 2023 Sep 20;3:e48620. [doi: 10.2196/48620] [Medline: 37728981]
4. Jeyaraman M, Ramasubramanian S, Kumar S, et al. Multifaceted role of social media in healthcare: opportunities, challenges, and the need for quality control. Cureus 2023 May;15(5):e39111. [doi: 10.7759/cureus.39111] [Medline: 37332420]
5. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917. [doi: 10.2196/17917] [Medline: 33978589]
6. Yin JDC. Media data and vaccine hesitancy: scoping review. JMIR Infodemiology 2022;2(2):e37300. [doi: 10.2196/37300] [Medline: 37113443]
7. Nuwarda RF, Ramzan I, Weekes L, Kayser V. Vaccine hesitancy: contemporary issues and historical background. Vaccines (Basel) 2022 Sep 22;10(10):1595. [doi: 10.3390/vaccines10101595] [Medline: 36298459]
8. Trangerud HA. "What is the problem with vaccines?" A typology of religious vaccine skepticism. Vaccine X 2023;14:100349. [doi: 10.1016/j.jvacx.2023.100349] [Medline: 37484867]
9. Motta M, Stecula D. Quantifying the effect of Wakefield et al. (1998) on skepticism about MMR vaccine safety in the U.S. PLoS ONE 2021;16(8):e0256395. [doi: 10.1371/journal.pone.0256395] [Medline: 34411172]
10. Sylvia Chou WY, Gaysynsky A, Cappella JN. Where we go from here: health misinformation on social media. Am J Public Health 2020 Oct;110(S3):S273-S275. [doi: 10.2105/AJPH.2020.305905] [Medline: 33001722]
11. Wong FHC, Leung DKY, Wong ELY, et al. The moderating role of community capacity for age-friendly communication in mitigating anxiety of older adults during the COVID-19 infodemic: cross-sectional survey. JMIR Infodemiology 2022;2(1):e33029. [doi: 10.2196/33029] [Medline: 35257090]
12. Sasse K, Mahabir R, Gkountouna O, Crooks A, Croitoru A. Understanding the determinants of vaccine hesitancy in the United States: a comparison of social surveys and social media. PLoS One 2024;19(6):e0301488. [doi: 10.1371/journal.pone.0301488] [Medline: 38843170]
13. Sigalo N, Awasthi N, Abrar SM, Frias-Martinez V. Using COVID-19 vaccine attitudes on Twitter to improve vaccine uptake forecast models in the United States: infodemiology study of tweets. JMIR Infodemiology 2023 Aug 21;3:e43703. [doi: 10.2196/43703] [Medline: 37390402]
14. Nowak GJ, Cacciatore MA. State of vaccine hesitancy in the United States. Pediatr Clin North Am 2023 Apr;70(2):197-210. [doi: 10.1016/j.pcl.2022.11.001] [Medline: 36841590]
15. Skafle I, Nordahl-Hansen A, Quintana DS, Wynn R, Gabarron E. Misinformation about COVID-19 vaccines on social media: rapid review. J Med Internet Res 2022 Aug 4;24(8):e37367. [doi: 10.2196/37367] [Medline: 35816685]
16. Sylvia Chou WY, Gaysynsky A. A prologue to the special issue: health misinformation on social media. Am J Public Health 2020 Oct;110(S3):S270-S272. [doi: 10.2105/AJPH.2020.305943] [Medline: 33001727]
17. Walsh-Buhi ER. Social media and cancer misinformation: additional platforms to explore. Am J Public Health 2020 Oct;110(S3):S292-S293. [doi: 10.2105/AJPH.2020.305949] [Medline: 33001721]
18. Gillard CJ, Al-Dahir S, Earls M, Singleton B. A culturally competent vaccine hesitancy educational model for community pharmacists to increase vaccine uptake, Louisiana, 2021-2022. Am J Public Health 2022 Nov;112(S9):S900-S903. [doi: 10.2105/AJPH.2022.307070] [Medline: 36446056]
19. How did COVID-19 affect people in Louisiana? USA Facts. URL: https://usafacts.org/answers/how-did-covid-19-affect-people-in-the-us/state/louisiana/ [accessed 2026-04-21]
20. Flu vaccination coverage, United States, 2022–2023 influenza season. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/fluvaxview/coverage-by-season/2022-2023.html [accessed 2026-04-21]
21. Walker B, Anderson A, Stoecker C, Shao Y, LaVeist TA, Callison K. COVID-19 and routine childhood and adolescent immunizations: evidence from Louisiana Medicaid. Vaccine 2022 Feb;40(6):837-840. [doi: 10.1016/j.vaccine.2021.12.022] [Medline: 35033386]
22. Hest R. Revised childhood vaccinations measure on state health compare shows vaccination rates vary by state, race/ethnicity, and insurance coverage. State Health Access Data Assistance Center. 2024. URL: https://www.shadac.org/news/childhood-vaccinations-rate-by-state-ethnicity-insurance [accessed 2026-04-21]
23. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820. [doi: 10.2196/21820] [Medline: 32589589]
24. U.S. Department of Health and Human Services Centers for Disease Control and Prevention. Social listening and monitoring tools. 2021. URL: https://archive.cdc.gov/www_cdc_gov/vaccines/covid-19/health-departments/addressing-vaccine-misinformation.html [accessed 2026-04-30]
25. Nugent GR, Anselmi D, Chin BN. Social media, conspiracy beliefs, and COVID-19 vaccines: a survey study of emerging and middle-aged adults in the United States. Soc Sci 2025;14(1):34. [doi: 10.3390/socsci14010034]
26. Galinkala P, Atkinson E, Campos-Castillo C. Age variation among US adults' social media experiences and beliefs about who is responsible for reducing health-related falsehoods: secondary analysis of a national survey. JMIR Aging 2024 Nov 27;7:e56761. [doi: 10.2196/56761] [Medline: 39602437]

XSL•FO
RenderX

27. Sheldon P, Antony MG, Ware LJ. Baby boomers' use of Facebook and Instagram: uses and gratifications theory and contextual age indicators. Heliyon 2021 Apr;7(4):e06670. [doi: 10.1016/j.heliyon.2021.e06670] [Medline: 33889780]
28. Torralbas Oslé JE, Corcho Rosales E. Personal exposure to social media and variations by gender among Cuban youth. Psychol Russ 2023;16(4):72-89. [doi: 10.11621/PIR.2023.0405]
29. Twenge JM, Martin GN. Gender differences in associations between digital media use and psychological well-being: evidence from three large datasets. J Adolesc 2020 Feb;79:91-102. [doi: 10.1016/j.adolescence.2019.12.018] [Medline: 31926450]
30. Scott CF, Bay-Cheng LY, Prince MA, Nochajski TH, Collins RL. Time spent online: latent profile analyses of emerging adults' social media use. Comput Human Behav 2017 Oct;75:311-319. [doi: 10.1016/j.chb.2017.05.026] [Medline: 34334933]
31. Prada E, Langbecker A, Catalan-Matamoros D. Public discourse and debate about vaccines in the midst of the Covid-19 pandemic: a qualitative content analysis of Twitter. Vaccine 2023 May 11;41(20):3196-3203. [doi: 10.1016/j.vaccine.2023.03.068] [Medline: 37080830]
32. Wawrzuta D, Klejdysz J, Jaworski M, Gotlib J, Panczyk M. Attitudes toward COVID-19 vaccination on social media: a cross-platform analysis. Vaccines (Basel) 2022 Jul 27;10(8):1190. [doi: 10.3390/vaccines10081190] [Medline: 35893839]
33. Boucher JC, Cornelson K, Benham JL, et al. Analyzing social media to explore the attitudes and behaviors following the announcement of successful COVID-19 vaccine trials: infodemiology study. JMIR Infodemiology 2021;1(1):e28800. [doi: 10.2196/28800] [Medline: 34447924]
34. Ng QX, Lee DY, Ng CX, Yau CE, Lim YL, Liew TM. Examining the negative sentiments related to influenza vaccination from 2017 to 2022: an unsupervised deep learning analysis of 261,613 Twitter posts. Vaccines (Basel) 2023 May 23;11(6):1018. [doi: 10.3390/vaccines11061018] [Medline: 37376407]
35. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 1;16(11):2586-2593. [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
36. Perrin A. Social media usage: 2005-2015.: Pew Research Center; 2015. URL: https://www.pewresearch.org/internet/2015/10/08/social-networking-usage-2005-2015/ [accessed 2026-04-21]
37. Jaidka K. Cross-platform- and subgroup-differences in the well-being effects of Twitter, Instagram, and Facebook in the United States. Sci Rep 2022 Feb 28;12(1):3271. [doi: 10.1038/s41598-022-07219-y] [Medline: 35228639]
38. Pascual-Ferrá P, Alperstein N, Barnett DJ. A multi-platform approach to monitoring negative dominance for COVID-19 vaccine-related information online. Disaster Med Public Health Prep 2022 Dec;16(6):2471-2479. [doi: 10.1017/dmp.2021.136] [Medline: 33938423]
39. Aleksandric A, Anderson HI, Melcher S, Nilizadeh S, Wilson GM. Spanish Facebook posts as an indicator of COVID-19 vaccine hesitancy in Texas. Vaccines (Basel) 2022 Oct 14;10(10):1713. [doi: 10.3390/vaccines10101713] [Medline: 36298580]
40. CrowdTangle. Meta. URL: https://transparency.meta.com/researchtools/other-datasets/crowdtangle [accessed 2025-01-31]
41. Bajaj SS, Stanford FC. Beyond Tuskegee - vaccine distrust and everyday racism. N Engl J Med 2021 Feb 4;384(5):e12. [doi: 10.1056/NEJMpv2035827] [Medline: 33471971]
42. 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]
43. Jaiswal J, Halkitis PN. Towards a more inclusive and dynamic understanding of medical mistrust informed by science. Behav Med 2019;45(2):79-85. [doi: 10.1080/08964289.2019.1619511] [Medline: 31343962]
44. Sanchez AV, Ison JM, Hemley H, et al. Designing the fostering inclusivity in research engagement for underrepresented populations in Parkinson's disease study. Contemp Clin Trials 2022 Apr;115:106713. [doi: 10.1016/j.cct.2022.106713] [Medline: 35202842]
45. Sanchez AV, Ison JM, Hemley H, Jackson JD. Diversifying the research landscape: assessing barriers to research for underrepresented populations in an online study of Parkinson's disease. J Clin Transl Sci 2024;8(1):e34. [doi: 10.1017/cts.2024.20] [Medline: 38384928]
46. Ison JM, Jackson JD, Hemley H, et al. Fostering inclusivity in research engagement for underrepresented populations in Parkinson's disease: the FIRE-UP PD study. Contemp Clin Trials 2024 Sep;144:107619. [doi: 10.1016/j.cct.2024.107619] [Medline: 38971301]
47. Rooney MK, Santiago G, Perni S, et al. Readability of patient education materials from high-impact medical journals: a 20-year analysis. J Patient Exp 2021;8:2374373521998847. [doi: 10.1177/2374373521998847] [Medline: 34179407]
48. Otado J, Kwagyan J, Edwards D, Ukaegbu A, Rockcliffe F, Osafo N. Culturally competent strategies for recruitment and retention of African American populations into clinical trials. Clin Transl Sci 2015 Oct;8(5):460-466. [doi: 10.1111/cts.12285] [Medline: 25974328]

Abbreviations

CDC: Centers for Disease Control and Prevention

XSL•FO
RenderX

HPV: human papillomavirus
MMR: measles, mumps, and rubella
RSV: respiratory syncytial virus
WHO: World Health Organization

Edited by T Purnat; submitted 01.May.2025; peer-reviewed by B Olujobi, H Fu, NA Rossi, T Amusa; revised version received 27.Aug.2025; accepted 15.Jan.2026; published 06.May.2026.

Please cite as:
Sanchez AV, Smith S, Sakhamuri S, Ardoin J, Chu H
Understanding Vaccine Hesitancy in Louisiana Through Social Media Listening and Community Feedback: Cross-Sectional Study
JMIR Infodemiology 2026;6:e76827
URL: https://infodemiology.jmir.org/2026/1/e76827
doi:10.2196/76827

© Angie Viviana Sanchez, Sadie Smith, Sahithya Sakhamuri, Julia Ardoin, Henry Chu. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 6.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Marketing Strategies and Factors Influencing the Popularity of Alcohol Videos from Official Brand Accounts on Douyin: Content Analysis Study

Yuchen Zhao[1], MB; Lingyun Zhang[1,2], MPH; Chenyu Qian[1], MPH; Wenjie Guo[1], MB; Weiyun Zhu[1], MB; Pinpin Zheng[1], PhD

[1]Department of Preventive Medicine and Health Education, School of Public Health, Institute of Health Communication, Fudan University, 138 Yixueyuan Road, Xuhui District, Shanghai, China

[2]Party Committee Office, Zhongshan Hospital Affiliated with Fudan University, Shanghai, China

Corresponding Author:
Pinpin Zheng, PhD
Department of Preventive Medicine and Health Education, School of Public Health, Institute of Health Communication, Fudan University, 138 Yixueyuan Road, Xuhui District, Shanghai, China

Abstract

Background: Alcohol consumption in China poses significant public health challenges. Alcohol marketing has been shown to increase public alcohol consumption, with social media platforms such as Douyin (TikTok in Mainland China) being among the main channels for alcohol marketing.

Objective: This study aimed to analyze the thematic content of alcohol advertising on the Douyin platform and to explore the factors influencing the popularity of these types of advertising.

Methods: Using data from the JINGDONG platform and alcohol industry reports, we identified 40 popular alcohol brands. For each brand, we located their official Douyin accounts and selected the top 20 most-liked videos posted between November 1, 2020, and November 1, 2021. In total, 659 videos from 37 brands were collected for analysis. Two trained researchers independently coded each video using a predefined codebook, which consisted of 7 sections and 20 items. Binary logistic regression was conducted with the grouping of the number of likes as the dependent variable, and the marketing strategies and warning elements of each video as independent variables.

Results: Among the 659 videos analyzed, 320 (48.6%) garnered more than 1000 likes. A significant portion of the videos was direct advertisements (281/659, 42.6%) and short skits (255/659, 38.7%), with 56.0% (369/659) featuring characters engaging in drinking-related behaviors or directly consuming alcohol. Additionally, many videos highlighted brand elements (510/659, 77.4%) and extended features (161/659, 24.4%). Cultural themes were also common, with 23.2% (153/659) of the videos promoting the enjoyment of life and 6.8% (45/659) emphasizing balance in life. However, age restrictions were missing for 26.9% (177/659) of the videos, and only 1.2% (8/659) included a health warning stating that “Drinking is harmful to health.” Certain marketing strategies were significantly associated with greater video popularity, including the use of short skits (odds ratio [OR] 2.77, 95% CI 1.42 - 5.41), highlighting brand elements (OR 2.96, 95% CI 1.59 - 5.51), and emphasizing life balance (OR 3.44, 95% CI 1.11 - 10.66). In contrast, the presence of age restrictions (OR 0.32, 95% CI 0.15 - 0.67) and explicit health warnings (OR 0.06, 95% CI 0.01 - 0.84) were associated with lower popularity. The period from July to September and November was the peak release period for alcohol advertisements on Douyin.

Conclusions: Alcohol marketing strategies on Douyin leverage experiential, brand-driven, collaborative, and cultural marketing techniques to enhance video attractiveness and create alcogenic environments. Moreover, effective age restrictions and health warnings are largely absent. It is essential to legislate and enforce stricter alcohol marketing regulations to reduce the health risks associated with alcohol marketing.

(JMIR Infodemiology 2026;6:e74221) doi:10.2196/74221

KEYWORDS

alcohol marketing; Douyin platform; alcogenic environments; advertisement; health warning

XSL•FO
RenderX

Introduction

Alcohol consumption is a leading risk factor for the global disease burden, which is associated with the risk of more than 200 diseases and injuries, causing approximately 3 million deaths globally each year, accounting for 5.3% of all deaths [1]. There is growing evidence that the public's alcohol consumption could be influenced by "alcogenic environments," which are settings in which alcohol is easily accepted, available, and affordable [2]. "Alcogenic environments" have been shown to be associated with harmful drinking behaviors, including alcohol addiction and underage drinking [3-5].

As an important part of the "commercial determinants of health," alcohol marketing is a key driver for creating an alcogenic environment [2,5]. The alcohol industry often promotes products through intensive advertising and even emphasizes the "function" of drinking [3]. This marketing strategy of widespread advertising could help enhance the acceptability and normalization of alcohol consumption at both the individual and community levels, resulting in an increase in the alcogenic environment [5-8].

The widespread use of digital media has provided new channels for alcohol marketing, such as websites, apps, and social media [9,10]. Alcohol marketing on social media platforms has grown rapidly over the past two decades, with the alcohol industry reporting that alcohol marketing on social media platforms (including YouTube and Facebook) has reached many more consumers [11]. For instance, a previous study revealed that music videos on YouTube conveyed approximately 10.06 billion alcohol-related impressions to the British population [12].

As a country with high alcohol consumption, alcohol marketing on social media in China is also severe [13]. The annual advertising investment of the Chinese liquor industry has exceeded 20 billion yuan (approximately US $3 billion), with the goal of creating an all-media marketing platform combining television, websites, and social media [14]. A Chinese alcohol industry report in 2021 informed that 89% of alcohol consumers received alcohol-related information through the internet, with more than 30% coming from short video platforms such as Douyin (TikTok in mainland China) [15].

Douyin, China's leading short video platform with more than 800 million daily active users, is primarily youth-centric, with 60% of its users aged younger than 30 years [16]. It has become a major hub for entertainment, social interaction, and marketing, leveraging algorithm-based recommendations to help users discover tailored content and enabling businesses to reach a broad audience with creative campaigns. Under the "Chinese liquor" tag, videos related to Chinese liquor on Douyin have a total of 65.63 billion views, with the most popular video receiving as many as 878,000 likes [17]. These videos feature diverse drinking scenes and emotions, using promotional strategies, such as collaborations with opinion leaders and brand partnerships [18].

According to the attention-interest-desire-action (AIDA) model, the consumer purchase decision-making process is characterized by a progression through four stages [19]: attention, interest, desire, and action. Previous research has demonstrated that higher levels of alcohol social media marketing exposure are associated with positive drinking expectancies and drinking behaviors [20,21], confirming that the AIDA model can be used to explain alcohol advertising exposure, where attention attracted by alcohol advertisements can translate into drinking behavior. However, regarding the "A (attention)" and "I (interest)" components of the AIDA model, likes can be viewed as a preliminary quantitative indicator of successful "interest" capture, and it remains unclear from previous research what types of alcohol-related social media advertisements effectively capture audience attractiveness.

The World Health Organization (WHO) has called for interventions to reduce the alcogenic environment, and the key interventions include restrictions on alcohol marketing [2]. However, China lacks comprehensive regulations governing social media alcohol advertising. It is essential to determine the marketing strategies of alcohol products to formulate more comprehensive regulations and supervision measures.

This study aimed to identify the marketing strategies of alcohol advertisements and the placement of warnings on the Douyin platform, the factors associated with the attractiveness of those alcohol advertisements, and the time trend of these alcohol videos. These findings may contribute to extending the application of the AIDA model in social media alcohol advertising, with a specific focus on exploring the mechanisms through which different types of alcohol marketing strategies facilitate the transition from the "attention" to the "interest" stage. These findings are also expected to provide policymakers with insights into how alcohol products are promoted and marketed on social media, which is crucial for enhancing the regulation of alcohol marketing.

Methods

Sampling and Data Collection

In 2021, the size of the online market for alcohol in China reached 136.31 billion yuan [22], of which JINGDONG [23] was one of the main Chinese e-commerce platforms. The sales rankings of alcohol brands on JINGDONG can reflect the popularity of those brands. By reviewing alcohol industry reports, we categorized the alcohol currently sold in China into six groups: Chinese liquor (spirits), beer, wine, yellow rice wine, fruit wine, and premixed cocktails. On the basis of this classification, we determined the alcohol brands for this study as follows:

- Database A: According to retail data from the JINGDONG platform in 2021, the top 5 alcohol brands in each category were selected, resulting in a total of 30 brands representing popular choices across all age groups. These brands represent products commonly purchased by general consumers.
- Database B: In addition to the brands aforementioned, some new brands are very popular among young people and women but lack large sales records. We identified the 10 most popular brands targeting these demographics by

XSL•FO
RenderX

reviewing industry reports on alcohol consumption by young people and women.

- Combining databases A and B, a total of 40 brands were included in the study. We then identified the official Douyin accounts for each brand. Three brands did not have official Douyin accounts, leaving 37 brands for inclusion (Table 1). Given the similar formats of Douyin advertisements within the same brand and for the feasibility of the study, we selected the top 20 most-liked videos from each account that were released between November 1, 2020, and November 1, 2021. For brands that had released fewer than 20 videos in the past year, all available videos were included. The selection of 2021 as the data collection window was based on a key transition in Douyin's commercialization process: before 2020, Douyin's strategy focused primarily on user acquisition, while from 2020 onward, its commercialization accelerated, with the platform achieving the top position in domestic advertising revenue that same year [24]. This makes 2020 - 2021 a suitable period for examining the marketing strategies of alcohol brands during the platform's commercialization push. In total, 659 videos were collected for further analysis.

Table . Samples of alcohol brands in this study.

Categories	Brands
Chinese liquor	MOUTAI, WuLiangYe, Luzhou Laojiao, Yanghe, Fen Jiu, and Jiang Xiaobai
Beer	Budweiser, Tsingtao, Snowflake, Yanjing, Harbin, and Corona
Yellow or rice wine	Guyuelongshan, Kuaijishan, Jimo rice wine, Nverhong, MIK, Suzhou Qiao
Wine	Penfolds, Greatwall, Jacob's Creek, Lafite, Chateau Monlot, Changyu/Torre Oria
Fruit wine or preconditioning of cocktails	Jin liquor, Jinro, UMEET, Huatian Xiangz, RIO, Breezer, Power Station, Jiushilang
Imported liquor	Absolut Vodka, Bacardi, Johnnie Walker, Ballantine, and Hennessy

Coding Method

Codebook development was a top-down or bottom-up process. This process included the following phases. First is the top-down phase—reviewing the previous studies and the related advertising marketing reports [9,25,26], which allowed us to identify the marketing strategies used by official alcohol accounts on Douyin. Afterward, a preliminary framework of the codebook was developed with 8 main dimensions, including basic information, content presentation, scene setting, brand and product appeal, promotion strategy, emotion, culture, and warning (Multimedia Appendix 1). Second is the bottom-up phase—reviewing 20 randomly selected videos of official alcohol accounts on Douyin, which helped us identify the strategies that were undetermined in the preliminary framework. Accordingly, we enriched the codebook with subdimensions derived from the video content, including the specific form of the presentation, the characters' drinking action, the scene, the product promotion strategy, the emotional tone, the cultural appeal, and different types of warnings. Third, 20 videos were randomly selected from the remaining samples for retested coding to verify the applicability of the current codebook. The resulting codebook consisted of 8 sections with 19 items, including basic information, content presentation, scene setting, brand and product appeal, promotion strategy, emotion, culture, and warning (Table 2).

Apart from the aforementioned elements, this research also collected the like counts of each video as the variable of video attractiveness. As 1000 is close to the median number of likes (Median=870, IQR: 135-7612), we categorized the number of likes into a high likes group and a low likes group using 1000 as the cutoff point.

The content analysis of the alcohol-related video started in December 2021 and took 1 month in total. A total of 659 videos were monitored. A total of 3 public health researchers participated in the coding work, and all researchers received training to unify the coding standards before the start of coding. After the training, the consistency of the researchers when coding the same video reached a kappa value of 92.8%, indicating a high level of consistency in their coding standards.

The specific coding work was conducted via an online questionnaire platform "Wenjuanxing." The codebook was preentered into the platform and presented in the form of choice questions. Researchers selected the corresponding options on the online questionnaire platform while viewing Douyin videos. All videos were downloaded to local computers and coded independently by 2 researchers. If there were inconsistencies, the 2 researchers were required to rereview the video to confirm it. If they still failed to reach an agreement, the third researcher was involved in the discussion to reach a consensus.

XSL•FO
RenderX

Table . Coding book of alcohol videos on the Douyin platform.

Dimensions, variable, and category	Definition or example
Basic information	
Durations	
≤30	—[a]
31-60	—
≥61	—
Brand category	
Traditional	Brands from database A
New	Brands from database B
Content presentation	
Form of presentation	
Advertisement	A direct promotional message aimed at selling a product
Short skit	A brief video performance using theatrical elements
Vlog	Short-form documentaries
Film and TV[b] show excerpts	The clips from film and TV shows
Other	Other forms
Characters' drinking action	
No characters and no drinking behavior	No characters appear
Characters appear without drinking behavior	Characters do not touch alcohol products
Characters appear with drinking-related behavior	Characters hold alcohol glasses or pour alcohol and clink glasses
Characters appear with drinking behavior directly	Characters drink alcohol directly
Scene setting	
Drinking alone	Contain scenes of drinking alone
Party	Contain scenes of drinking at parties
Natural scenery	Contain scenes of natural scenery
Cultural or sports activities	Contain scenes of cultural or sports activities
Brand and product appeal	
Brand elements	Contain brand name, brand logo, or brand mascot
Product elements	Contain product name or product logo
Intrinsic product features	Contain information about the odor, color, taste, and materials of the product
Extended product features	Contain information about the origin, production process, vintage, and creative drinking methods
Promotion strategy	
Product promotion strategy	
Key opinion leaders	Invite celebrities to endorse
Cross-border brand cooperation	Collaborate with other brands to promote
Interaction with audience	Engage with the audience to attract fans
Not mentioned	—
Cues refer to women's interests	Contain cues that refer to women's interests, such as flowers, perfume
Cues refer to youth's interests	Contain cues that refer to youth interests, such as cartoons, cosplay
Emotion	
Emotional tone	

XSL•FO
RenderX

Dimensions, variable, and category	Definition or example
Positive	A favorable or optimistic emotion conveyed in the video, such as pleasure, moving
Neutral	An emotion neither leans positive nor negative, such as calmness
Negative	An unfavorable or unpleasant emotion conveyed in the video, such as sadness, loss
Culture	
Cultural appeal	
Historical inheritance	Highlighting the alcohol's history, heritage, and tradition
Festival celebration	Emphasizing drinking as a way to celebrate the holiday
Balance in life	Emphasize the philosophical concept of balance in life gained through drinking
Ambition and striving	Drinking symbolizes ambition and striving for success
Enjoy one's life	Emphasizing drinking as a way to enjoy life
Not mentioned	—
Warning	
Age restriction	
Yes or no	Contain age restrictions, such as "minors under the age of 18 are prohibited from drinking alcohol"
Health warnings	
Drinking is harmful to health	Contain health warning related to "drinking is harmful to health"
Please drink responsibly	Contain health warning related to "please drink responsibly"
Not mentioned	—

[a]Not applicable.

[b]TV: television.

Statistical Analysis

Frequencies and proportions are reported for the basic information, the marketing elements, and the warnings of Douyin videos. Chi-square tests were conducted to examine the association between different marketing strategies and the grouping of like counts, as well as the association between the warnings in videos and the grouping of like counts. As the dependent variable, "grouping of like counts," is a binary variable, we conducted a binary logistic regression, with the low-likes group serving as the reference. The marketing elements and warning elements of each video, which were recorded through coding, served as the independent variables. The enter method was adopted for the model to explore the factors affecting the popularity of Douyin videos among the public.

Adjusted odds ratios (ORs) and their 95% CIs were used to quantify the effects. To evaluate the model's goodness of fit, the Nagelkerke R^2 was used to determine the explanatory power of the model. To investigate the seasonal variations in Douyin videos, this study compiled the release times of the sampled videos and generated a scatter plot depicting the monthly distribution of video releases. IBM SPSS software (version 20.0) was used to carry out all the analyses.

Ethical Considerations

According to Article 32 of China's National Health Commission, Ministry of Education, and Ministry of Science and Technology Document No. 4 in 2023 "Notice on Issuing the Measures for Ethical Review of Human Life Science and Medical Research," research using legally obtained publicly available data, data generated through observation without interfering with public behavior, or anonymized information data is exempt from ethical review [27]. This study uses legally accessible public data from social media platforms, involves no individual user data, and does not interfere with public behavior. Therefore, this study qualifies for an exemption from ethical review.

Results

The Marketing Strategies of Alcohol Advertisements and the Placement of Warnings

Among the 659 Douyin videos analyzed, the number of likes ranged from 2 to 440,000, with a median of 870 (IQR: 135-7612). Using 1000 likes as the cutoff point, 320 videos (48.6%) were classified as high-liked videos, with more than 1000 likes, whereas 339 videos (51.4%) were classified as low-liked videos, with fewer than 1000 likes.

Most videos were presented as advertisements (n=281, 42.6%) and short skits (n=255, 38.7%), with 56.0% (n=369) of the characters engaging in drinking-related behavior or drinking

XSL•FO
RenderX

directly. The most frequent scenes were parties (n=170, 25.8%) and natural scenery (n=112, 17.0%). A total of 254 (38.5%) videos showed the product's intrinsic features, such as taste and odor, whereas 161 (24.4%) videos highlighted extended features, such as creative ways of mixing and drinking Rio cocktails with Sprite. Some videos conveyed drinking-related culture, mainly including enjoying one's life (n=153, 23.2%) and historical inheritance (n=65, 9.9%). For example, "Rainy days go better with blueberry wine" and "Tradition is our unwavering craftsmanship" (Figure 1).

Additionally, 36.6% (n=241) of the videos included elements favored by women, such as flowers and perfume, and 16.1% (n=106) contained elements appealing to teenagers, including e-sports and anime. However, not all videos included age restrictions (n=482, 73.1%), and only 1.2% (8/659) contained health warnings "Drinking is harmful to health" (Table 3).

XSL•FO
RenderX

Figure 1. Screens depicting the marketing strategy for alcohol advertisements on the Douyin platform. (A) Use of a short skit. (B) Creative method of drinking the product (adding grapefruit sauce to a cocktail). (C) Use of key opinion leaders. (D) The promotion of a life balance philosophy (drinking is a time management philosophy that balances patience and happiness).

XSL•FO
RenderX

Table . Characteristics comparison of alcohol videos with different popularities.

Dimensions, variable, and category	Values, n (%)	Low likes, n (%)	High likes, n (%)	Chi-square (*df*)	*P* value
Basic Information					
Durations				12.1 (2)	.002
0-30	340 (51.6)	184 (54.3)	156 (48.8)		
31-60	150 (22.8)	87 (25.7)	63 (19.7)		
>60	169 (25.6)	68 (20.1)	101 (31.6)		
Brand category				9.6 (1)	.002
New	299 (45.4)	134 (39.5)	165 (51.6)		
Traditional	360 (54.6)	205 (60.5)	155 (48.4)		
Alcohol category				161.6 (5)	<.001
Chinese liquor	118 (17.9)	42 (12.4)	76 (23.8)		
Beer	111 (16.8)	8 (2.4)	103 (32.2)		
Wine	103 (15.6)	80 (23.6)	23 (7.2)		
Imported liquor	74 (11.2)	53 (15.6)	21 (6.6)		
Fruit wine or preconditioning of cocktails	134 (20.3)	69 (20.4)	65 (20.3)		
Yellow or rice wine	119 (18.1)	87 (25.7)	32 (10.0)		
Content presentation					
Form of presentation				7.6 (4)	.11
Advertisement	281 (42.6)	147 (43.4)	134 (41.9)		
Short skit	255 (38.7)	122 (36.0)	133 (41.6)		
Film and television show excerpts	20 (3.0)	7 (2.1)	13 (4.1)		
Vlog	70 (10.6)	42 (12.4)	28 (8.8)		
Other	33 (5.0)	21 (6.2)	12 (3.8)		
Characters' drinking action				28.1 (3)	<.001
No characters and no drinking behavior	106 (16.1)	79 (23.3)	27 (8.4)		
Characters appear without drinking behaviors	184 (27.9)	92 (27.1)	92 (28.8)		
Characters appear with drinking-related behaviors	265 (40.2)	119 (35.1)	146 (45.6)		
Characters appear with drinking behaviors	104 (15.8)	49 (14.5)	55 (17.2)		
Scene setting					
Drinking alone				3.1 (1)	.08
Yes	99 (15.0)	59 (17.4)	40 (12.5)		
No	560 (85.0)	280 (82.6)	280 (87.5)		
Party				16.0 (1)	<.001
Yes	170 (25.8)	65 (19.2)	105 (32.8)		
No	489 (74.2)	274 (80.8)	215 (67.2)		

XSL•FO
RenderX

Dimensions, variable, and category	Values, n (%)	Low likes, n (%)	High likes, n (%)	Chi-square (*df*)	*P* value
Natural scenery				0.6 (1)	.45
Yes	112 (17.0)	54 (15.9)	58 (18.1)		
No	547 (83.0)	285 (84.1)	262 (81.9)		
Cultural or sports activities				3.5 (1)	.06
Yes	75 (11.4)	31 (9.1)	44 (13.8)		
No	584 (88.6)	308 (90.9)	276 (86.3)		
Brand and product appeal					
Brand elements				45.9 (1)	<.001
Yes	510 (77.4)	226 (66.7)	284 (88.8)		
No	149 (22.6)	113 (33.3)	36 (11.3)		
Product elements				2.0 (1)	.16
Yes	399 (60.5)	214 (63.1)	185 (57.8)		
No	260 (39.5)	125 (36.9)	135 (42.2)		
Intrinsic product features				48.2 (1)	<.001
Yes	254 (38.5)	174 (51.3)	80 (25.0)		
No	405 (61.5)	165 (48.7)	240 (75.0)		
Extended product features				28.0 (1)	<.001
Yes	161 (24.4)	112 (33.0)	49 (15.3)		
No	498 (75.6)	227 (67.0)	271 (84.7)		
Promotion strategy					
Product promotion				22.7 (3)	<.001
Key opinion leaders	53 (8.0)	14 (4.1)	39 (12.2)		
Cross-border brand cooperation	32 (4.9)	23 (6.8)	9 (2.8)		
Interaction with audience	12 (1.8)	3 (0.9)	9 (2.8)		
Not mentioned	562 (85.3)	299 (88.2)	263 (82.2)		
Cues refer to women’s interests				4.4 (1)	.04
Yes	241 (36.6)	111 (32.7)	130 (40.6)		
No	418 (63.4)	228 (67.3)	190 (59.4)		
Cues refer to teenagers’ interests				20.5 (1)	<.001
Yes	172 (26.1)	63 (18.6)	109 (34.1)		
No	487 (73.9)	276 (81.4)	211 (65.9)		
Emotion					
Emotional tone				3.0 (2)	.22
Positive	263 (39.9)	128 (37.8)	135 (42.2)		
Neutral	359 (54.5)	195 (57.5)	164 (51.3)		
Negative	37 (5.6)	16 (4.7)	21 (6.6)		
Culture					
Cultural appeal				40.4 (6)	<.001
Historical inheritance	65 (9.9)	43 (12.7)	22 (48.8)		

XSL•FO
RenderX

Dimensions, variable, and category	Values, n (%)	Low likes, n (%)	High likes, n (%)	Chi-square (*df*)	*P* value
Festival celebration	49 (7.4)	35 (10.3)	14 (6.9)		
Balance in life	45 (6.8)	8 (2.4)	37 (4.4)		
Ambition and striving	33 (5.0)	17 (5.0)	16 (19.7)		
Enjoy one's life	153 (23.2)	90 (26.5)	63 (11.6)		
Gift-giving culture	21 (3.2)	9 (2.7)	12 (5.0)		
Not mentioned	292 (44.5)	137 (40.4)	156 (3.8)		
Warning					
Age restriction				19.4 (1)	<.001
Yes	482 (73.1)	273 (80.5)	209 (65.3)		
No	177 (26.9)	66 (19.5)	111 (34.7)		
Health warnings				17.0 (1)	<.001
Drinking is harmful to health	8 (1.2)	6 (1.8)	2 (0.6)		
Please drink responsibly	68 (10.3)	50 (14.7)	18 (5.6)		
Not mentioned	583 (88.5)	283 (83.5)	300 (93.8)		

The Factors Associated With the Attractiveness of Alcohol Advertisements

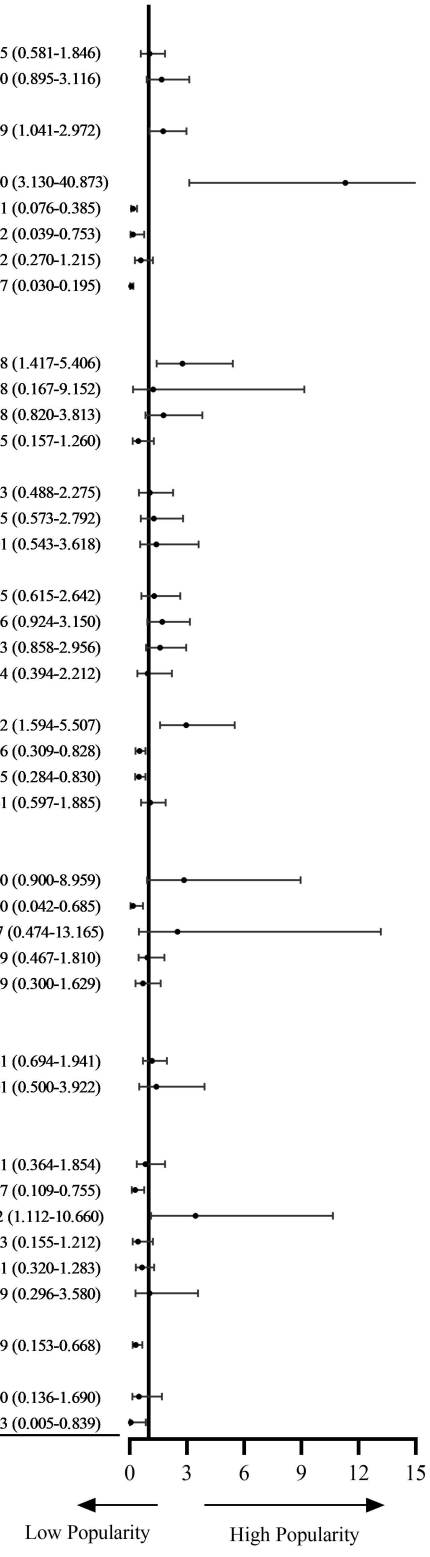
Compared with videos from traditional brands, videos from new brands were more likely to receive more likes (OR 1.759, 95% CI 1.041 - 2.972). Compared with Chinese liquor videos, beer videos garnered more likes (OR 11.310, 95% CI 3.130 - 40.873), whereas wine, imported liquor, and yellow or rice wine videos received fewer likes (OR 0.171, 95% CI 0.076 - 0.385; OR 0.172, 95% CI 0.039 - 0.753; and OR 0.077, 95% CI 0.030 - 0.195).

Compared with direct advertisements, videos presented as short skits were more likely to receive likes (OR 2.768, 95% CI 1.417 - 5.406). Videos conveying the culture of "balance in life" received more likes (OR 3.442, 95% CI 1.112 - 10.660). Additionally, videos that displayed brand-related information were more attractive (OR 2.962, 95% CI: 1.594 - 5.507), whereas those showcasing product elements or intrinsic features garnered fewer likes (OR 0.506, 95% CI 0.309 - 0.828; and OR 0.485, 95% CI 0.284 - 0.830). Furthermore, videos with age restrictions and the health warning "Drinking is harmful to health" tended to receive fewer likes (OR 0.319, 95% CI 0.153 - 0.668; OR 0.063, 95% CI 0.005 - 0.839). The total Nagelkerke R^2 of this logistic regression was 0.561 (Figure 2).

XSL•FO
RenderX

Figure 2. Multivariate logistic regression of popularity of alcohol-related videos. OR: odds ratio.

Dimension	Variable	Category	P	OR (95%CI)
Basic information				
	Duration	0s to 30s (Ref.)		
		31 to 60s	.91	1.035 (0.581-1.846)
		More than 60s	.11	1.670 (0.895-3.116)
	Brand category	Traditional (Ref.)		
		New	.04	1.759 (1.041-2.972)
	Alcohol category	Chinese liquor(Ref.)		
		Beer	<.001	11.310 (3.130-40.873)
		Wine	<.001	0.171 (0.076-0.385)
		Imported liquor	.02	0.172 (0.039-0.753)
		Fruit wine/ cocktails	.12	0.572 (0.270-1.215)
		Yellow/rice wine	<.001	0.077 (0.030-0.195)
Content presentation				
	Form of presentation	Advertisement (Ref.)		
		Short skit	.003	2.768 (1.417-5.406)
		Film and TV show excerpts	.84	1.238 (0.167-9.152)
		Vlog	.15	1.768 (0.820-3.813)
		Other	.13	0.445 (0.157-1.260)
	Characters' drinking action	No characters and no drinking behavior (Ref.)		
		Characters appear without drinking behaviors	.90	1.053 (0.488-2.275)
		Characters appear with drinking related behaviors	.56	1.265 (0.573-2.792)
		Characters appear with drinking behaviors directly	.49	1.401 (0.543-3.618)
Scene setting				
	Drinking alone	Yes vs No	.51	1.275 (0.615-2.642)
	Party	Yes vs No	.09	1.706 (0.924-3.150)
	Natural scenery	Yes vs No	.14	1.593 (0.858-2.956)
	Cultural and sports activities	Yes vs No	.88	0.934 (0.394-2.212)
Brand and product appeal				
	Brand elements	Yes vs No	.001	2.962 (1.594-5.507)
	Product elements	Yes vs No	.007	0.506 (0.309-0.828)
	Intrinsic product features	Yes vs No	.008	0.485 (0.284-0.830)
	Extended product features	Yes vs No	.84	1.061 (0.597-1.885)
Promotion strategy				
	Product promotion strategy	Not mentioned (Ref.)		
		Key opinion leaders	.08	2.840 (0.900-8.959)
		Cross-border brand cooperation	.01	0.170 (0.042-0.685)
		Interaction with audience	.28	2.497 (0.474-13.165)
	Elements attracting women	Yes vs No	.81	0.919 (0.467-1.810)
	Elements attracting teenagers	Yes vs No	.41	0.699 (0.300-1.629)
Emotion				
	Emotional tone	Neutral (Ref.)		
		Positive	.60	1.161 (0.694-1.941)
		Negative	.52	1.401 (0.500-3.922)
Culture				
	Cultural appeal	Not mentioned (Ref.)		
		Historical inheritance	.64	0.821 (0.364-1.854)
		Festival celebration	.01	0.287 (0.109-0.755)
		Balance in life	.03	3.442 (1.112-10.660)
		Ambition and striving	.11	0.433 (0.155-1.212)
		Enjoy one's life	.21	0.641 (0.320-1.283)
		Gift-giving culture	.96	1.029 (0.296-3.580)
Warning				
	Age restriction	Yes vs No	.002	0.319 (0.153-0.668)
	Health warnings	Not mentioned (Ref.)		
		Please drink responsibly	.25	0.480 (0.136-1.690)
		Drinking is harmful to health	.04	0.063 (0.005-0.839)

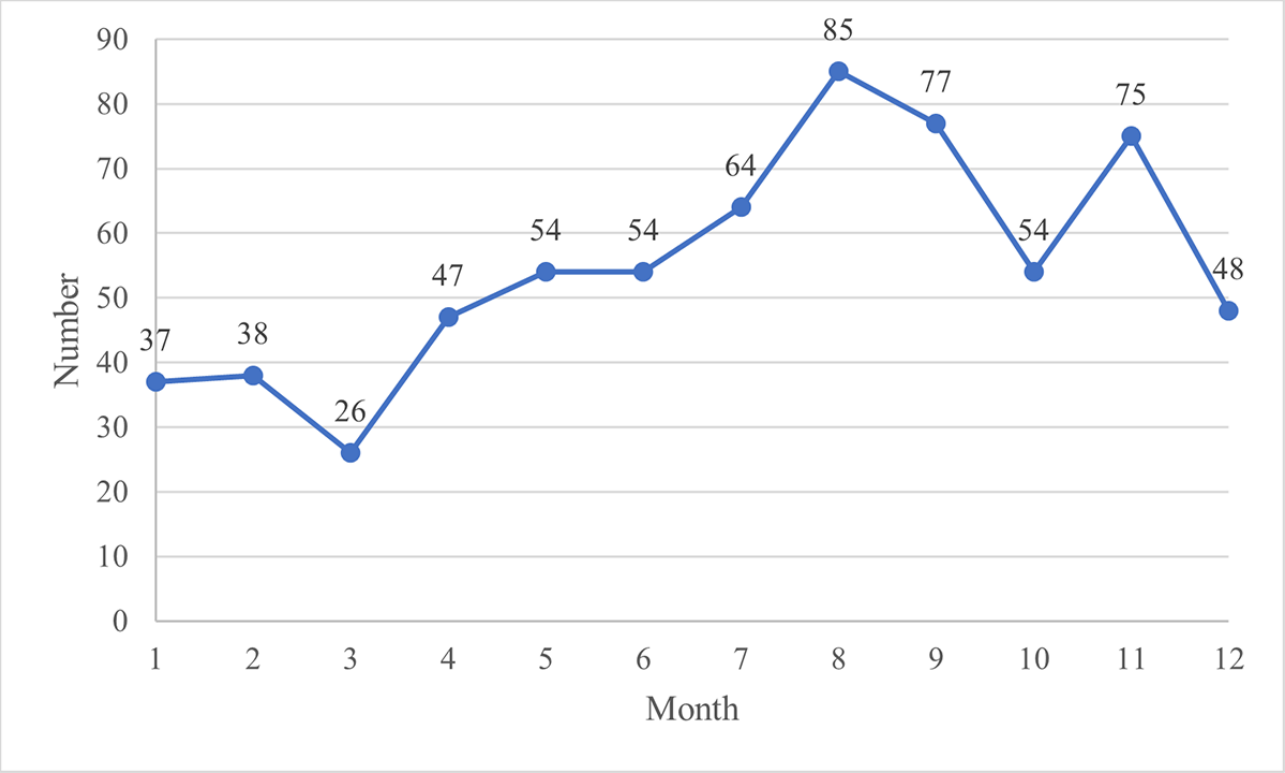


Temporal Trend of Alcohol Video Releases on Douyin

The period from July to September of that year was the peak season for alcohol advertisements on Douyin, with the highest monthly number reaching 85. In addition, a release peak for Douyin alcohol videos was recorded in November, with the number reaching 75 (Figure 3).

XSL•FO
RenderX

Figure 3. Distribution of Douyin alcohol advertisement release times.



Discussion

Principal Findings

To the best of our knowledge, this is the first study to analyze the thematic content of alcohol advertising on the Douyin platform. Through the analysis of 659 Douyin videos, we identified several alcohol marketing strategies, including diverse forms of presentation, highlighting the brand elements and incorporating cultural elements into the videos. In addition, we explored the factors influencing the popularity of those videos and reported that the use of several marketing approaches was positively correlated with the attractiveness of the videos. All these alcohol marketing practices may accelerate public alcohol consumption through enhancing alcogenic environment. These findings emphasize the urgent need for strong policy formulation and enforcement to reduce the negative influence of alcohol marketing on social media platforms.

Alcohol Types and Temporal Trends

According to the results, beer advertisements were more popular than Chinese liquor advertisements, whereas wine, imported liquor, and rice or yellow wine advertisements received less attention. We speculate that this is due to the high overlap between the characteristics of the beer-consuming audience and Douyin users. According to the Douyin user portrait report, the main audience of Douyin consists of young people aged 19 to 30 years from non–first-tier cities [28]. Chinese youth primarily prefer beer, followed by Chinese liquor, whereas wine, imported liquor, and yellow wine are far less popular [29]. The temporal trend of Douyin alcohol videos also reflects young Chinese people's preferences. Unlike the traditional peak season of Chinese liquor during the Spring Festival season (December to February) [30], alcohol advertisements on Douyin are heavily launched during the summer vacation period (July to September) as well as the month of the Double 11 Shopping Festival to cater to young people's demand for beer consumption. As in previous studies, all these factors indicate that social media platforms such as Douyin have become popular alcohol marketing channels targeting youth [31-33].

Alcohol Marketing Strategies on Douyin

The first alcohol marketing strategy on the Douyin platform is experiential marketing, a prime example of which is the utilization of short skits [34]. Unlike traditional advertising, these skits occur within an entertainment context, becoming part of an experience that immerses the viewer in the storyline [35]. This approach may blur the lines between audiences' personal lives and commercial messages while frequently depicting alcohol-related behaviors, which may encourage higher drinking frequency among viewers [31,36,37]. Another marketing strategy is brand marketing, which emphasizes brand elements such as the logo or mascot, rather than simple introductions of products. These elements could convey brand emotions and values, prompting people to hold a positive attitude toward a brand or product, thus increasing the probability of purchasing [38,39]. Brand marketing is not only common in China but has also been validated by previous studies as one of the prevalent strategies in global alcohol marketing [40].

The third strategy is collaborative marketing, in which the involvement of celebrities becomes the main strategy. Positive characteristics associated with the celebrity can be transferred to the product. In particular, the celebrity endorsement of young people could increase their recollection of drinking images [32,41]. In addition, many alcohol brands choose to collaborate with other brands or activities, including the National Basketball

XSL•FO
RenderX

Association and music festivals, which could also broaden their audience beyond traditional alcohol consumers. The last strategy is cultural marketing. Many alcohol brands incorporate cultural elements into their promotions [42]. Similar to previous studies, this is also the most commonly used marketing strategy in alcohol advertisements on Chinese television [43]. Previous studies have demonstrated that adapted cultural value appeals are more persuasive and attractive in advertisements [44]. On the other hand, alcohol advertising could further strengthen the "alcohol culture," enhance the alcogenic environment, and ultimately promote alcohol consumption.

This study extends the application of the AIDA model to alcohol advertising on the Douyin platform. Multivariate analysis revealed that when audiences are exposed to alcohol videos using experiential, brand, collaborative, and cultural marketing strategies, these videos are more likely to gain audience likes, thereby facilitating the transition from "attention" to "interest." To prevent the transition of audience attention to purchasing action, reducing the use of these marketing strategies in alcohol videos is crucial, thereby limiting the formation of "interest."

Age Restriction and Health Warning

Notably, many videos include cues related to teenagers' interests, such as idols, cartoons, and e-sports. Although there is currently no regulation explicitly requiring alcohol advertisements to include age restriction warnings, Douyin's platform policy clearly mandates that alcohol advertisements must contain warnings such as "Users under the age of 18 are prohibited from purchasing this product." [45] However, not all videos contain those age restriction signs. Although the presence of age restrictions may reduce the number of likes, the actual effectiveness of these restrictions remains unclear. It is generally easy to access social media platforms, and even children younger than 13 years can "legally" use social media and view alcohol advertisements, despite these age restrictions in place [46]. It is urgent to standardize the "teenage mode" on these social media platforms to mitigate the influence of alcohol marketing on teenagers.

Moreover, the lack of health warnings on social media platforms is a serious issue. Only 11.5% of the videos include any form of health warning, and most of these are ambiguous messages, "drink responsibly." In fact, "drink responsibly" messages are associated with increased alcohol consumption [47]. This is because such messages tend to enhance a prodrinking social norm. When "responsible drinking" messages are placed in alcohol advertisements, the alcohol industry can give the impression of fulfilling corporate responsibility without decreasing sales [48,49]. As reported in previous studies, only explicit health warnings that inform consumers about the carcinogenic effects of alcohol have a significant effect, rather than ambiguous messages [49]. Consequently, the 2022‐2030 WHO Global Alcohol Action Plan has called for ensuring that the labeling of alcoholic beverages is appropriate [50] and that essential health protection information is displayed.

Policy Implications

Although China's Advertising Law stipulates that all alcohol advertisements must not encourage drinking, depict drinking behaviors, or emphasize the functional benefits of alcohol consumption, and additionally prohibits the dissemination of alcohol advertisements through mass media targeting minors [51], alcohol marketing remains highly prevalent on social media platforms such as Douyin, where minors can still be easily exposed to such content. This suggests that deficiencies in both the legal provisions on alcohol advertising and their enforcement remain.

With respect to laws governing alcohol advertising, first, the definition of alcohol advertising should be refined to cover not only direct advertisements but also embedded stealth advertising, such as short skits and placements. Thus, legislative restrictions should include a comprehensive ban on all forms of alcohol commercial communication, recommendation, or activity. While these factors may not directly encourage drinking, they could help enhance the alcogenic environment. Moreover, the definition of "media targeting minors" remains ambiguous, making it possible for minors to be exposed to alcohol advertisements through general-audience channels, especially social media [43]. It is imperative to either explicitly prohibit alcohol advertising on social media or establish a robust "teenage mode." With respect to age restriction warnings and health warnings for alcohol advertising, mandatory regulations should not be limited to platforms such as Douyin but must be incorporated into advertising laws and strictly enforced. Finally, health warnings should be provided to avoid vague expressions such as "drink responsibly" and instead provide specific examples of the health risks associated with alcohol consumption.

With respect to law enforcement, the responsibility for monitoring and enforcing the existing regulations on advertising in China lies primarily with only the current commercial administrative department. In the future, other administrative departments, including health, food, and drug departments, could also participate to establish a comprehensive enforcement network [21]. Previous studies have demonstrated that the low penalties in the case of violation and the lack of effective detection are also key factors hindering its effective enforcement [21]. Enhancing penalty severity, improving regulatory channels, and encouraging public participation in supervision can significantly strengthen the enforcement of the law. With comprehensive legislation and implementation of restrictions, reducing the alcogenic environment caused by alcohol marketing is among the most cost-effective ways.

Limitations

The data collection period for this study (2021) may present certain limitations. With the rapid evolution of the digital marketing environment, the promotional strategies of alcohol brands on social media platforms, such as Douyin, may have become more covert and innovative. In recent years, emerging approaches such as algorithm-driven personalized content placement, interactions with virtual spokespersons, and cross-platform integrated marketing campaigns have increased in prevalence. These strategies are often integrated more deeply into users' everyday browsing experiences, further blurring the boundary between commercial promotion and organic content. Future monitoring of alcohol advertising should focus on such

XSL•FO
RenderX

covert marketing techniques to mitigate the potential impact of alcogenic environments.

There are also several limitations of this study. First, this study only gathered alcohol advertisements from Douyin over a 1-year period, failing to gather data over an extended period to assess the changing trends in marketing strategies over the years. Future studies could conduct longer-term longitudinal research to analyze temporal trends of marketing strategies in alcohol advertising. Second, like counts were used as the only variable representing popularity. The specific expressions and interactions of the audience in the comment section were not included and need further exploration. Despite these limitations, this study lays a foundation for future research on alcohol marketing content on Chinese social media platforms and provides evidence for strengthening the regulation of alcohol marketing on social media platforms.

Conclusions

In conclusion, this study summarized several alcohol marketing strategies on the Douyin platform, including experiential marketing, brand building, cross-border collaboration, and cultural connection. These strategies may enhance video attractiveness and appeal to teenagers, serving as key factors in transforming “attention” into “interest” (2 basic elements in the AIDA model) of alcohol advertisements. However, effective age restrictions and explicit health warnings are rarely shown in these alcohol-related Douyin videos. Urgent actions should include closing existing legal loopholes, such as refining the definition of alcohol advertising, strengthening protections for minors, and requiring specific health warnings, along with enhancing multiagency collaboration and imposing stricter penalties to decrease the alcogenic environment.

Acknowledgments

The authors are grateful to all study participants for their participation.

Funding

This work was supported by the National Natural Science Foundation of China (72474052).

Data Availability

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

Authors' Contributions

PZ, YZ, and LZ conceptualized and designed the study. YZ and LZ were responsible for data curation and formal analysis under the supervision of PZ. YZ wrote the original draft. CQ, WG, WZ, and PZ contributed to reviewing and editing the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The primary coding book of alcohol videos on Douyin platform.
[DOCX File, 19 KB - infodemiology_v6i1e74221_app1.docx]

References

1. Alcohol. WHO. URL: https://www.who.int/news-room/fact-sheets/detail/alcohol [accessed 2025-09-09]
2. A health promotion approach for reducing youth exposure to alcogenic environments: snapshot series on alcohol policies and practice. WHO. 2023. URL: https://www.who.int/publications/i/item/9789240073296 [accessed 2025-09-09]
3. Hill KM, Foxcroft DR, Pilling M. “Everything is telling you to drink”: understanding the functional significance of alcogenic environments for young adult drinkers. Addict Res Theory 2018 Nov 2;26(6):457-464. [doi: 10.1080/16066359.2017.1395022]
4. Huckle T, Huakau J, Sweetsur P, Huisman O, Casswell S. Density of alcohol outlets and teenage drinking: living in an alcogenic environment is associated with higher consumption in a metropolitan setting. Addiction 2008 Oct;103(10):1614-1621. [doi: 10.1111/j.1360-0443.2008.02318.x] [Medline: 18821871]
5. Murphy A, Roberts B, Ploubidis GB, Stickley A, McKee M. Using multi-level data to estimate the effect of an “alcogenic” environment on hazardous alcohol consumption in the former Soviet Union. Health Place 2014 May;27:205-211. [doi: 10.1016/j.healthplace.2014.02.015] [Medline: 24662529]
6. Petticrew M, Shemilt I, Lorenc T, et al. Alcohol advertising and public health: systems perspectives versus narrow perspectives. J Epidemiol Community Health 2017 Mar;71(3):308-312. [doi: 10.1136/jech-2016-207644] [Medline: 27789756]

XSL•FO
RenderX

7. Stockings E, Bartlem K, Hall A, et al. Whole-of-community interventions to reduce population-level harms arising from alcohol and other drug use: a systematic review and meta-analysis. Addiction 2018 Nov;113(11):1984-2018. [doi: 10.1111/add.14277] [Medline: 29806876]
8. Moreno MA, Christakis DA, Egan KG, Brockman LN, Becker T. Associations between displayed alcohol references on Facebook and problem drinking among college students. Arch Pediatr Adolesc Med 2012 Feb;166(2):157-163. [doi: 10.1001/archpediatrics.2011.180] [Medline: 21969360]
9. Monitoring and restricting digital marketing of unhealthy products to children and adolescents. WHO. URL: https://www.who.int/europe/activities/monitoring-and-restricting-digital-marketing-of-unhealthy-products-to-children-and-adolescents [accessed 2025-09-09]
10. Lobstein T, Landon J, Thornton N, Jernigan D. The commercial use of digital media to market alcohol products: a narrative review. Addiction 2017 Jan;112(S1):21-27. [doi: 10.1111/add.13493] [Medline: 27327239]
11. Bouckley B. Anheuser-busch: facebook beats any US broadcast network for consumer reach. Beverage Daily. 2013. URL: https://www.beveragedaily.com/Article/2013/11/12/Anheuser-Busch-Facebook-consumer-reach-beats-any-US-broadcast-network/ [accessed 2025-09-09]
12. Cranwell J, Opazo-Breton M, Britton J. Adult and adolescent exposure to tobacco and alcohol content in contemporary YouTube music videos in Great Britain: a population estimate. J Epidemiol Community Health 2016 May;70(5):488-492. [doi: 10.1136/jech-2015-206402] [Medline: 26767404]
13. Manthey J, Shield KD, Rylett M, Hasan OSM, Probst C, Rehm J. Global alcohol exposure between 1990 and 2017 and forecasts until 2030: a modelling study. Lancet 2019 Jun 22;393(10190):2493-2502. [doi: 10.1016/S0140-6736(18)32744-2] [Medline: 31076174]
14. Peng C. Digital transformation and new retail model exploration of liquor enterprise marketing [Article in Chinese]. Enterp Reform Manag 2021(17):112-113. [doi: 10.13768/j.cnki.cn11-3793/f.2021.1744]
15. China liquor consumption insight report 2021. Tencent News. 2021. URL: https://jiu.ifeng.com/c/89c8yf4DYxr [accessed 2025-09-09]
16. 2024 Douyin user portrait and audience analysis. TikTok. 2025. URL: https://www.jingliren.org/rdxx/hlw/311578.html [accessed 2025-09-09]
17. #Chinese liquor. Tiktok. URL: https://www.douyin.com/hashtag/1583910072238157 [accessed 2025-09-09]
18. White book on Chinese liquor consumption 2023. Chinese Alcoholic Drinks Association. 2023. URL: https://file.tencentads.com/web/pdf/index/f90b46878551cdba [accessed 2025-09-09]
19. Tristanto TA, Hurriyati R. AIDA model as a marketing strategy to influence consumer buying interest in the digital age. BIRCI 2023;4:12575-12586. [doi: 10.33258/birci.v4i4.3319]
20. Chan RHW, Dong D, Chong MKC, et al. Alcohol social media marketing and drinking behaviors among Chinese young adults: Mediation by drinking expectancies. Drug Alcohol Depend 2025 Oct 1;275:112818. [doi: 10.1016/j.drugalcdep.2025.112818] [Medline: 40769088]
21. Ji N, Xu Q, Zeng X, Casswell S, Bai Y, Liu S. Alcohol advertising exposure and drinking habits among Chinese adolescents in 2021: a national survey. Am J Public Health 2024 Aug;114(8):814-823. [doi: 10.2105/AJPH.2024.307680] [Medline: 38870435]
22. China liquor industry development research report 2021. iiMedia research. 2021. URL: https://baijiahao.baidu.com/s?id=1718016690969263242&wfr=spider&for=pc [accessed 2025-09-09]
23. JD Global. URL: https://global.jd.com/ [accessed 2025-12-31]
24. The rise of Douyin and TikTok: the commercially accelerating short video empire after 2020. Vzkoo. 2024. URL: https://www.vzkoo.com/read/20241112138 85f4bdbc1f915258ac52d.html [accessed 2025-12-05]
25. Barker AB, Bal J, Murray RL. A content analysis and population exposure estimate of Guinness branded alcohol marketing during the 2019 Guinness Six Nations. Alcohol Alcohol 2021 Aug 30;56(5):617-620. [doi: 10.1093/alcalc/agab039] [Medline: 34080614]
26. Atkinson AM, Sumnall H, Meadows B. “We’re in this together”: a content analysis of marketing by alcohol brands on Facebook and Instagram during the first UK Lockdown, 2020. Int J Drug Policy 2021 Dec;98:103376. [doi: 10.1016/j.drugpo.2021.103376] [Medline: 34364199]
27. Notice on issuing the measures for ethical review of human life science and medical research. China’s National Health Commission, Ministry of Education, and Ministry of Science and Technology. 2023. URL: https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm [accessed 2025-12-22]
28. TikTok user profile report 2024. CSDN. 2024. URL: https://blog.csdn.net/2301_76223496/article/details/137100176?spm=1001.2101.3001.6650.3 [accessed 2025-09-09]
29. GE Lianying LR, Ning LI, Jing ZHU, Dan XU. Investigation and analysis on drinking culture of youth [Article in Chinese]. Modern Food 2022;28(21):224-228. [doi: 10.16736/j.cnki.cn41-1434/ts.2022.21.053]
30. Mid-2025 research report on China’s baijiu market. : Chinese Alcoholic Drinks Association; 2025 URL: https://assets.kpmg.com/content/dam/kpmg/cn/pdf/zh/2025/06/mid-term-research-report-on-the-chinese-baijiu-market-2025.pdf [accessed 2025-09-09]

XSL•FO
RenderX

31. Barry AE, Padon AA, Whiteman SD, et al. Alcohol advertising on social media: examining the content of popular alcohol brands on Instagram. Subst Use Misuse 2018 Dec 6;53(14):2413-2420. [doi: 10.1080/10826084.2018.1482345] [Medline: 29889647]
32. Hendriks H, Wilmsen D, van Dalen W, Gebhardt WA. Picture me drinking: alcohol-related posts by Instagram influencers popular among adolescents and young adults. Front Psychol 2019;10:2991. [doi: 10.3389/fpsyg.2019.02991] [Medline: 32038379]
33. Guégan E, Zenone M, Mialon M, Gallopel-Morvan K. #Bartender: portrayals of popular alcohol influencer's videos on TikTok©. BMC Public Health 2024 May 23;24(1):1384. [doi: 10.1186/s12889-024-18571-1] [Medline: 38783213]
34. Mart SM. Alcohol marketing in the 21st century: new methods, old problems. Subst Use Misuse 2011;46(7):889-892. [doi: 10.3109/10826084.2011.570622] [Medline: 21599504]
35. Cowley E, Barron C. When product placement goes wrong: the effects of program liking and placement prominence. J Advert 2008 Apr;37(1):89-98. [doi: 10.2753/JOA0091-3367370107]
36. Eagle L, Dahl S. Product placement in old and new media: examining the evidence for concern. J Bus Ethics 2018 Feb;147(3):605-618. [doi: 10.1007/s10551-015-2955-z]
37. D'Angelo J, Kerr B, Moreno MA. Facebook displays as predictors of binge drinking: from the virtual to the visceral. Bull Sci Technol Soc 2014;34(5-6):159-169. [doi: 10.1177/0270467615584044] [Medline: 26412923]
38. Beukeboom CJ, Kerkhof P, De Vries M. Does a virtual like cause actual liking? How following a brand's Facebook updates enhances brand evaluations and purchase intention. J Interact Mark 2015 Nov;32(1):26-36. [doi: 10.1016/j.intmar.2015.09.003]
39. Toldos-Romero MP, Orozco-Gómez MM. Brand personality and purchase intention. Eur Bus Rev 2015 Aug 10;27(5):462-476. [doi: 10.1108/EBR-03-2013-0046]
40. Jernigan D, Ross CS. The alcohol marketing landscape: alcohol industry size, structure, strategies, and public health responses. J Stud Alcohol Drugs Suppl 2020 Mar(19):13-25. [doi: 10.15288/jsads.2020.s19.13] [Medline: 32079559]
41. Jernigan DH, Padon A, Ross C, Borzekowski D. Self-reported youth and adult exposure to alcohol marketing in traditional and digital media: results of a pilot survey. Alcohol Clin Exp Res 2017 Mar;41(3):618-625. [doi: 10.1111/acer.13331] [Medline: 28219114]
42. Brodmerkel S, Carah N. Alcohol brands on Facebook: the challenges of regulating brands on social media. J Public Aff 2013 Aug;13(3):272-281. [doi: 10.1002/pa.1466]
43. Tang Y, Lei N, Hu D, et al. Estimated exposure to televised alcohol advertisements among children and adolescents. JAMA Netw Open 2025 Jul 1;8(7):e2521819. [doi: 10.1001/jamanetworkopen.2025.21819] [Medline: 40674047]
44. Hornikx J, Janssen A, O'Keefe DJ. Cultural value adaptation in advertising is effective, but not dependable: a meta-analysis of 25 years of experimental research. Int J Bus Commun 2023. [doi: 10.1177/23294884231199088]
45. 【Alcoholic beverages】product release and promotion guidelines. The Douyin E-commerce Operations Team. 2025. URL: https://school.jinritemai.com/doudian/web/article/aHSh8Zmd66rx#:~:text [accessed 2025-12-05]
46. Barry AE, Johnson E, Rabre A, Darville G, Donovan KM, Efunbumi O. Underage access to online alcohol marketing content: a YouTube case study. Alcohol Alcohol 2015 Jan;50(1):89-94. [doi: 10.1093/alcalc/agu078] [Medline: 25411395]
47. Moss AC, Albery IP, Dyer KR, et al. The effects of responsible drinking messages on attentional allocation and drinking behaviour. Addict Behav 2015 May;44:94-101. [doi: 10.1016/j.addbeh.2014.11.035] [Medline: 25577316]
48. Maani Hessari N, Petticrew M. What does the alcohol industry mean by "Responsible drinking"? A comparative analysis. J Public Health 2018 Mar 1;40(1):90-97. [doi: 10.1093/pubmed/fdx040] [Medline: 28398571]
49. Davies E, Lewin J, Field M. Am I a responsible drinker? The impact of message frame and drinker prototypes on perceptions of alcohol product information labels. Psychol Health 2024 Aug;39(8):1005-1022. [doi: 10.1080/08870446.2022.2129055] [Medline: 36190181]
50. Global alcohol action plan 2022-2030. WHO. 2024. URL: https://www.who.int/publications/i/item/9789240090101 [accessed 2025-09-09]
51. Advertisement Law of the People's Republic of China. State Administration for Market Regulation. 2021. URL: https://www.samr.gov.cn/zw/zfxxgk/fdzdgknr/fgs/art/2023/art_5474cf75173c45d6a0379730fb4e8d97.html [accessed 2025-09-09]

Abbreviations

AIDA: attention-interest-desire-action
OR: odds ratio
WHO: World Health Organization

XSL•FO
RenderX

Edited by T Mackey; submitted 20.Mar.2025; peer-reviewed by A Pal, Y Yang; accepted 09.Dec.2025; published 06.Jan.2026.

Please cite as:
Zhao Y, Zhang L, Qian C, Guo W, Zhu W, Zheng P
Marketing Strategies and Factors Influencing the Popularity of Alcohol Videos from Official Brand Accounts on Douyin: Content Analysis Study
JMIR Infodemiology 2026;6:e74221
URL: https://infodemiology.jmir.org/2026/1/e74221
doi:10.2196/74221

© Yuchen Zhao, Lingyun Zhang, Chenyu Qian, Wenjie Guo, Weiyun Zhu, Pinpin Zheng. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 6.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Artificial Intelligence Methods to Evaluate the Effect of the National Cytomegalovirus Awareness Month on the Content and Sentiment of Social Media Posts: Infodemiology Study

Tracy R Rosebrock[1], PhD, MPH; Zhen Yang[2], MS, MScAc; Lauren D'Arco[1], BA; Tapan Pathak[2], MS; Rebecca Vislay-Wade[2], MA, MS, PhD; Karen Fowler[3], MS, PhD; John Diaz-Decaro[2], MS, PhD; Colin Kunzweiler[2], MS, PhD

[1]Department of Health Science, Department of Biology, School of Arts and Sciences, Stonehill College, North Easton, MA, United States

[2]Moderna Therapeutics (United States), Cambridge, MA, United States

[3]Department of Pediatrics, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Tracy R Rosebrock, PhD, MPH
Department of Health Science, Department of Biology
School of Arts and Sciences
Stonehill College
Shields Science Center
320 Washington Street
North Easton, MA, 02357
United States
Phone: 1 5085651097
Email: trosebrock@stonehill.edu

Abstract

Background: The month of June has been recognized as the National Cytomegalovirus (CMV) Awareness Month since 2011 in the United States. Established by government resolution, the goal is to increase awareness and reduce the incidence of congenital CMV infection, a leading cause of preventable birth defects and developmental disabilities. Social media is a powerful tool to support public health by making health information easily accessible. With an estimated 246 million users in the United States and more than half of adults seeking health information through such platforms, social media offers an unparalleled opportunity to promote CMV awareness and prevention.

Objective: This study aimed to evaluate social media messaging before, during, and after the National CMV Awareness Month to assess how the campaign influenced messaging patterns and sentiment related to specific CMV health topics.

Methods: Publicly available posts on Twitter/X from May to August 2023 that contained at least one of the five most used CMV-related hashtags were collected using a media monitoring platform. The dataset was preprocessed using a customized Bidirectional Encoder Representations from Transformers tokenizer and a language detection package to remove irrelevant and non-English posts. Validated and artificial intelligence (AI) methods (Cohen κ=0.69) were used to determine the thematic content of posts (N=14,900), such as awareness and prevention messaging, and to characterize the sentiment. Changes in post characteristics were measured in relation to the National CMV Awareness Month.

Results: CMV-relevant post volume increased by 55% during the campaign month and returned to precampaign levels in July. Overall, academic/university researchers were the most frequent authors, pediatrics was the most frequent population discussed, and vaccines were the most frequently mentioned prevention. Significant associations were observed between the month of post publication and the target audience (χ^2_2=144.3, *P*<.001), awareness or prevention messaging (χ^2_2=107.8, *P*<.001), and post sentiment (χ^2_4=163.6, *P*<.001). The intended audience of posts shifted toward the general population from scientists/health care professionals during the campaign month (adjusted Pearson residuals, *P*=.009). Awareness messaging increased in June 2023, particularly in relation to CMV transmission and disease burden, while prevention messaging decreased (adjusted Pearson residuals, *P*=.008). Finally, although posts were generally neutral in sentiment, a significant shift occurred toward a positive sentiment during the campaign month (adjusted Pearson residuals, *P*=.006), a sentiment that was more likely to engage the user (Kruskal-Wallis; χ^2_2=194.31, *P*<.001).

XSL•FO
RenderX

Conclusions: The National CMV Awareness Month in 2023 shifted the digital CMV conversation toward public-facing messaging and raised awareness efforts. Although posts related to CMV prevention generally conveyed a positive sentiment, prevention messaging declined during the campaign. These findings highlight opportunities for future CMV social media initiatives to balance awareness with prevention through evaluation and strategic design using AI models to strengthen CMV public health communication and engagement.

(JMIR Infodemiology 2026;6:e80922) doi:10.2196/80922

KEYWORDS

cytomegalovirus; social media; public health, health communication; sentiment analysis; artificial intelligence

Introduction

Background

Human cytomegalovirus (CMV) is a ubiquitous beta-herpesvirus with an estimated seroprevalence of 83% worldwide and 63% in the United States and a disease burden that disproportionally impacts disadvantaged and minoritized communities [1-4]. The virus is transmitted through direct contact with infectious body fluids, through organ/stem cell transplants, and transplacentally, which can result in congenial CMV infection [5,6]. Infection with CMV is lifelong and is typically subclinical in healthy individuals, though CMV may silently contribute to chronic conditions, such as cardiovascular disease [7], cognitive decline [8], neurologic disorders [9,10], anxiety and depression [11], immunosenescence [12], Guillain-Barré syndrome [13], certain malignancies (eg, glioblastoma multiforme) [14], and all-cause mortality [15]. In those who are immunocompromised, CMV infection can lead to acute disease and death [16,17].

When CMV is transmitted from mother to fetus during pregnancy, it is referred to as congenital cytomegalovirus (cCMV). In the United States, 1 in 200 infants is born with cCMV [18]. The majority of infants born with cCMV are asymptomatic without recognizable signs or symptoms at birth, while a subset (~10%) may be impacted by a wide range of signs and symptoms at birth [19]. Regardless of the presence or absence of symptoms at birth, all infants born with cCMV are at risk for developing long-term sequelae, such as sensorineural hearing loss, vision impairment, cerebral palsy, and developmental delays [20]. CMV is also a recognized cause of stillbirth and intrauterine fetal demise [21-24]. Despite the significant burden of cCMV [25,26], just 13% of women are aware of CMV, cCMV, or simple hygiene prevention practices that can reduce risk [27]. Given the limited effectiveness of pharmaceutical interventions and the lack of a licensed vaccine to prevent infection, public health education is of critical importance.

In 2011, the US Senate issued a resolution declaring June to be the National CMV Awareness Month, with explicit objectives to raise awareness of the dangers of CMV and to reduce the incidence of cCMV infections through education [28]. Today, eHealth, or the delivery of health information digitally, is recognized as an important frontier for public health education [29]. Social media is an important tool for public health education in the United States, where ~246 million people, or 73% of the population, are active users [30]. Among US adults, 55% report using social media at least occasionally to seek health information, and for the ~95 million US users of Twitter/X, the average daily use is greater than 30 minutes per day on the platform [31,32]. These platforms offer a valuable opportunity to disseminate accurate information and raise awareness about cCMV and prevention.

The National CMV Awareness Month is primarily promoted by the National Cytomegalovirus Foundation (NCMVF), advocacy groups of families affected by cCMV, and state and national public health agencies [33,34]. Shared goals are to increase public awareness and educate on prevention behaviors, while advocacy groups also promote expanded CMV screening and research. The awareness campaign is visible on social media through educational graphics, family stories, and hashtags, such as #stopCMV and #CMVawareness, to spread information and engage the public. Although CMV-specific campaign activity is described by advocacy organizations, prior public health infodemiological research has not systematically evaluated how these campaigns function on social media platforms.

Natural language processing (NLP) and sentiment analysis are novel, powerful, and essential tools in public health for monitoring public opinions toward health-related topics and identifying potential areas and concerns. For example, recent research has used Bidirectional Encoder Representations from Transformers (BERT) to analyze social media posts to understand public opinions toward the impact of the COVID-19 pandemic on social life [35] and sentiments expressed during an outbreak in Uganda [36]. Despite coordinated social media campaigns dedicated to CMV awareness, to date, sentiment analysis has not been conducted regarding this disease.

Beyond BERT-based approaches, recent infodemiology studies have increasingly incorporated large language models (LLMs) to analyze high-volume Twitter/X data for public health insight. For example, artificial intelligence (AI) models have been used to classify tweets related to conjunctivitis outbreaks and estimate epidemic signals, and LLM-driven sentiment and substance-use detection models have been applied to opioid-related social media data [37,38]. Such studies demonstrate the expanding role of generative and transformer-based models in characterizing eHealth discourse, providing methodological precedent for applying LLMs to evaluate public health messaging on Twitter/X.

Although survey-based studies have assessed CMV awareness and attitudes toward prenatal and neonate CMV screening [39,40], formal infodemiology or eHealth investigations, particularly those evaluating social media messaging related to CMV awareness and prevention, have not been conducted. This study represents the first systematic evaluation of CMV-related

XSL•FO
RenderX

discourse on social media, characterizing Twitter/X messaging during the National CMV Awareness Month and assessing how this communication aligns with goals set forth by the US Senate and public health stakeholders. Although prior infodemiology research has analyzed other public health campaigns or disease-related discourse using transformer models and LLMs, no studies have examined CMV or cCMV infections, and none have assessed a nationally recognized awareness-month campaign. By leveraging an LLM to classify, summarize, and evaluate campaign-related tweets, this study used emerging LLM-based methods to fill an important gap in the literature on digital CMV education and awareness.

Study Objective

The aim of this study was to evaluate the messaging of CMV-related posts on Twitter/X before, during, and after the National CMV Awareness Month in June 2023 (ie, May-August 2023).

Methods

Data Aggregation

Investigators collected social media posts from the Twitter/X platform (Twitter became X on July 23, 2023) for May-August 2023, representing the months immediately preceding and following the National CMV Awareness Month in June using Keyhole, a subscription-based, commercial social listening and analytics platform. Keyhole provides real-time tracking, historical backfill, and reach/impression estimates based on post volume and author follower counts.

To be eligible for this initial download, social media posts must have included one or more of the following five hashtags: #stopCMV, #cCMV, #CMV, #CMVawareness, and #cytomegalovirus. No additional filters were applied. These hashtags were identified by two independent reviewers, who manually assessed Twitter/X posts referencing "cytomegalovirus" or "CMV" using the site's search function. The most recent posts were assessed weekly from March 1 to April 30, 2023. From relevant posts that included hashtags, the five hashtags used in this study were the most frequently used and captured all posts in the assessment window that used hashtags.

In addition to the social media post text, user information (eg, username, biography, number of followers, and location) and dissemination-related metrics (eg, number of reposts, number of likes, and number of comments) were downloaded for each post. More than 30,000 social media posts were downloaded based on the aforementioned criteria.

Data Preparation

Non-English language posts were detected using the language-detecting Python package, *langdetect*, and removed during preliminary data-cleaning procedures. Among the more than 20,000 English-language social media posts that remained, posts were first cleaned to remove noise and other nuisance text (eg, URLs, emojis, and hashtag [#] and mention [@] symbols). A fine-tuned text classifier was next applied to identify relevant social media posts and exclude irrelevant social media posts from analyses [41]. Briefly, a customized BERT tokenizer was used to complete word-level tokenization of all social media posts. Reading social media posts from both left and right, the BERT tokenizer can understand the language context and flow of words based on a given word's surroundings. By default, BERT tokenizes, or breaks down, words into subword units. For example, the term "congenital" may be segmented into the subword units "con," "##gen," and "##ital" (the "##" prefix indicates that the subword is a continuation of the previous token unit). However, during exploratory data analysis, it was noticed that certain keywords, and in particular abbreviations, were segmented into individual characters (eg, "CMV" was segmented into "C," "##M," and "##V). To ensure meaningful pretraining, the BERT tokenizer was customized by manually adding a dictionary of domain-specific words (eg, "congenital," "cCMV," and "CMV") that would be identified as complete words, as opposed to subword units. We compared model performance using the default bert-base-cased tokenizer versus the customized tokenizer, each fine-tuned on the same manually labeled dataset. Both models achieved comparable accuracy (95%), indicating that the custom dictionary did not materially affect overall performance. Additional methodological details are provided in Multimedia Appendix 1.

Descriptive Analyses

Numerous descriptive analyses were conducted. The date of social media posts was analyzed to describe the number of posts per month for the May-August 2023 period. The location of each relevant, English-language social media post (when such information was available) was summarized at the country level (data not shown), as well as at the state level for posts originating in the United States. User categories (eg, university/academic researchers, news/journal/public health/education, and physicians/health care/hospitals) were summarized for the top 20 users with the most social media posts or most followers during the study period. The biography data supplied in the author's Twitter/X biosketch were used to determine their user category, or if a biography statement was not provided, the author was located using a web search and categorized using available data. Metrics of impact, including the number of followers, reposts, and "likes," were also summarized.

Theme (Aspect) Classification

Following the identification of relevant social media posts, investigators developed a master prompt for Moderna's internal ChatGPT AI tool (mChat), as shown in Multimedia Appendix 2, to annotate the specific aspects contained within each post. mChat serves as a pass-through to the OpenAI application programming interface (API), does not modify model behavior, and directly calls the OpenAI API using company credentials.

Aspects were first prespecified by two independent investigators and were grouped according to four broad categories: population discussed (the subject of the social media post; eg, "women of reproductive age," "parents," and "pediatric patients"); awareness and knowledge (a list of clinical and nonclinical terms related to CMV; eg, "seroconversion," "newborn screening," "National CMV Foundation," and "parental education"); prevention (a list of terms describing preventive

XSL•FO
RenderX

measures related to CMV; eg, "hygiene measures," "antiviral treatment," and "vaccines"); and general CMV information (generic terms; eg, "safety," "efficacy," and "tolerability"). In addition, the perceived target audience (eg, general population, scientists/health care professionals) was annotated (see Table S1 in Multimedia Appendix 3 for a complete list of prespecified aspects within each thematic category). Investigators manually classified prespecified aspects for approximately 90 social media posts. The ChatGPT model then used the manually classified social media posts for few-shot learning for aspect classification. Although 90 manually classified posts represented a small subset of the >10,000 posts analyzed, the approach ensured complete coverage of all prespecified aspects with representation in at least 2 example posts, making the number of manually classified posts commensurate with the list of aspects of interest (Table S1 in Multimedia Appendix 3).

To load the full dataset stored in a Microsoft Excel spreadsheet, the tweets were imported into data frames using the Python *pandas* library, and only the tweet-text column was used as input for annotation. The master prompt (Multimedia Appendix 2) contained (1) step-by-step instructions for identifying CMV-related aspects, segmenting text, and assigning sentiment, while maintaining aspect integrity, and (2) the set of approximately 90 manually annotated tweets used as few-shot examples. The model temperature was set to 0 to promote deterministic outputs. Responses were requested in JSON format (keys included aspects, aspect segment, sentiment toward each aspect, and overall sentiment; see sentiment methods in the *Sentiment and other Statistical Analyses* section). Each tweet was processed with up to five retries to ensure valid JSON formatting. No manual prompt revisions were made between batches, and no additional model fine-tuning was performed. Reproducibility was supported by fixing model parameters (temperature=0), using a single master prompt for all tweets, processing tweets independently, and allowing up to five retries for valid JSON output. To enhance transparency, the prompt instructed ChatGPT to include additional keys in the JSON output, identifying the specific text segments that contributed to each aspect or sentiment label. This allowed investigators to trace which words or phrases informed each annotation. Because annotations were produced via prompt-based generation rather than model training, no explicit class-balancing techniques were applied.

Sentiment and Other Statistical Analyses

Social media posts and specific aspects identified within the posts were assigned a sentiment (ie, positive, neutral, or negative). Four independent, blinded reviewers assigned a sentiment to 97 randomly selected post texts outside of the dataset with moderate agreement, with an interrater reliability score of 0.56 (Fleiss κ; "moderate" defined as 0.41-0.6 [42]). The ChatGPT model was then provided with the scored posts for few-shot learning and asked to assign a sentiment to social media posts included in the analysis dataset. Following assignment, text from 50 posts scored by ChatGPT was evaluated for sentiment by the same four independent and blinded reviewers. Moderate interrater reliability was measured at 0.51 (Fleiss κ). The interrater reliability score between the sentiment assigned by most reviewers and ChatGPT was 0.69 (Cohen κ), indicating substantial agreement between reviewers and ChatGPT ("substantial" defined as 0.61-0.8 [42]). Of the 10 posts, or 20%, where human-AI agreement was not observed, 90% of posts were scored neutral by one party and either positive or negative by the other, indicating that minor discrepancies rather than large errors (positive versus negative) explain the discordance.

Statistical analyses (Fleiss κ, Cohen κ, chi-square tests with Bonferroni correction, Kruskal-Wallis with Bonferroni correction) were conducted using Excel or Datatab. Chi-square tests were applied only to independent categorical variables. Individual posts often contained multiple dependent variables (eg, multiple hashtags; multiple aspects within a category, such as target audience, awareness, or prevention messaging). These attributes were excluded from tests requiring variable independence. Adjusted Pearson residuals were calculated for chi-square cross-tabulations, and Bonferroni correction was applied to adjust *P* values for multiple comparisons to set the critical threshold for significance.

Ethical Considerations

This study used only publicly available text from social media posts on Twitter/X. The research team did not interact with post authors or collect private information. All data were processed to remove or avoid inclusion of identifiable information and are reported only in aggregate. This study did not constitute human subjects research and therefore did not require review or approval by an institutional review board per Federal Regulations for the Protection of Human Research Subjects (45 CFR 46.104(d)(4)(i); [43]).

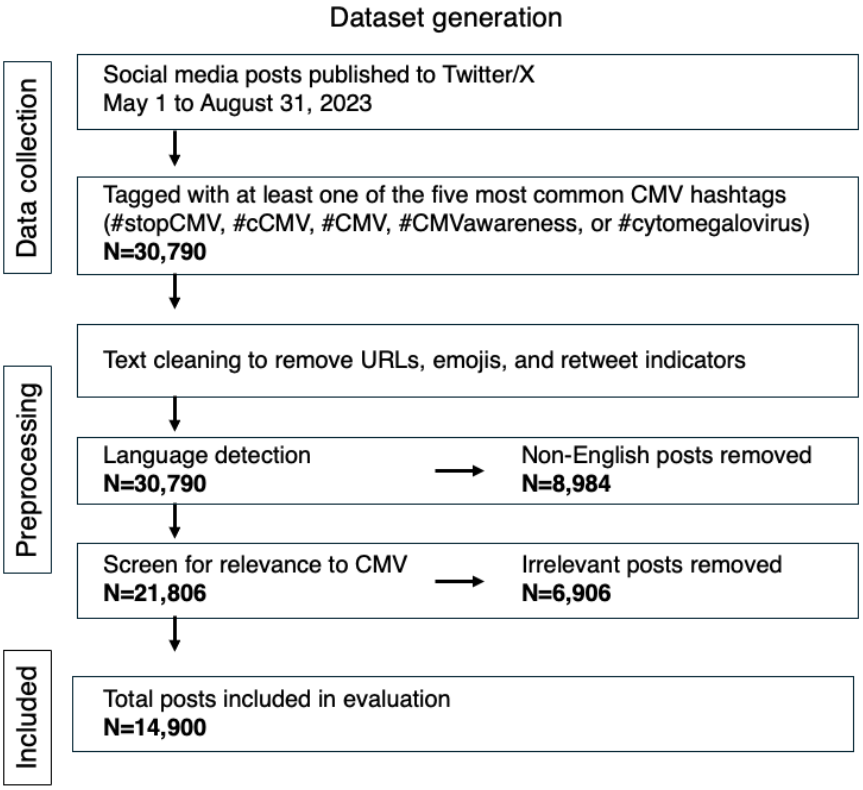
Results

Data Aggregation and Processing

Between May 1 and August 31, 2023, a total of 30,790 public posts published to Twitter/X were tagged with one or more of the five most common hashtags used to reference CMV or CMV disease. Several hundred posts were cotagged with hashtags unrelated to CMV (eg, #SHREKINU, a cryptocurrency) or used a CMV-related hashtag to indicate a non-CMV topic (eg, #CMV=commercial motor vehicle). Additionally, nearly one-third of posts were written in a language other than English. To focus subsequent analyses, the 30,790 posts were preprocessed to (1) remove noise elements, such as URLs and emojis; (2) filter out non-English posts; and (3) extract posts relevant to CMV or CMV disease (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of the dataset collection, preprocessing, and inclusion steps. cCMV: congenital cytomegalovirus; CMV: cytomegalovirus.



Descriptive Analyses

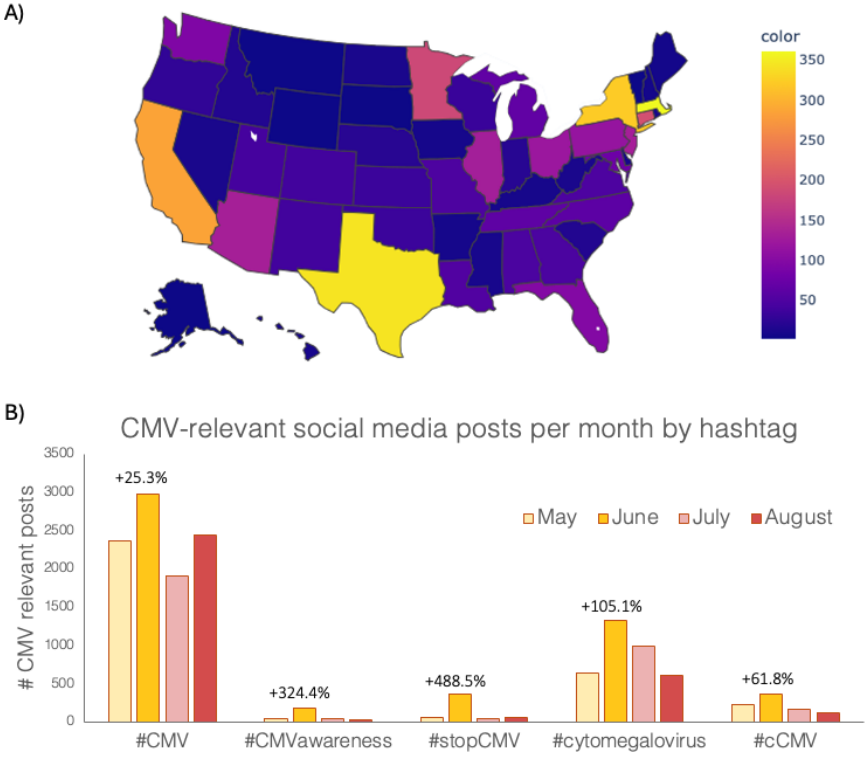
Overall, 14,900 posts were analyzed. There were 3336 (22.4%) social media posts in May 2023, which increased by 55% to 5180 (34.8%) posts in June 2023 (coinciding with the National CMV Awareness Month in the United States), and decreased during both July 2023 (n=3124, 21%) and August 2023 (n=3260, 21.9%), as shown in Table S2 in Multimedia Appendix 3. Social media posts most frequently originated from the United States (n=3858, 25.9%). The next most frequent countries included Canada (n=594, 4%), the United Kingdom (n=535, 3.6%), India (n=384, 2.6%), and Australia (n=370, 2.5%). In the United States, among posts with known geographic locations, Massachusetts, Texas, and New York had the highest number of posts (Figure 2A). A more detailed analysis of posts grouped by hashtag revealed additional trends. The hashtag #CMV was the most used (n=9674, 64.9%), followed by #cytomegalovirus (n=3558, 23.9%), #cCMV (n=859, 5.8%), #stopCMV (n=529, 3.6%), and #CMVawareness (n=280, 1.9%), as shown in Figure 2B. With respect to the National CMV Awareness Month, an increase in posts was observed from May to June for all five hashtags. The largest proportional increases occurred with #stopCMV (488.5%, from n=61, 0.4%, to n=359, 2.4%, posts) and #CMVawareness (+324.4%; from n=41, 0.3%, to n=174, 1.2%, posts), possibly driven by national advocacy organizations.

To understand who is publishing CMV-relevant content on social media, the authors of posts containing each CMV-related hashtag were analyzed. Because a single post may include multiple hashtags, authors could appear across several hashtag groups. For each hashtag, authors were ranked by the total number of posts they published, and the top 20 most frequent authors were identified. These authors were then categorized by author category based on their profile affiliation. To describe overall trends, the top 20 authors from each hashtag were combined into a single dataset, with duplicate users removed. The most common author category was “university/academic researchers,” followed by “news/journal/public health/education and “physicians/health care/hospitals” (Table S3 in Multimedia Appendix 3).

Hashtag-specific trends included the nearly equal representation of most CMV stakeholders under #CMV, in contrast to skewed author distributions for #CMVawareness and #cytomegalovirus. The top 20 authors scored by total followers were also assessed. Unsurprisingly, the most common follower category was “news/journal/public health/education” (Table S3 in Multimedia Appendix 3).

XSL•FO
RenderX

Figure 2. Descriptive statistics. Quantification of CMV-relevant public posts (A) published to Twitter/X at the US state level and (B) sorted by hashtag by month (May-August 2023). Noted percentages reference the change in post number from May to June (National CMV Awareness Month). cCMV: congenital cytomegalovirus; CMV: cytomegalovirus.



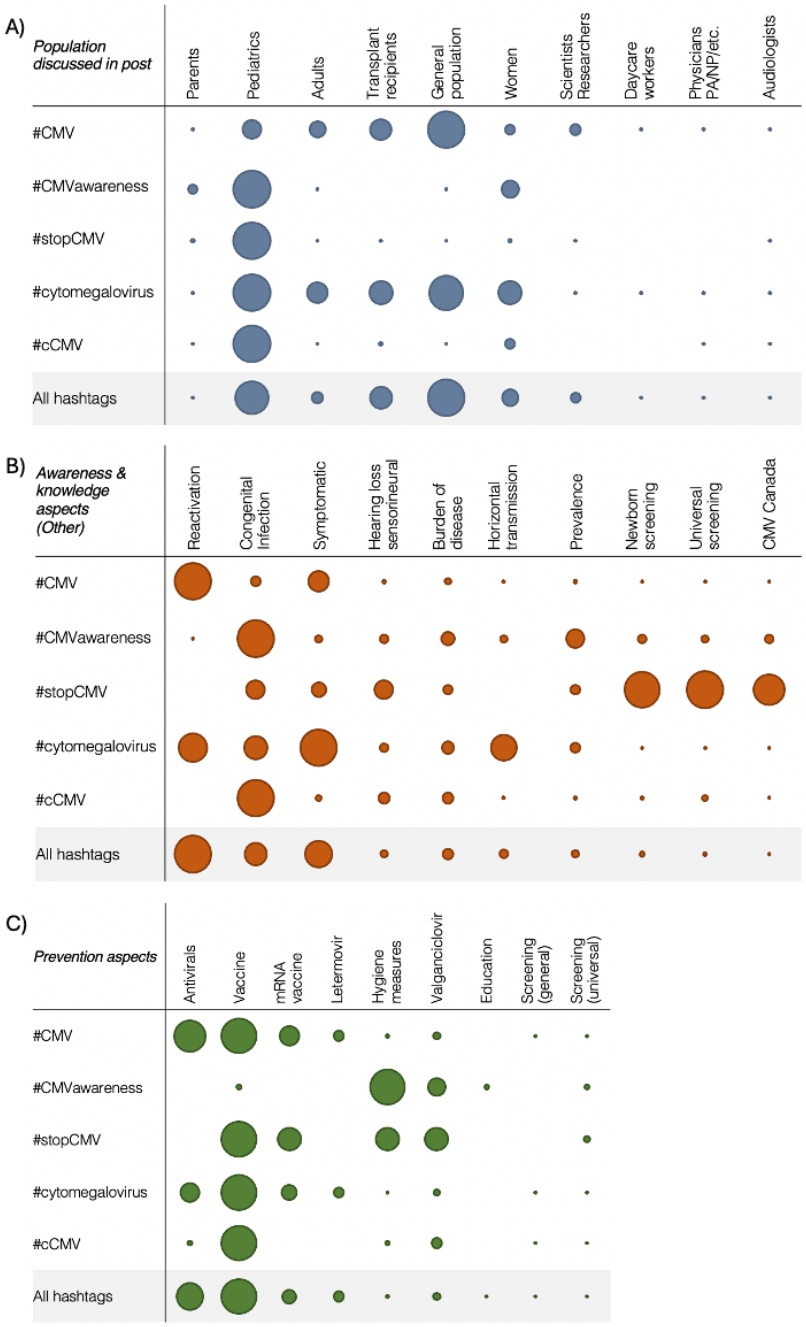
Classification of Thematic (Aspects) Content

Social media posts were also evaluated for thematic content. Five broad thematic categories (target audience, population discussed, awareness and knowledge, prevention, and general CMV information) were created and populated with specific aspects to be identified within the posts. Overall, the "scientific/health care professionals" category was the most frequent target audience of social media posts (n=7575, 50.8%); however, this varied by hashtag (Table S4 in Multimedia Appendix 3). The "general population" category was the most frequent population discussed across all five hashtags (n=2727, 18.3%), followed by "pediatrics" (n=2456, 16.5%), as shown in Figure 3A and in Table S5 in Multimedia Appendix 3. Transplant recipients (n=1636, 11%) were the third-most frequent population discussed, though they were nearly exclusively mentioned in #CMV and #cytomegalovirus posts. "Women" (n=1141, 7.7%), "adults" (n=784, 5.3%), and "scientists" (n=691, 4.6%) were also frequently discussed; all other prespecified aspects related to the "population discussed" category were identified in fewer than 2235 (15%) of the 14,900 social media posts evaluated.

Approximately 50 aspects were prespecified for the "awareness and knowledge" category (Table S1 in Multimedia Appendix 3). Perhaps expectedly, "CMV" was overwhelmingly the most frequently identified aspect (n=10,929, 73.3%, posts), with "cCMV" being the second-most frequently identified aspect (n=1598, 10.7%), as shown in Table S6 in Multimedia Appendix 3. The frequency of aspects other than "CMV" and "cCMV" varied by hashtag, with the third-most frequent aspect being "reactivation" for #CMV and #cytomegalovirus, "congenial infection" for #CMVawareness and #cCMV, and "newborn/universal screening" for #stopCMV (Table S7 in Multimedia Appendix 3 and Figure 3B). Although "prevention" aspects were included in just 2593 (17.4%) of the 14,900 posts, the most frequent preventions identified included "vaccines" (n=1115, 7.5%), "antiviral treatment" (n=849, 5.7%), "mRNA vaccine" (n=453, 3%), and "Letermovir" (n=332, 2.2%), as shown in Figure 3C and in Table S8 in Multimedia Appendix 3. "Hygiene measures" and the specific antiviral treatment "Valganciclovir" were less common overall (n=114, 0.8%, and n=214, 1.4%, respectively), though these were the most frequent "prevention" aspects within #CMVawareness posts. In the final category, which included terms reflecting "general CMV information," "immune response" was the most frequently identified aspect (n=533, 3.6%), which was two times greater than "side effects," the second-most frequently identified aspect in this category (Table S9 in Multimedia Appendix 3). In general, aggregate results combining data for all five hashtags were consistent with results for the individual hashtags #CMV and #cytomegalovirus; however, as described here, distinct trends regarding thematic content were observed when quantified by individual hashtags.

XSL•FO
RenderX

Figure 3. Relative comparison of the number of posts by aspect for each hashtag or aggregated. (A) Population discussed in a post, (B) awareness aspects other than CMV or cCMV, and (C) CMV prevention aspects. The largest circle in a row represents the highest number of posts with that aspect for that hashtag. All other circles are proportional to this reference. cCMV: congenital cytomegalovirus; CMV: cytomegalovirus; mRNA: messenger RNA.



The Impact of the National CMV Awareness Month on Thematic Content of Posts

We next assessed the effect of the National CMV Awareness Month (June 2023) on the thematic content of Twitter/X posts. In our sample of 12,910 (86.6%) posts from May to July 2023, a significant association was observed between the month of publication and the target audience (χ^2_2=144.3, *P*<.001), with an above-expected increase in the general population in June and a decrease in scientists/health care professionals (adjusted Pearson residuals; see Table S10 in Multimedia Appendix 3). This shift from scientists/health care professionals to the general population from May to June was lost in July (Tables S11-S13 in Multimedia Appendix 3). The population discussed also shifted during the National CMV Awareness Month, with a 110% increase in the "pediatrics" group (n=577, 4.5%, posts in May to n=1213, 9.4%, in June), a 134% increase in "women" (n=280, 2.2%, posts in May to n=656, 5.1%, in June), and a decrease of –32% in "transplant recipients" (n=569, 4.4%, posts in May to n=386, 3%, in June), as shown in Tables S11-S13 in Multimedia Appendix 3. Given the goal of the National CMV Awareness Month to bring awareness about CMV and prevent infection, we next tested for an association between the month of publication and the number of posts that contained these attributes. A significant association was observed between "awareness" and "prevention" and the publication month

XSL•FO
RenderX

(χ^2_2=107.8, *P*<.001), with a significant increase in posts containing awareness messaging compared to that expected and a significant decrease in prevention messaging in June (adjusted Pearson residuals; see Table S10 in Multimedia Appendix 3). To understand which awareness and prevention aspects shifted during the National CMV Awareness Month, we analyzed individual aspects. Within the "awareness and knowledge" thematic category, the absolute number of posts increased across all aspect groups from May to June 2023 (Tables S11-S13 in Multimedia Appendix 3). We then assessed whether the proportional representation of individual aspects also shifted during the National CMV Awareness Month. The proportion of "awareness and knowledge" posts mentioning "burden of disease" increased from 5.7% (59/3773) of posts in May to 9.8% (173/6421) in June (+71%), and posts mentioning "horizontal transmission" increased from 3.8% (39/3773) of posts in May to 10.4% (183/6421) in June (+174%). In contrast, mentions of "universal screening" decreased from 5.6% (58/3773) of posts in May to 3.6% (64/6421) in June (–35%), as shown in Tables S11-S13 in Multimedia Appendix 3. Shifts in the proportional representation of aspects within the "prevention" category also occurred from May to June 2023. The proportion of "prevention" posts with mentions of "vaccine" increased from 24.5% (207/845) of posts in May to 42.9% (368/857) in June (+75%). A proportional decrease in posts that mentioned "antivirals" (–26%; 268/845, 31.7%, posts in May to 202/857, 23.6%, posts in June) and "hygiene measures" (–37%; 50/845, 5.9%, posts in May to 32/857, 3.7%, posts in June) also occurred (Tables S11-S13 in Multimedia Appendix 3). Finally, as expected, the "fundraising" aspect increased within the "general" category of posts during the National CMV Awareness Month (+157%; 37/286, 4.4%, posts in May to 95/329, 11.2%, posts in June), as shown in Tables S11-S13 in Multimedia Appendix 3.

Sentiment Analyses

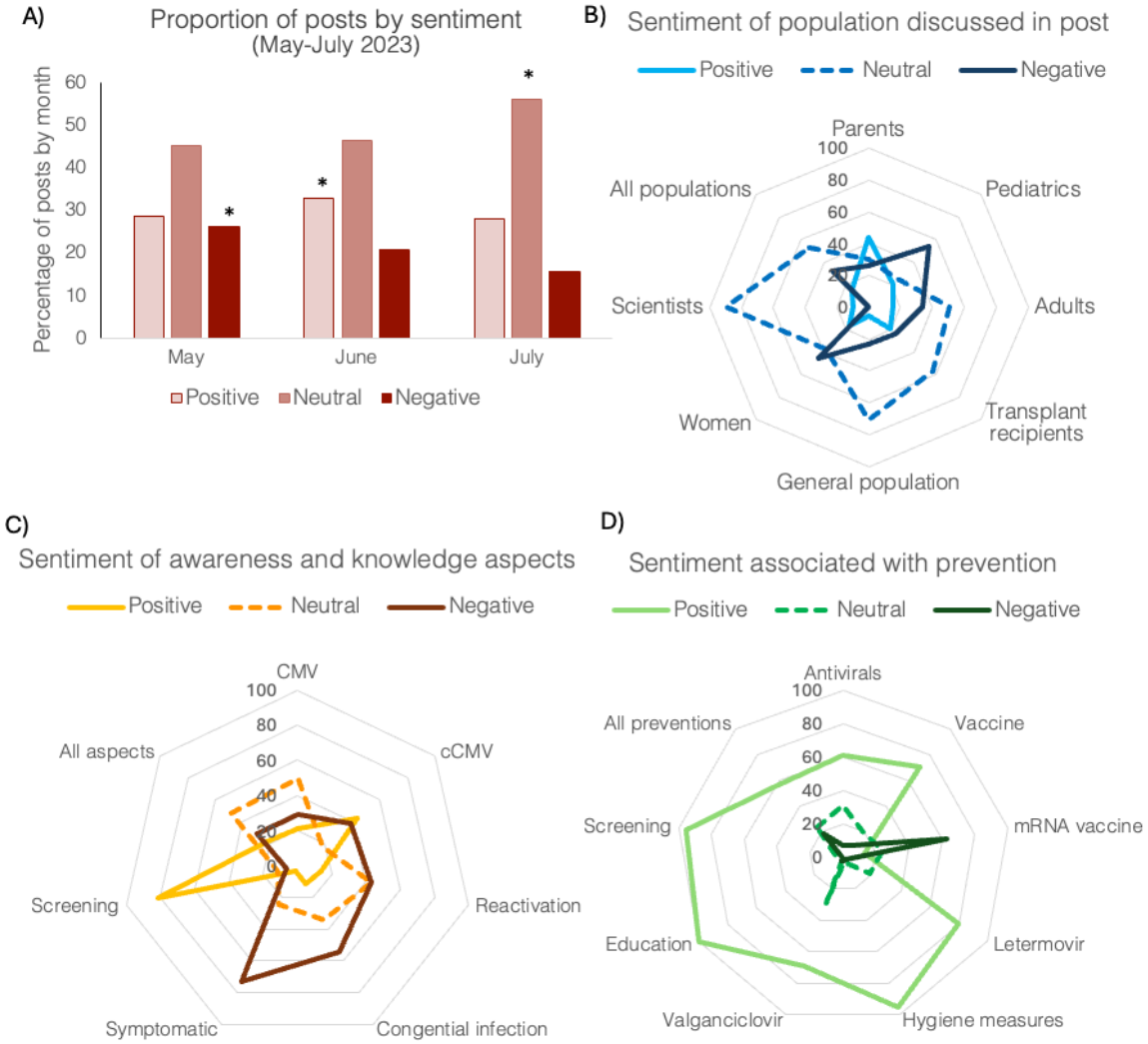
Through sentiment analysis, the majority of social media posts were classified as neutral. This was true for most of the broad thematic categories of prespecified aspects (Table S14 in Multimedia Appendix 3 and Figure 4A), though the trend reversed for aspects pertaining to "prevention" for which the number of posts classified as positive (n=1288, 8.6%) was several times greater than the number of posts classified as negative (n=418, 2.8%), as shown in Table S17 in Multimedia Appendix 3. Overall, more social media posts were classified as positive relative to negative (n=3531, 23.7%, vs n=2436, 16.3%, respectively; see Table S14 in Multimedia Appendix 3). In our sample of 12,910 (86.6%) posts from May to July 2023, a significant association was observed between sentiment type and the month of publication (χ^2_4=163.6, *P*<.001). An examination of all possible combinations of posts revealed posts during the National CMV Awareness Month to be significantly more likely to be positive than expected (adjusted Pearson residuals; see Table S14 in Multimedia Appendix 3 and Figure 4A).

The sentiment associated with independent aspects was also scored. A significant association was observed between sentiment type and the specific target audience, either the general population or scientists/health care professionals (χ^2_2= 481.2, *P*<.001). Although most posts were classified as neutral (Table S15 in Multimedia Appendix 3), "general population" posts were significantly more likely to be positive or negative in sentiment than expected, while "scientists/health care professionals" posts were significantly more likely to be neutral (adjusted Pearson residuals; see Table S15 in Multimedia Appendix 3).

The sentiment varied depending on the specific "population" discussed," the specific "awareness and knowledge" aspect, or the type of "prevention." For example, among the populations discussed, the sentiment of the "scientists" aspect was disproportionally neutral (620/691, 89.7%); in contrast, the sentiment of the "parents" aspect was more likely classified as positive (99/225, 44%), and the most common sentiment for the "pediatrics" aspect was negative (1336/2476, 54%), as shown in Table S16 in Multimedia Appendix 3 and Figure 4B. Examples of aspects included in the "awareness and knowledge" category that diverged from the average sentiment included increased positive association with "screening" (266/327, 81.3%) and increased negative association with "symptomatic" (712/977, 72.9%) and "congenital infection" (437/800, 54.6%), as shown in Table S17 in Multimedia Appendix 3 and Figure 4C. Lastly, "education," "screening," and "hygiene measures" aspects included in the "prevention" category were scored as nearly universally positive (education: 23/23, 100%;screening: 22/23, 95.7%; and hygiene measures: 109/114, 95.6%), as shown in Table S18 in Multimedia Appendix 3 and Figure 4D.

XSL•FO
RenderX

Figure 4. Evaluation of post sentiment from May to July 2023. (a) Proportion of posts scored by overall sentiment per month, (b) by population discussed, (c) by awareness and knowledge, and (d) by type of prevention. Post attributes that occurred infrequently (<10% of the total aspects) were not included. Posts that included more than one attribute are represented in each individual aspect that the posts included. *Significant increase over expected (*P*<.05), chi-square, adjusted Pearson residuals. cCMV: congenital cytomegalovirus; CMV: cytomegalovirus; mRNA: messenger RNA.

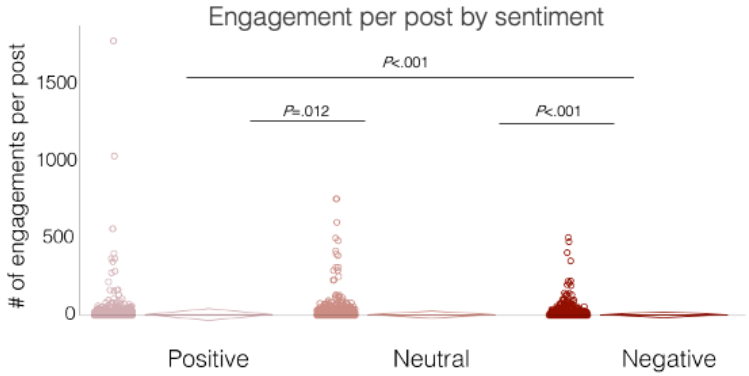


User Engagement

In our final analysis, we assessed user engagement with CMV-relevant posts, including "likes," retweets, and comments. Overall, users engaged with 3863 (27.3%) of 14,136 posts from May to August 2023. Post engagement was slightly higher on average if scored with a positive sentiment (1230/3895, 31.6%) compared to a neutral (1923/6445, 29.8%) or a negative (710/3796, 18.7%) sentiment (Table S19 in Multimedia Appendix 3). To understand potential differences in the magnitude of engagement with respect to sentiment, we compared median engagement per post across sentiment groupings. Engagement was found to be significantly different between sentiments (Kruskal-Wallis; χ^2_2=194.31, *P*<.001), with a higher rank mean for positive posts compared to neutral and negative posts (Table S19 in Multimedia Appendix 3 and Figure 5). Therefore, posts with a positive sentiment were more likely to engage the audience and to a greater degree, while those with negative sentiment were least likely to do so.

XSL•FO
RenderX

Figure 5. Evaluation of post engagement with respect to sentiment (May-August 2023). A dot plot of the number of engagements per post categorized by sentiment, with the SD displayed to the right. **P* values are adjusted with Bonferroni correction, Kruskal-Wallis, Dunn-Bonferroni test. One outlier value (3835, negative) was removed from the graph to improve visualization of data.



Discussion

Principal Findings

The objective of our study was to monitor the volume and thematic content of social media posts on Twitter/X before, during, and after the National CMV Awareness Month in June 2023 to understand the virtual impact of the campaign. We first used a language detection model and a customized BERT tokenizer to extract English language tweets and to remove posts that were not relevant to CMV or CMV disease. Analysis of the remaining 14,900 CMV-relevant posts revealed a peak in posts during the National CMV Awareness Month, a trend observed across all five of the most frequently used CMV-related hashtags, with the highest volume of posts originating from the United States. As expected, these data confirm an active campaign initiative by multiple CMV stakeholders in the United States. We sought to further characterize who is generating information and how effectively their messages are disseminated by identifying the key users and classifying these authors by affiliation. These data point to a potential opportunity to enhance collaboration between advocacy organizations, academic researchers (who were observed to be the most prolific authors), and media outlets (observed to have the largest number of followers) to expand messaging and to target messaging to specific audiences.

Our analyses also examined the thematic content, or aspects, of social media posts, along with the sentiment of each post. A subset of posts (outside of the primary dataset) was annotated by researchers for these attributes, and this dataset was provided to a ChatGPT model, which annotated CMV-relevant posts in substantial agreement with blinded reviewers. Highly mentioned aspects were typically shared between hashtags, though hashtag-specific trends point to specific conversations being siloed to separate channels. For example, “screening” was mentioned primarily under the #stopCMV hashtag, while “transplant recipients” were discussed under the #CMV and #cytomegalovirus hashtags. The National CMV Awareness Month shifted CMV conversations toward the general audience from scientists and health care professionals and were more likely to contain awareness messaging indicating an effort by stakeholders to increase attention with respect to pediatric populations, women, and the burden of disease, most likely reflecting advocacy around cCMV. The sentiment of CMV-relevant social media posts was overall neutral, though it shifted meaningfully toward positive during the campaign month. The overwhelmingly positive sentiment associated with CMV the “prevention” aspects “education,” “screening,” and “hygiene measures” speaks to the enthusiastic advocacy of the CMV community for various interventions and preventive measures. Unexpectedly, prevention messaging decreased significantly during the awareness month even as posts containing prevention aspects were observed to have a positive sentiment, which correlates with higher community engagement.

The methodology used in this study was implemented with multiple checks to ensure accuracy in processing and analysis and is described in sufficient detail to support reproducibility. This allows our methodology to be easily used to evaluate future CMV awareness campaigns to identify long-term trends or shifts in CMV thematic content and sentiment. We envision the possibility of monitoring social media posts as new interventions become available, as is done commonly for sentiment analysis around vaccines against influenza and SARS-CoV-2. Our methods are adaptable and can be expanded to monitor messaging once a vaccine against CMV becomes available or as new legislation is proposed around newborn screening for cCMV. These surveillance and data collection steps provide a foundation for more rigorous interpretation through the lens of health communication and health behavior theory to inform future advocacy and awareness campaigns.

Limitations

This study has several limitations. We examined posts from a single social media platform (ie, Twitter/X) using Keyhole. Users of Twitter/X are not necessarily representative of the broader US population or all CMV stakeholders; moreover, user demographics likely differ on other platforms, such as Instagram, Reddit, Facebook, and TikTok. Furthermore, although Keyhole is widely used in commercial practice, it is not commonly used in public health studies, in part due to cost. The platform aligns well with real-world social listening practice (eg, real-time tracking), making it suitable for evaluating a public health awareness campaign; however, data were collected, filtered, and summarized by Keyhole rather than accessed as raw posts directly from the Twitter/X API, which may affect completeness and reproducibility.

XSL•FO
RenderX

We were also unable to analyze nearly 9000 non-English tweets, which may differ in terms of author category, aspects, and sentiment relative to the English-language posts analyzed here. Furthermore, the manual classification of aspects and sentiment by researchers is a subjective process, and the list of aspects generated by the authors, while detailed and broad, may not reflect all possible themes discussed in posts. Although moderate in agreement, the interrater reliability in sentiment scores between the four independent raters reflects a limitation in the sentiment analysis. Complete disagreement between the reviewers occurred in 4 of 50 test tweets (positive or negative, excluding neutral), which may reflect the inherently mixed nature of disease-related posts that combine positive messaging around interventions or milestones and negative messaging that describes the underlying need for the intervention. Accurate annotation and classification by AI were dependent on these subjective processes. They may have also been impacted by the constrained length of tweets, which can limit content, explanatory details, and other cues. Lastly, although the pre- and postperiods surrounding June 2023 we examined are limited, we were able to evaluate the immediate impact of the National CMV Awareness Month. For these reasons, the results reported here should be viewed as exploratory and interpreted with this lens. Additional research leveraging the expertise of diverse stakeholders is needed to design and evaluate long-term public health information and future CMV awareness campaigns.

Strengths

This study also has several notable strengths. To the best of our knowledge, it represents the first systematic infodemiologic evaluation of the National CMV Awareness Month and the first large-scale characterization of conversations concerning CMV on Twitter/X. By analyzing nearly 15,000 CMV-relevant posts, this study leveraged a larger and more comprehensive dataset than would have been feasible to analyze through manually annotation alone. Furthermore, analysis from multiple lenses, such as user characteristics, thematic content, and sentiment, provides a multidimensional view of CMV discussions that has not been unavailable to researchers, advocates, or other relevant stakeholders.

A second strength is the practical demonstration of a multistep analytical pipeline that combines human annotation, model customization, and iterative validation. Furthermore, the application of few-shot prompting with ChatGPT enabled efficient classification of aspects and sentiment at scale. Importantly, human-AI agreement of post sentiment was substantial, increasing confidence in the reliability of the automated post annotations. Finally, the study's design, which evaluated posts before, during, and after the campaign month, allows for a direct observation of how a nationally recognized campaign shifts online attention, thematic content, and sentiment.

Conclusion

To the best of our knowledge, this study is the first to examine and report the volume, thematic content, and sentiment of virtual CMV-related conversations on Twitter/X before, during, and after the National CMV Awareness Month. The use of AI permitted detailed evaluation of many thousands of social media posts. The results of our analyses enable us to predict potential collaborations between key users to achieve greater dissemination and impact during future campaigns. In addition, the detailed analyses presented here provide a more complete characterization of the conversations and culture within distinct CMV-related hashtags and highlight the thematic content that can be amplified in future campaigns. The complexity of health communication via social media poses distinct challenges to public health investigators and practitioners when planning and executing information and awareness campaigns. Although this study demonstrates the analytic capabilities of AI, the generative capabilities of the ChatGPT model could also be used to draft campaign messaging to enhance specific themes or emotional undertones.

Acknowledgments

We are grateful to Giovana Vieira and Rofail Wassef for their kind contribution of scoring the sentiment for social media posts. This research was supported by startup funds provided to TR.

Data Availability

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

Authors' Contributions

Conceptualization was handled by KF and TRR; data curation, ZY and TRR; formal analysis, ZY, TRR, LD, and TP; funding acquisition, TRR; investigation, TRR, ZY, LD, TP, and KF; methodology, ZY, TP, TRR, JDD, and RVW; project administration, CK, TRR, JDD, and RVW; resources, TRR; supervision, CK, KF, TRR, and JDD; validation, TRR and ZY; visualization, TRR and ZY; writing—original draft, TRR, CK, ZY, and LD; and writing—review and editing, TRR, CK, ZY, LD, RVW, KF, and JDD. No advocacy organization provided funding for this study. The views expressed are those of the authors and do not necessarily represent those of their institutions.

Conflicts of Interest

TRR is an employee of Stonehill College, and LD was an undergraduate student. TRR received institutional startup funds from Stonehill College that supported this work. ZY, TP, and RVW were employees of Moderna Therapeutics, Incorporated, at the

XSL•FO
RenderX

time this study was conducted and may have held company stock or stock options during that period. JDD and CK are employees of Moderna Therapeutics, Incorporated, and may hold company stock or stock options. KF receives consulting fees from Moderna Therapeutics and KF reports no competing interests. The authors have no financial relationships with Twitter/X, Keyhole, OpenAI, or the National Cytomegalovirus Foundation relevant to this work.

Multimedia Appendix 1
Expanded methodology.
[DOCX File , 20 KB - infodemiology_v6i1e80922_app1.docx]

Multimedia Appendix 2
ChatGPT-4 model master prompt used to annotate aspects and sentiments and identify tweet segments associated with those aspects and sentiments.
[DOCX File , 17 KB - infodemiology_v6i1e80922_app2.docx]

Multimedia Appendix 3
Thematic categories, social media post counts, top 20 social media authors, adjusted Pearson residuals and chi-square test results, and engagement with respect to sentiment.
[DOCX File , 61 KB - infodemiology_v6i1e80922_app3.docx]

References

1. Dana Flanders W, Lally C, Dilley A, Diaz-Decaro J. Estimated cytomegalovirus seroprevalence in the general population of the United States and Canada. J Med Virol 2024 Mar 26;96(3):e29525 [FREE Full text] [doi: 10.1002/jmv.29525] [Medline: 38529529]
2. Lantos PM, Permar SR, Hoffman K, Swamy GK. The excess burden of cytomegalovirus in African American communities: a geospatial analysis. Open Forum Infect Dis 2015 Dec;2(4):ofv180 [FREE Full text] [doi: 10.1093/ofid/ofv180] [Medline: 26716106]
3. Zuhair M, Smit GSA, Wallis G, Jabbar F, Smith C, Devleesschauwer B, et al. Estimation of the worldwide seroprevalence of cytomegalovirus: a systematic review and meta-analysis. Rev Med Virol 2019 May 31;29(3):e2034 [FREE Full text] [doi: 10.1002/rmv.2034] [Medline: 30706584]
4. Lantos PM, Hoffman K, Permar SR, Jackson P, Hughes BL, Kind A, et al. Neighborhood disadvantage is associated with high cytomegalovirus seroprevalence in pregnancy. J Racial Ethn Health Disparities 2018 Aug;5(4):782-786 [FREE Full text] [doi: 10.1007/s40615-017-0423-4] [Medline: 28840519]
5. Clinical overview of CMV and congenital CMV: cytomegalovirus (CMV) and congenital CMV infection. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/cytomegalovirus/hcp/clinical-overview/index.html [accessed 2024-12-16]
6. Stagno S, Pass RF, Dworsky ME, Alford CA. Maternal cytomegalovirus infection and perinatal transmission. Clin Obstet Gynecol 1982 Sep;25(3):563-576 [FREE Full text] [doi: 10.1097/00003081-198209000-00014] [Medline: 6290121]
7. Wang H, Peng G, Bai J, He B, Huang K, Hu X, et al. Cytomegalovirus infection and relative risk of cardiovascular disease (ischemic heart disease, stroke, and cardiovascular death): a meta‐analysis of prospective studies up to 2016. J Am Heart Assoc 2017 Jul 06;6(7):e005025 [FREE Full text] [doi: 10.1161/JAHA.116.005025] [Medline: 28684641]
8. Aiello AE, Haan M, Blythe L, Moore K, Gonzalez JM, Jagust W. The influence of latent viral infection on rate of cognitive decline over 4 years. J Am Geriatr Soc 2006 Jul;54(7):1046-1054 [FREE Full text] [doi: 10.1111/j.1532-5415.2006.00796.x] [Medline: 16866674]
9. Burgdorf KS, Trabjerg BB, Pedersen MG, Nissen J, Banasik K, Pedersen OB, et al. Large-scale study of toxoplasma and cytomegalovirus shows an association between infection and serious psychiatric disorders. Brain Behav Immun 2019 Jul;79:152-158 [FREE Full text] [doi: 10.1016/j.bbi.2019.01.026] [Medline: 30685531]
10. Barnes LL, Capuano AW, Aiello AE, Turner AD, Yolken RH, Torrey EF, et al. Cytomegalovirus infection and risk of Alzheimer disease in older black and white individuals. J Infect Dis 2015 Jan 15;211(2):230-237 [FREE Full text] [doi: 10.1093/infdis/jiu437] [Medline: 25108028]
11. Phillips AC, Carroll D, Khan N, Moss P. Cytomegalovirus is associated with depression and anxiety in older adults. Brain Behav Immun 2008 Jan;22(1):52-55 [FREE Full text] [doi: 10.1016/j.bbi.2007.06.012] [Medline: 17703915]
12. Gadoth A, Ourfalian K, Basnet S, Kunzweiler C, Bohn RL, Fülöp T, et al. Potential relationship between cytomegalovirus and immunosenescence: evidence from observational studies. Rev Med Virol 2024 Jul;34(4):e2560 [FREE Full text] [doi: 10.1002/rmv.2560] [Medline: 38866595]
13. Orlikowski D, Porcher R, Sivadon-Tardy V, Quincampoix J, Raphaël JC, Durand M, et al. Guillain-Barré syndrome following primary cytomegalovirus infection: a prospective cohort study. Clin Infect Dis 2011 Apr 01;52(7):837-844 [FREE Full text] [doi: 10.1093/cid/cir074] [Medline: 21427390]

XSL•FO
RenderX

14. Mercado NB, Real JN, Kaiserman J, Panagioti E, Cook CH, Lawler SE. Clinical implications of cytomegalovirus in glioblastoma progression and therapy. NPJ Precis Oncol 2024 Sep 29;8(1):213 [FREE Full text] [doi: 10.1038/s41698-024-00709-4] [Medline: 39343770]
15. Savva GM, Pachnio A, Kaul B, Morgan K, Huppert FA, Brayne C, Medical Research Council Cognitive Function and Ageing Study. Cytomegalovirus infection is associated with increased mortality in the older population. Aging Cell 2013 Jun;12(3):381-387 [FREE Full text] [doi: 10.1111/acel.12059] [Medline: 23442093]
16. Leeaphorn N, Garg N, Thamcharoen N, Khankin EV, Cardarelli F, Pavlakis M. Cytomegalovirus mismatch still negatively affects patient and graft survival in the era of routine prophylactic and preemptive therapy: a paired kidney analysis. Am J Transplant 2019 Feb;19(2):573-584 [FREE Full text] [doi: 10.1111/ajt.15183] [Medline: 30431703]
17. Stern L, Withers B, Avdic S, Gottlieb D, Abendroth A, Blyth E, et al. Human cytomegalovirus latency and reactivation in allogeneic hematopoietic stem cell transplant recipients. Front Microbiol 2019 May 28;10:1186 [FREE Full text] [doi: 10.3389/fmicb.2019.01186] [Medline: 31191499]
18. CMV fact sheet for healthcare providers. Centers for Disease Control and Prevention. URL: https://stacks.cdc.gov/view/cdc/137322 [accessed 2025-12-23]
19. Boppana SB, van Boven M, Britt WJ, Gantt S, Griffiths PD, Grosse SD, et al. Vaccine value profile for cytomegalovirus. Vaccine 2023 Nov 03;41 Suppl 2:S53-S75 [FREE Full text] [doi: 10.1016/j.vaccine.2023.06.020] [Medline: 37806805]
20. Lombardi G, Garofoli F, Stronati M. Congenital cytomegalovirus infection: treatment, sequelae and follow-up. J Matern Fetal Neonatal Med 2010 Oct;23 Suppl 3:45-48 [FREE Full text] [doi: 10.3109/14767058.2010.506753] [Medline: 20807160]
21. Song X, Li Q, Diao J, Li J, Li Y, Zhang S, et al. Association between first-trimester maternal cytomegalovirus infection and stillbirth: a prospective cohort study. Front Pediatr 2022 Jun;10(11):803568-803604 [FREE Full text] [doi: 10.3389/fped.2022.803568] [Medline: 35372174]
22. Pesch M, Mowers J, Huynh A, Schleiss M. Intrauterine fetal demise, spontaneous abortion and congenital cytomegalovirus: a systematic review of the incidence and histopathologic features. Viruses 2024 Sep 30;16(10):1552 [FREE Full text] [doi: 10.3390/v16101552] [Medline: 39459885]
23. Pereira L, Petitt M, Fong A, Tsuge M, Tabata T, Fang-Hoover J, et al. Intrauterine growth restriction caused by underlying congenital cytomegalovirus infection. J Infect Dis 2014 May 15;209(10):1573-1584 [FREE Full text] [doi: 10.1093/infdis/jiu019] [Medline: 24403553]
24. Manicklal S, Emery VC, Lazzarotto T, Boppana SB, Gupta RK. The “silent” global burden of congenital cytomegalovirus. Clin Microbiol Rev 2013 Jan;26(1):86-102 [FREE Full text] [doi: 10.1128/cmr.00062-12]
25. Diaz-Decaro J, Myers E, Mucha J, Neumann M, Lewandowski W, Kaczanowska M, et al. A systematic literature review on the humanistic burden of cytomegalovirus. Curr Med Res Opin 2023 May;39(5):739-750 [FREE Full text] [doi: 10.1080/03007995.2023.2191477] [Medline: 36938652]
26. Diaz-Decaro J, Myers E, Mucha J, Neumann M, Lewandowski W, Kaczanowska M, et al. A systematic literature review of the economic and healthcare resource burden of cytomegalovirus. Curr Med Res Opin 2023 Jul;39(7):973-986 [FREE Full text] [doi: 10.1080/03007995.2023.2222583] [Medline: 37395088]
27. Cannon MJ, Westbrook K, Levis D, Schleiss MR, Thackeray R, Pass RF. Awareness of and behaviors related to child-to-mother transmission of cytomegalovirus. Prev Med 2012 May;54(5):351-357 [FREE Full text] [doi: 10.1016/j.ypmed.2012.03.009] [Medline: 22465669]
28. Congressional record - Senate S4106. United States Senate. 2011. URL: https://www.congress.gov/112/crec/2011/06/23/CREC-2011-06-23-pt1-PgS4106-3.pdf [accessed 2025-12-23]
29. eHealth: health topics. World Health Organization Regional Office for the Eastern Mediterranean. URL: https://www.emro.who.int/health-topics/ehealth/ [accessed 2024-12-16]
30. Social media in America: key stats in 2025. Sprinklr. 2025. URL: https://www.sprinklr.com/blog/social-media-in-america/ [accessed 2025-11-16]
31. Schumacher S, Sparks G, Montalvo III J, Kirzinger A, Hamel L. KFF health information and trust tracking poll: health information and advice on social media. KFF. 2025 Aug 7. URL: https://www.kff.org/public-opinion/kff-health-information-and-trust-tracking-poll-health-information-and-advice-on-social-media/ [accessed 2025-11-16]
32. Gottfried J. Americans’ social media use. Pew Research Center. 2024 Jan 31. URL: https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/ [accessed 2025-11-16]
33. June is national CMV awareness month - week 1 “CMV is common”. National CMV Foundation. 2016 Jun. URL: https://www.nationalcmv.org/cmv-research/blog/june-2016/national-cmv-awareness-month [accessed 2025-11-16]
34. About cytomegalovirus: cytomegalovirus (CMV) and congenital CMV infection. Centers for Disease Control and Prevention. 2025 Jan 17. URL: https://www.cdc.gov/cytomegalovirus/about/index.html [accessed 2025-11-16]
35. Singh M, Jakhar AK, Pandey S. Sentiment analysis on the impact of coronavirus in social life using the BERT model. Soc Netw Anal Min 2021;11(1):33 [FREE Full text] [doi: 10.1007/s13278-021-00737-z] [Medline: 33758630]
36. Mirugwe A, Ashaba C, Namale A, Akello E, Bichetero E, Kansiime E, et al. Sentiment analysis of social media data on Ebola outbreak using deep learning classifiers. Life (Basel) 2024 May 30;14(6):708 [FREE Full text] [doi: 10.3390/life14060708] [Medline: 38929691]

XSL•FO
RenderX

37. Ahmad M, Batyrshin I, Sidorov G. Sentiment analysis using a large language model–based approach to detect opioids mixed with other substances via social media: method development and validation. JMIR Infodemiology 2025 Jun 19;5(1):e70525 [FREE Full text] [doi: 10.2196/70525] [Medline: 40536906]
38. Deiner MS, Deiner NA, Hristidis V, McLeod SD, Doan T, Lietman TM, et al. Use of large language models to assess the likelihood of epidemics from the content of tweets: infodemiology study. J Med Internet Res 2024 Mar 01;26:e49139 [FREE Full text] [doi: 10.2196/49139] [Medline: 38427404]
39. Tastad KJ, Schleiss MR, Lammert SM, Basta NE. Awareness of congenital cytomegalovirus and acceptance of maternal and newborn screening. PLoS One 2019 Aug 26;14(8):e0221725 [FREE Full text] [doi: 10.1371/journal.pone.0221725] [Medline: 31449545]
40. Binda S, Pellegrinelli L, Terraneo M, Caserini A, Primache V, Bubba L, et al. What people know about congenital CMV: an analysis of a large heterogeneous population through a web-based survey. BMC Infect Dis 2016 Sep 26;16(1):513 [FREE Full text] [doi: 10.1186/s12879-016-1861-z] [Medline: 27671033]
41. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online 2019 [FREE Full text] [doi: 10.48550/arXiv.1810.04805]
42. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174 [FREE Full text] [Medline: 843571]
43. Code of Federal Regulations: § 46.104 Exempt research. U.S. Department of Health and Human Services. 2026 Jan 16. URL: https://www.ecfr.gov/current/title-45/part-46/section-46.104#p-46.104(d)(4)(i)

Abbreviations

AI: artificial intelligence
API: application programming interface
BERT: Bidirectional Encoder Representations from Transformers
cCMV: congenital cytomegalovirus.
CMV: cytomegalovirus
LLM: large language model
NCMVF: National Cytomegalovirus Foundation

Edited by T Mackey; submitted 19.Jul.2025; peer-reviewed by M Marian, A Rasool; comments to author 30.Sep.2025; revised version received 10.Dec.2025; accepted 17.Dec.2025; published 22.Jan.2026.

Please cite as:
Rosebrock TR, Yang Z, D'Arco L, Pathak T, Vislay-Wade R, Fowler K, Diaz-Decaro J, Kunzweiler C
Using Artificial Intelligence Methods to Evaluate the Effect of the National Cytomegalovirus Awareness Month on the Content and Sentiment of Social Media Posts: Infodemiology Study
JMIR Infodemiology 2026;6:e80922
URL: https://infodemiology.jmir.org/2026/1/e80922
doi:10.2196/80922
PMID:

©Tracy R Rosebrock, Zhen Yang, Lauren D'Arco, Tapan Pathak, Rebecca Vislay-Wade, Karen Fowler, John Diaz-Decaro, Colin Kunzweiler. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 22.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Automated Risk Assessment of Opioid Use: Analysis Using Pre-Trained Transformers on Social Media Data

Muhammad Ahmad[1], Mphil; Rita Orji[2], Prof Dr; Maaz Amjad[3], PhD; Abubakar Siddique[4], PhD; Nailya Kubysheva[5], PhD; Ildar Batyrshin[1], Prof Dr; Grigori Sidorov[1], Prof Dr

[1]Instituto Politécnico Nacional, Centro de Investigación en Computación, Mexico City, Mexico

[2]Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada

[3]Department of Computer Science, Texas Tech University, Lubbock, TX, United States

[4]School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand

[5]Kazan Federal University, Kazan, Russian Federation

Corresponding Author:
Grigori Sidorov, Prof Dr
Instituto Politécnico Nacional
Centro de Investigación en Computación
Juan de Dios Bátiz S/N, Nueva Industrial Vallejo, Gustavo A. Madero
Mexico City, 07738
Mexico
Phone: 52 5591887293
Email: sidorov@cic.ipn.mx

Abstract

Background: The illegal use of opioids has emerged as a major global public health concern, contributing to widespread addiction and a growing number of overdose-related deaths. In response, the US federal government has invested billions of dollars in combating the opioid epidemic through treatment, prevention, and law enforcement initiatives. Despite these efforts, there remains an urgent need for automated tools capable of detecting overdose cases and assessing the risk levels of substances—tools that can enable faster, more effective responses with less reliance on human intervention. Social media, particularly Reddit, has become a valuable source of self-reported data on opioid misuse, offering rich insights into user experiences and symptoms.

Objective: This research aimed to develop an advanced automated tool for detecting opioid overdose risks and classifying substances into high-risk and low-risk categories by analyzing social media posts.

Methods: A multistage methodology was used to achieve the objectives of this work. First, a new dataset was constructed from Reddit posts and manually annotated. Each post was labeled according to the risk level of the mentioned substance, using contextual indicators and user-reported experiences as the basis for classification. To ensure reliability and annotator consistency, detailed annotation guidelines were developed and applied throughout the labeling process. Second, a bidirectional encoder representation from transformers for biomedical text mining (BioBERT)–based classification framework was implemented and enhanced with a custom attention mechanism to capture relevant semantic information for more accurate predictions. Third, the model's performance was evaluated using 5-fold cross-validation and compared against several baseline approaches, including traditional supervised learning, deep learning, and transfer learning methods. In total, 14 experiments were conducted to evaluate comparative effectiveness. To further assess the contribution of the attention layer, the best-performing model was also evaluated against a version incorporating the standard self-attention mechanism, using a train-test split. Finally, a paired *t* test was conducted to statistically assess the performance difference between the BioBERT-based model and the strongest baseline, extreme gradient boosting (XGBoost), providing validation of the observed improvements.

Results: The proposed BioBERT model with custom attention achieved an F_1-score of 0.99 in cross-validation, outperforming the best baseline, XGBoost (F_1-score=0.97), with a relative improvement of 2.06%. A paired *t* test conducted across the 5 folds (n=5) confirmed that the performance gain was statistically significant (*P*=.003), providing strong evidence that the improvement reflects genuine advances in overdose risk detection.

Conclusions: This paper demonstrates the potential of leveraging social media data and advanced natural language processing models to build reliable systems for opioid overdose risk detection. The BioBERT model with custom attention shows state-of-the-art

XSL•FO
RenderX

performance and robustness, offering a powerful tool to support timely intervention and harm reduction strategies in the ongoing opioid crisis.

(JMIR Infodemiology 2026;6:e77783) doi:10.2196/77783

KEYWORDS

opioid overdose; chronic pain; data mining; social media; deep learning; transformer; BERT; drug abuse; Reddit; AI; artificial intelligence

Introduction

Background

Chronic pain has become a major public health challenge worldwide, affecting more than 25 million adults in the United States [1,2], with prevalence rates ranging from 11% to 40% [3]. Treatment of chronic pain is a complex process, and most treatment advice proposes a combination of psychological and pharmacological methods [4-6], and pharmacological treatments include both prescribed and illicit medications [3].

Nonprescribed medications, such as paracetamol, codeine, tramadol, ibuprofen, and aspirin, are frequently used for mild pain management when medical consultation is unnecessary, and they are usually considered harmless and effective [7]. Paracetamol is commonly used to relieve mild pain and is considered safe when administered according to recommended guidelines. However, there has been a surge in reports showing its misuse [7,8]. A study observing nonprescribed drug usage, such as paracetamol, showed that 24% of adults admitted to exceeding the recommended dosage [9]. Another study found that many people do not read the pharmacological instructions, and more than 50% do not know the active ingredient in the medication that they are taking [10]. Codeine and tramadol, classified as weak opioids, can sometimes be obtained as nonprescribed drugs, despite their potential for side effects [11].

Meanwhile, the misuse of prescription opioids and illicit drugs, such as heroin and fentanyl, has resulted in a record number of opioid-related overdoses and fatalities. According to the Centers for Disease Control and Prevention (CDC), opioids were involved in approximately 70% of all drug overdose deaths in the United States [12]. This crisis presents significant health risks [13], strains health care systems, contributes to social unrest, and intensifies economic challenges. Addressing this issue requires multifaceted strategies that go beyond traditional clinical settings, incorporating preventive measures and rapid responses to emerging risks.

Moreover, despite being illegal, the online sale of prescription medications has been documented [11,14]. Twitter (now known as X; developed by X Corp) has emerged as a platform where illegal transactions involving these drugs occur. Sales on Twitter are often less regulated and monitored compared to those on other online platforms [15]. Similarly, Reddit has also been identified as a place where individuals engage in discussions or transactions related to prescription drugs, often with minimal oversight [16,17].

Social media is an important source of user-generated content that offers valuable insights into opioid misuse. Reddit is a popular social media platform and is ranked as the 9th-largest social media platform in the United States [18]. It provides enough space to express self-reported experiences, with a maximum character limit for a text post of 40,000 characters. These discussions offer a unique opportunity to better understand the hidden patterns and risks associated with opioid misuse, which often go overlooked in clinical settings.

In recent years, natural language processing (NLP) has arisen as a powerful tool for addressing and understanding the opioid crisis through social media discourse. Nowadays, social media platforms such as X (formerly Twitter), Facebook (Meta Platforms, Inc), Reddit (Reddit, Inc), YouTube (Google LLC), and Instagram (Meta Platforms, Inc) have become real-time sources of information where people openly discuss their experiences, share their views, and behaviors related to opioid use. These platforms comprise a lot of textual, image, and video data that can be analyzed for sentiment analysis [19-22], named entity recognition [23-25], topic modeling [21,26], text classification [27,28], and opioid-related risks, such as overdose incidents, patterns of misuse, and emerging trends in substance abuse [29,30]. Through advanced NLP methods, it is possible to automatically categorize and evaluate the risk levels of posts related to opioids, providing a timely and scalable method to identify individuals at risk and inform public health interventions.

Unlike traditional clinical settings, social media provides a unique outlet where individuals can candidly share their experiences with opioid use, often motivated by anonymity, community support, and reduced stigma. These self-reported narratives offer critical, real-time insights into opioid misuse patterns that may go unreported in clinical data. While prior studies have used NLP to analyze social media content, most rely on unsupervised methods (eg, topic modeling and sentiment analysis) or weakly labeled data. These approaches often lack the granularity and clinical validity needed for precise risk assessment. Whereas, advanced transformer-based models applied in earlier studies typically do not incorporate domain-specific annotation schemes, limiting their effectiveness in identifying high-risk content.

To address these limitations, this work aims to manually annotate a new dataset using expert-informed risk guidelines, enabling supervised training of models to directly classify opioid-related posts as high-risk or low-risk. Moreover, this study integrates recent advances in NLP with clinically grounded annotation and empirical benchmarking across both traditional and deep learning models, including transformers. This approach supports the development of a scalable system for real-time detection of opioid misuse risk on social media, facilitating timely public health interventions.

XSL•FO
RenderX

To further improve the prediction accuracy of deep learning models, this work introduces a novel architectural enhancement by integrating a custom attention mechanism in bidirectional encoder representations from transformers for biomedical text mining (BioBERT)–based contextualized representations. The attention layer dynamically assigns weights to token embeddings, enabling the model to focus more effectively on the most informative parts of the input text. Such targeted representation learning is expected to enhance the model's ability to capture nuanced patterns associated with opioid misuse. It is hypothesized that this strategy will play a key role in achieving superior empirical performance in detecting high-risk opioid substance use.

This study makes the following contributions: (1) the creation of a manually annotated dataset specifically designed to classify opioid substance use into high-risk and low-risk categories. The corpus is annotated by identifying key indicators and symptoms from social media posts; (2) the development of comprehensive annotation guidelines in collaboration with domain experts to ensure consistent and accurate identification of misuse patterns. These guidelines support early intervention efforts and can inform public health strategies for risk reduction and timely response; and (3) the introduction of a novel architectural enhancement inspired by transformer models, in which a custom attention mechanism is integrated into BioBERT's contextualized representations. This approach is expected to improve the model's ability to capture subtle, high-risk patterns associated with opioid misuse.

Prior Work

This section provides an overview and critical analysis of existing studies on substance use monitoring, opioid-related discourse, and public health surveillance.

Social Media Analysis for Substance Use Monitoring

Several studies have used NLP techniques to monitor drug-related discourse on social media. For instance, Garg et al [17] manually annotated Reddit posts related to fentanyl using a clinically informed codebook comprising 12 risk categories. Subsequently, they trained machine learning models to classify risky content, achieving 76% accuracy and sensitivity. Their work also uncovered slang terms for fentanyl and its analogues, aiding early risk detection. However, their focus remained limited to a single substance and a binary risk classification.

Similarly, Dunn et al [31] conducted a thematic analysis of Reddit discussions on veterinary drug misuse, particularly xylazine and carfentanil. Their qualitative and AI-assisted approach highlighted the potential misuse, motivations, and adverse effects. While valuable for trend detection, their approach lacks automated, supervised classification. In contrast, this work broadens the scope across multiple opioids and introduces a supervised binary classification (high risk vs low risk) using a manually annotated dataset, enhancing the generalizability and automation of risk detection.

Jha et al [32] developed a computational method to identify and characterize the stages of opioid addiction using users' social media posts. They combined recurrent neural networks, information-theoretic word association analysis, and context-based word embeddings to detect addiction stage–specific language patterns. To identify users at high risk of relapse, they applied propensity score matching and logistic regression techniques. Their approach demonstrated high accuracy in distinguishing addiction stages and relapse risk, achieving F_1-scores of 0.88 and 0.79, respectively.

Smith et al [33] investigated the potential of Reddit as a real-time data source for monitoring the US opioid epidemic, focusing on heroin, prescription, and synthetic opioids. They developed an NLP-based pipeline to identify opioid-related content. Moreover, they created a large user cohort of over 1.6 million Reddit users, assigning each to a US state. By tracking opioid-related posts over time, they compared Reddit-based trends with CDC overdose data and National Forensic Laboratory Information System drug reports. Incorporating Reddit data into overdose prediction models significantly improved accuracy, highlighting the value of social media for timely public health surveillance.

Sentiment and Emotion Analysis in Opioid-Related Discourse

Emotion and sentiment play a central role in substance use discussions. Gandy et al [34] evaluated multiple sentiment analysis tools—including VADER, LIWC-22, and ChatGPT 4.0—on YouTube comments about the opioid epidemic. While VADER performed best for identifying negative sentiment, and LIWC-22 for prevalence estimation, their study did not support real-time prediction or classification of risk. Yang et al. [35] used a GAN-based model to predict opioid relapse through sentiment features extracted from Reddit posts. By converting emotional content into "sentiment images," their model captured relapse indicators such as "joy" and "negative" emotions effectively. However, they aimed at predicting relapse, not general risk detection. Similarly, Lokala et al [24] applied deep learning to detect substance use disorder discussions, correlating emotions such as withdrawal and addiction with synthetic opioid mentions. Although their work achieved strong performance (F_1-score=82.12), it emphasized emotion tracking rather than comprehensive risk classification. In contrast, this study integrates an explainable transformer-based model (BioBERT with custom attention) to move beyond emotion detection toward direct, interpretable risk classification of opioid-related content.

Topic Modeling and Public Health Surveillance

Li et al [36] developed a comprehensive pipeline for analyzing COVID-19 drug discourse on Twitter using named entity recognition, sentiment analysis, topic modeling, and drug network mapping. Although their framework processed an extensive dataset of 169 million tweets, it lacked ground-truth annotations and supervised classification, thereby constraining its predictive and validation capabilities.

In another study, Zhang et al [37] examined Reddit posts to understand how teens frame substance use, revealing 7 social-emotional themes such as normalization, coping, and stigmatization. Their focus on emotional framing and qualitative analysis is insightful but not predictive. The novel system proposed here fills these gaps by offering a

XSL•FO
RenderX

ground-truth–labeled, supervised classification system capable of delivering real-time, interpretable predictions. This approach enhances both the predictive validity and practical applicability of social media analytics for public health monitoring and early intervention.

While prior studies have made significant contributions across thematic analysis, sentiment tracking, and topic modeling, they often lack supervised classification or are constrained to a single substance. Building upon these foundations, this work proposes a generalizable, supervised deep learning framework for real-time, risk-level classification of opioid-related Reddit posts, empowered by expert annotations and explainable attention mechanisms.

Methods

Dataset Construction and Preprocessing

To systematically organize high-risk and low-risk opioid substances, a comprehensive list of relevant keywords was compiled. It includes substance names as well as general terms related to opioid use. This strategy helped to ensure broad coverage and improved generalization in data collection. The complete list of keywords is provided in Table 1.

Table 1. Keywords used to extract the dataset.

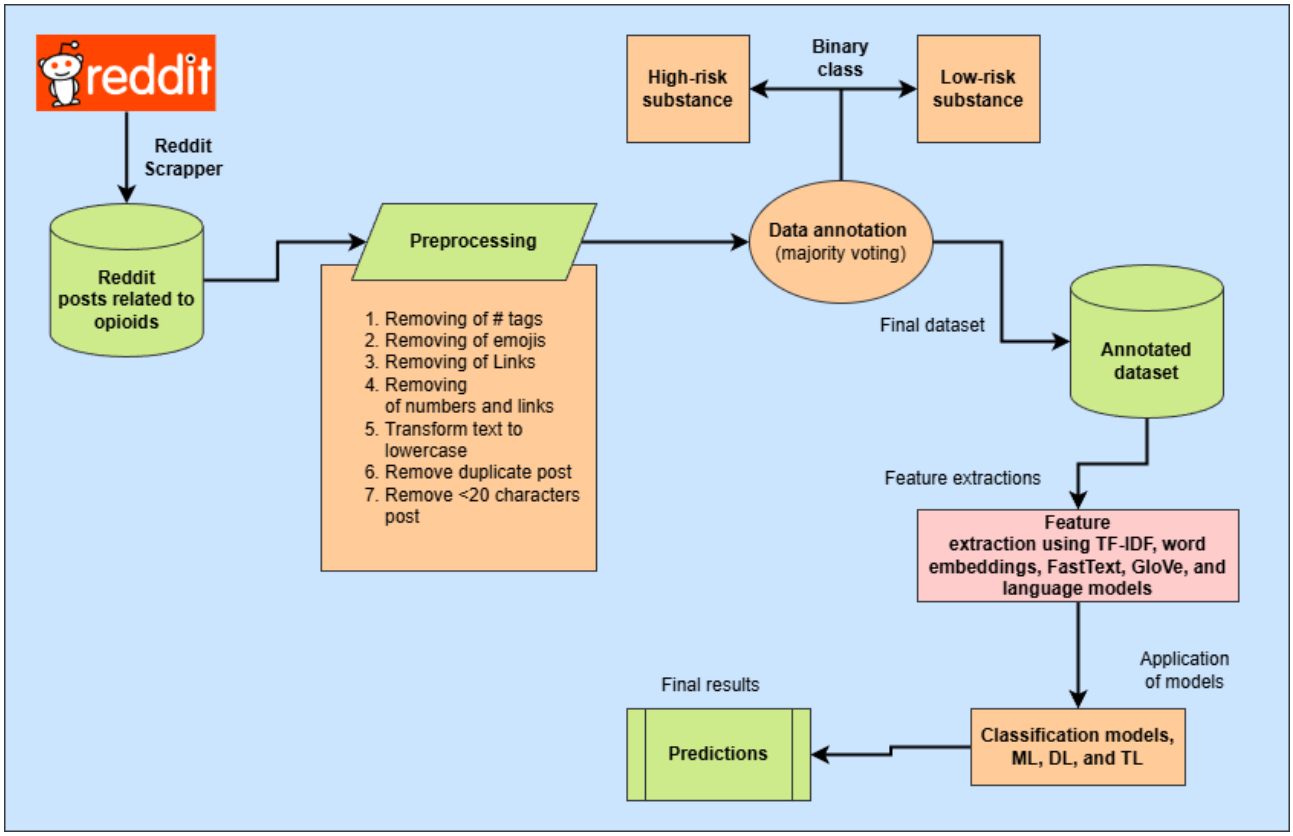
Category	Keywords used
High-risk substance use	overdose, fentanyl, heroin, oxycodone, opioid abuse, withdrawal, painkillers, substance dependence, addiction, relapse, injecting opioids, drug misuse, narcotics, high dosage, opioid crisis
Low-risk substance use	pain management, prescribed medication, controlled dosage, opioid awareness, recovery, tapering off, medical prescription, safe use, harm reduction, medication-assisted treatment (MAT), doctor-recommended, responsible use
General keywords	opioid, pain relief, medication, painkillers, chronic pain, prescription drugs, opioid discussion, rehab, opioid safety, self-medication

A Python-based data collection pipeline was developed using a keyword dictionary to extract Reddit posts via the Pushshift.io application programming interface (developed by Jason Baumgartner). Covering the period from January 1, 2022, to August 8, 2024, a total of 30,000 posts were gathered from several opioid-related subreddits, including r/opiates [38] and r/OpiatesRecovery [39], r/heroine [40], and r/addiction [41]. The posts were divided into 2 categories, namely, high-risk opioid use (eg, overdose cases and addiction symptoms) and low-risk opioid use (eg, responsible medication usage and recovery experiences). Moreover, each post was manually labeled either as high-risk or low-risk opioid use. Consequently, the dataset provides a strong foundation for training machine learning models to identify risk patterns in opioid-related social media discourse and to improve public health interventions.

A comprehensive text preprocessing pipeline was created to standardize the content and minimize noise in the collected posts. It converted all text to lowercase, expanded contractions, and translated common abbreviations using a custom-built dictionary specifically adapted for adolescent languages (eg, converting “idk” to “I do not know”). Moreover, it removed all the HTML tags, numeric values, punctuation marks, and special characters. Furthermore, it normalized elongated or repeated characters in words (eg, transforming “loooove” into “love”), filtered standard and domain-specific stop words using an extended version of the Natural Language Toolkit. English stop word list, and applied lemmatization to reduce words to their root forms. Duplicate entries were eliminated based on the cleaned content. Posts with fewer than a specific number of characters (eg, 20) were discarded to maintain the quality and relevance of the dataset. Finally, posts irrelevant to the selected topic or written in languages other than the specified ones were removed manually. Consequently, a cleaned dataset consisting of 4739 posts was prepared for analysis. Figure 1 illustrates the overall architecture and methodology of the proposed approach.

XSL•FO
RenderX

Figure 1. Proposed architecture and design. DL: deep learning; GloVe: Global Vectors for Word Representation; ML: machine learning; TL: transfer learning; TF-IDF: term frequency–inverse document frequency.



Annotation

Annotation is the process of assigning a predefined label to each sample to create a structured dataset that can be used to train a machine learning model. In opioid substance classification, annotation plays an important role in correctly categorizing each sample as either high-risk or low-risk based on its content.

Annotation Guidelines

To ensure consistency and reliability in the annotation process, the following 13 detailed annotation guidelines were developed to support the binary classification of Reddit posts:

1. Explicit mentions of high-risk substances: read each sample carefully and check whether the user consumes substances such as heroin, fentanyl, oxycodone, morphine, methadone, or other strong opioids that are labeled as high-risk.
2. Explicit mentions of low-risk substances: read carefully and check whether the social media posts mention medications such as paracetamol, ibuprofen, aspirin, tramadol (in moderate doses), or nonopioid pain relievers that are labeled as low-risk.
3. Overdose mentions: posts describing overdose symptoms, emergencies, or near-fatal incidents are categorized as high-risk.
4. Illegal drug transactions: if a user mentions buying, selling, or obtaining opioids illegally (especially from online sources), the post is labeled as high-risk.
5. Prescription use vs misuse: if any posts referring to prescription adherence are labeled as low-risk, whereas discussions about taking more than the prescribed dose are considered high-risk.
6. Polysubstance use: if a post mentions mixing opioids with alcohol, benzodiazepines, or other sedatives, it is classified as high-risk due to increased overdose potential.
7. Self-medication for pain management: if a user in the post mentions self-medicating with low-risk pain relievers such as ibuprofen, aspirin, or paracetamol, it is classified as low-risk.
8. Withdrawal symptoms and addiction struggles: if any user in the posts reports withdrawal symptoms, cravings, or dependence on opioids, the post is labeled as high-risk.
9. Recreational use: if any user in the post explicitly mentions opioid use for recreational purposes or to get high, it is categorized as high-risk.
10. Seeking help or support: if the post seeks medical advice, rehabilitation, or addiction recovery support, it is labeled based on the context—if high-risk substances are mentioned, it is high-risk; otherwise, it is labeled as low-risk.
11. Medical advice discussions: if users discuss doctor-prescribed medications for pain management, the post is classified as low-risk unless signs of misuse are present.
12. Harm reduction strategies: if a post provides information on safe opioid use, naloxone administration, or overdose prevention, it is categorized as low-risk.
13. Slang and codewords: posts containing slang terms for high-risk opioids (eg, “H” for heroin and “china white” for fentanyl) are considered high-risk, while those referring to nonopioid alternatives remain low-risk.

XSL•FO
RenderX

Table 2 provides representative examples of Reddit posts categorized as high-risk and low-risk substance use based on the above guidelines. These examples illustrate the distinctions in language patterns, context, and behavioral cues that guided the annotation process.

Table 2. Examples of high-risk and low-risk substance use posts taken from the dataset.

Post	Risk level
The first time I tried heroin, I took a high dose, and it was the worst experience of my life. I felt an intense rush, but then everything went dark. I couldn't think straight, and I felt like I was going to pass out. It scared me how quickly I lost control, and I swore I'd never touch it again. It's a dangerous substance, and I regret ever trying it.	High risk
I was prescribed oxycodone after surgery, and at first, it helped with the pain. But when I took a higher dose for pain management, it made me feel dizzy and out of it, like I wasn't fully there. It was too much, and I started to realize how easily you can lose control.	High risk
I've been using kratom for a while to help with chronic pain, as recommended by my doctor. The first time I tried it, it gave me a mild sense of relief—like a calm energy boost, without any highs or intense effects. It's been a good alternative for me, and I can take it without feeling out of control. My doctor keeps track of my usage, and it feels safer than other options.	Low risk
I've been using paracetamol for years to manage minor aches and pains, like headaches or body aches. It's effective and works fast, without making me feel drowsy or out of it. I take it as directed by my doctor and never exceed the recommended dose, as it's important to avoid liver damage. It's a safe, low-risk option for occasional pain relief.	Low risk

Annotation Selection

To ensure high-quality annotations for the dataset, a 2-stage selection process was devised to identify the most reliable annotators. In the first round, 500 Reddit posts were provided to 8 annotators to label the posts based on the above-mentioned annotation guidelines. After evaluating their responses, it was found that only 5 annotators assigned nearly identical labels, indicating consistency in their understanding of the task. In the second round, another set of 500 posts was assigned to the 5 annotators who showed consistency in their understanding. Their responses were again evaluated; only 3 annotators demonstrated high-quality labeling with strong agreement, making them the most reliable for dataset construction. To track annotator performance, Google Forms were used for individual assessments. Google Forms played a critical role in (1) standardizing the annotation process, ensuring all annotators follow the same guidelines; (2) collecting and analyzing responses, identifying discrepancies in labeling; and (3) monitoring consistency over time, assisting in selecting the most reliable annotators. The final 3 annotators were master's students with strong backgrounds in machine learning and NLP. While they were not clinical or addiction medicine professionals, the annotation guidelines were developed in consultation with domain experts to ensure clinical relevance. Annotators received training based on these guidelines, and ambiguous cases were resolved through collaborative discussion and expert input when necessary. To address any uncertainties or disagreements in annotation, we scheduled weekly meetings to discuss and resolve ambiguities by majority voting, as provided in Figure 1. Through this rigorous selection and validation process, it was ensured that the dataset was robust, accurate, and reliable for further processing and model training.

Interannotator Agreement

Interannotator agreement is a process used to measure the consistency and reliability of annotations between different annotators. It supports the assessment of whether the annotations are reliable or whether there are inconsistencies that need to be resolved. In this paper, Cohen κ was calculated to evaluate the agreement between annotators for classifying Reddit samples into a binary class. In this study, the Cohen κ value was 0.79, which shows a substantial agreement between the annotators, as shown in Table 3. This level of agreement is sufficient for ensuring that the annotations are consistent and reliable for further analysis and model training.

Table 3. Interpretation of the κ values.

Cohen κ value range	Interpretation
1.0	Perfect agreement
0.80-1.00	Substantial agreement
0.60-0.80	Moderate agreement
0.40-0.60	Fair agreement
<0.40	Poor agreement

Dataset Statistics

A comprehensive statistical analysis was performed to gain a deeper understanding of the dataset's composition and linguistic characteristics. Multimedia Appendix 1 provides a word cloud comprising keywords extracted from posts in the dataset related to the topic of opioid use and overdose. The word cloud visually highlights the most frequent terms, emphasizing the critical themes discussed in the dataset. Multimedia Appendix 2 provides the distribution of labels for each class used in the

XSL•FO
RenderX

corpus for sentiment analysis. The chart visually represents the frequency of each sentiment class in the dataset. Multimedia Appendix 3 provides dataset statistics related to high-risk or low-risk substance use. It includes 4739 posts, totaling 344,827 words and 1,567,169 characters, with a vocabulary of 14,905 unique words. On average, each sentence contains 16.65 words, and each post consists of about 4.37 sentences. The average word length is 4.54 characters, and each post typically contains 330.7 characters. These metrics help analyze language complexity, engagement, and potential differences in how individuals discuss substance use risk contexts.

Feature Extraction

To extract relevant and valuable features for opioid overdose risk detection, a combination of traditional and modern feature extraction techniques was used: term frequency–inverse document frequency (TF-IDF), pretrained word embeddings (FastText and Global Vectors for Word Representation [GloVe]), and contextualized embeddings derived from transformer-based language models. The TF-IDF approach represents textual data by weighing each word based on its frequency in a document and its rarity across the entire corpus. This technique effectively downplays common but uninformative words while emphasizing more meaningful terms for classification. TF-IDF consists of 2 components, term frequency (TF) and inverse document frequency (IDF). The TF measures the relative frequency of a word with respect to the total number of words in the document:



The IDF of a term reflects the inverse proportion of documents containing that term. Terms with technical jargon hold greater significance than words found in only a small percentage of all documents. The IDF is calculated using the following equation:



Finally, the TF-IDF score for each term is obtained by multiplying its TF and IDF values, as shown in equation 3:

TF – IDF = TF × IDF (3)

This representation enhances the ability of machine learning models to capture the most informative linguistic patterns associated with opioid-related risk discourse.

Two dense vector representations based on pretrained word embeddings are used: FastText and GloVe. These techniques provide fixed-dimensional representations while capturing semantic and syntactic properties. The FastText technique represents words as bags of character n-grams, enabling better handling of out-of-vocabulary or misspelled words common in noisy social media text. Specifically, it extends Word2Vec by representing words as bags of character n-grams. The embedding for a word “w” is calculated as follows:



- G (w) is the set of character n-grams in the word w.
- Vg is the vector representation of each n-gram g.

The GloVe generate word embedding based on the global word co-occurrence matrix of words, effectively modeling semantic relationships. The cost value for GloVe is computed using the ratio of co-occurrence probabilities:



Where:

- Xi,jis the number of times word j occurs in the context of word i.
- V is the vocabulary size.
- Vi and Vj are the embeddings for words i and j.
- bi and bjare bias terms for the words.
- f(Xi,j) is a weighting function to down-weight the influence of very frequent words.

The FastText and GloVe techniques are selected for study due to their complementary strengths: FastText addresses linguistic variations and out-of-vocabulary words in Reddit discussions on opioid use, while GloVe captures broader semantic relationships across the corpus. These embeddings provide richer representations than sparse features and were used to train deep learning architectures such as a convolutional neural network (CNN) and a bidirectional long short-term memory (BiLSTM). It is noted that Word2Vec is not used because it lacks subword-level understanding and relies mainly on local context windows, making it less robust for the lexical noise present in a dataset.

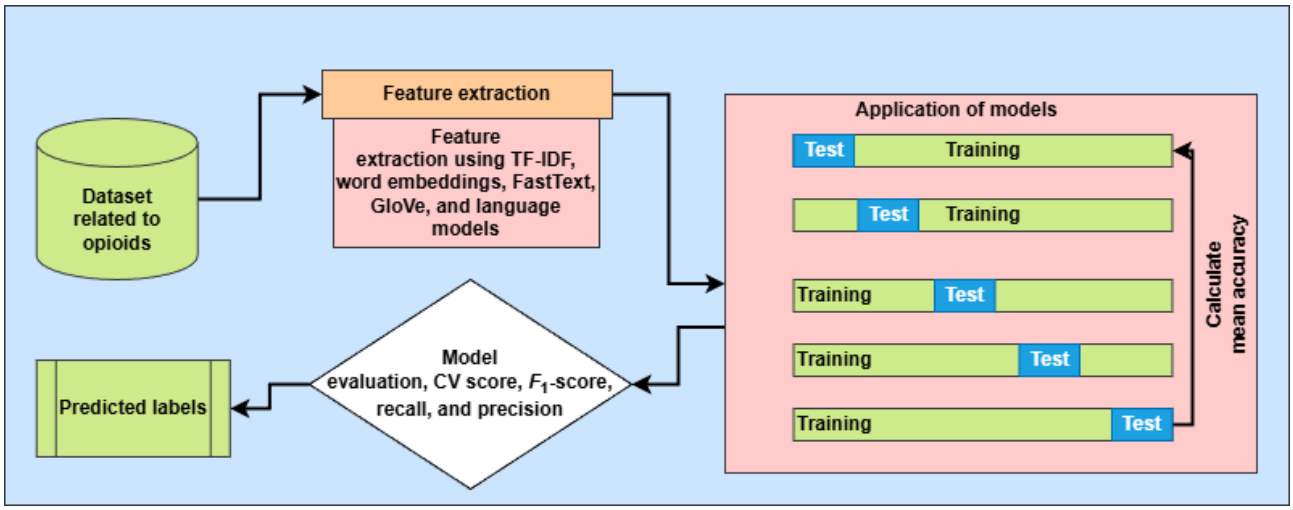
To further enhance contextual understanding, transformer-based language models are used to extract contextualized embeddings. These models generate dynamic representations of words based on their surrounding context, addressing challenges such as polysemy, ambiguous expressions, and informal language.

Application of Machine Learning Models

A diverse set of machine learning models is applied to the opioid dataset for thorough analysis and accurate classification of social media posts (Figure 2). This includes transformer-based architectures with a custom attention mechanism, such as BioBERT, Robustly Optimized BERT Pretraining Approach, XLM-RoBERTa (Cross-lingual Language Model–Robustly Optimized BERT Pretraining Approach), and ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately), traditional machine learning models (decision tree [DT], extreme Gradient Boosting [XGB], multinomial naive Bayes [MNB], and adaptive boosting [AdaBoost]), and commonly used deep learning models (BiLSTM and CNN). Each model in the set captures different aspects of the analysis and classification tasks. For example, DT offers interpretability and handles nonlinear decision boundaries; XGB and AdaBoost provide ensemble-based robustness for high-dimensional features; MNB serves as an efficient probabilistic baseline; CNN captures local n-gram patterns, while BiLSTM models long-range dependencies in both directions.

XSL•FO
RenderX

Figure 2. Application of machine learning models. CV: cross-validation; GloVe: Global Vectors for Word Representation; TF-IDF: term frequency–inverse document frequency.



The contribution of this work lies in a methodologically rigorous evaluation pipeline for assessing opioid risk levels in social media posts, integrating multiple learning paradigms and focusing on real-world Reddit data. Central to this approach is a fine-tuned BioBERT model with a custom attention mechanism, capturing nuanced linguistic and contextual cues.

To enhance the transformer models' ability to capture task-specific linguistic cues in opioid-related discourse, a custom attention layer is used on top of the transformer's contextualized representations, which—unlike standard self-attention—assigns greater weights to domain-relevant tokens (eg, "OD," "relapsed," "detox," and "pill count") through a learnable scoring function. This enables the model to emphasize medically significant phrases, emotional expressions, and slang terms, ultimately improving both the interpretability and precision of risk-level predictions.

Table 4 provides the hyperparameters and their grid search values across models, demonstrating careful attention to optimization and reproducibility. For transformer-based models, the custom attention mechanism dynamically weighs token embeddings. Hyperparameters include a learning rate of 3×10 , 3 epochs, batch size 16, AdamW optimizer, and CrossEntropyLoss for multiclass classification. The attention layer consists of 2 linear layers ($768 \rightarrow 512 \rightarrow 1$) with a Tanh activation, introducing nonlinearity and enabling token-level focus.

XSL•FO
RenderX

Table 4. Optimum values for the hyperparameters of the proposed models.

Learning approach and models	Hyperparameters	Grid search values
Transformer		
BERT[a], RoBERTa[b], XLM-RoBERTa[c], and ELECTRA[d] (with custom attention)	Learning rate, epochs, batch size, optimizer, and loss function	3 × 10 , 3, 16, AdamW, and CrossEntropy-Loss
Machine learning		
XGB[e]	n_estimators, max_depth, and learning_rate	100, 6, and 0.3
MNB[f]	Alpha (smoothing parameter) and fit prior (fit_prior)	alpha: 0.1, 0.5, 1.0 and fit_prior: true, false
DT[g]	Maximum depth of the tree (max_depth), minimum samples for splitting (min_samples_split), minimum samples for leaf nodes (min_samples_leaf), and maximum features for splitting (max_features)	max_depth: 5, 10, 20, and none; min_samples_split: 2, 5, and 10; min_samples_leaf: 1, 2, and 4; and max_features: “sqrt,” “log2,” and none
AdaBoost[h]	Number of estimators (n_estimators), learning rate (learning_rate), and base estimator (base_estimator)	n_estimators: 50, 100, and 200; learning_rate: 0.01, 0.1, and 1.0; base_estimator: DecisionTreeClassifier(max_depth=1)
Deep learning		
BiLSTM[i]	Learning rate, epochs, Embedding_dim, batch size, and LSTM[j] units	0.1, 5, 300, 32, and 128
CNN[k]	Learning rate, epochs, Embedding_dim, batch size, filters, and kernel size	0.1, 5, 300, 32, 128, and 5

[a]BERT: Bidirectional Encoder Representations from Transformers.

[b]RoBERTa: Robustly Optimized BERT Pretraining Approach.

[c]XLM-RoBERTa: Cross-lingual Language Model–Robustly Optimized BERT Pretraining Approach.

[d]ELECTRA: Efficiently Learning an Encoder that Classifies Token Replacements Accurately.

[e]XGB: extreme Gradient Boosting.

[f]MNB: multinomial naive Bayes.

[g]DT: decision tree.

[h]AdaBoost: adaptive boosting.

[i]BiLSTM: bidirectional long short-term memory.

[j]LSTM: long short-term memory.

[k]CNN: convolutional neural network.

Traditional machine learning models are tuned as follows: XGB used 100 estimators, a max depth of 6, and a learning rate of 0.3; MNB used alpha values from 0.1 to 1.0 and fit_prior was set to true or false; DT explored max_depth (5, 10, 20, and none), min_samples_split (2, 5, and 10), and other parameters governing tree complexity; AdaBoost used n_estimators (50, 100, and 200), learning_rate (0.01, 0.1, and 1.0), and a base DT with max_depth 1.

Deep learning models are optimized with BiLSTM using a learning rate of 0.1, 5 epochs, an embedding dimension of 300, a batch size of 32, and 128 long short-term memory units. CNN has similar hyperparameters with 128 filters and a kernel size of 5. These settings are selected to balance learning capacity, computational efficiency, and model performance, optimized through grid search on validation data.

To assess generalization, the proposed model is trained on posts from 2022 to 2023 and evaluated on posts from 2024. The 5-fold cross-validation (CV) and paired *t* tests are applied to ensure replicable and statistically robust performance. This strategy bridges domain-specific insights with advanced computational techniques, yielding results relevant for public health surveillance and intervention strategies.

Ethical Considerations

This study was conducted in accordance with established ethical standards for research involving online data. The dataset was compiled exclusively from publicly accessible social media platforms. “Publicly available” refers to content that can be accessed freely by any internet user without login requirements, private group membership, or special permissions.

As the research relied solely on publicly available data and did not involve interaction with human participants, intervention, or access to private or restricted information, formal Institutional Review Board (IRB) approval was not required under standard ethical guidelines governing research on publicly accessible online content.

To ensure the protection of user privacy and minimize the risk of re-identification, rigorous safeguards were implemented. No personally identifiable information was collected or stored. All direct identifiers (eg, usernames, profile details, account IDs) were removed during data preprocessing, and the dataset was

XSL•FO
RenderX

fully anonymized prior to analysis. Verbatim quotes were avoided or paraphrased to prevent traceability through search engines, and any contextual details that could potentially enable identification were excluded. All analyses were conducted and reported in aggregate form.

These procedures were undertaken to ensure responsible, privacy-conscious, and trustworthy use of publicly available online data.

Results

Traditional Machine Learning Results

This section provides the experimental results of the machine learning models applied to opioid risk detection using social mdia data. The traditional machine learning models, such as DT, XGB, AdaBoost, and MNB, are evaluated based on their performance in classifying high-risk and low-risk opioid-related posts.

Table 5 provides the results of 4 different machine learning models—XGB, MNB, DT, and AdaBoost—on a binary classification task to identify the opioid use risk as either high or low. The performance of each model is evaluated based on 4 standard evaluation metrics, that is, precision, recall, F_1-score, and CV score. Across all metrics, XGB outperforms all other models with a CV score of 0.97, showing its higher capability in distinguishing between high-risk and low-risk cases. DT follows with scores ranging from 0.94 to 0.95, indicating solid but slightly lower performance. AdaBoost also achieves comparable performance, maintaining steady scores of 0.94 across all metrics. On the other hand, MNB shows the weakest performance, especially in recall and F_1-score (0.93), showing limitations in capturing the nuances of the classification problem. Overall, XGB appears to be the best-performing model for this task in identifying opioid use risk.

Table 5. Results for machine learning models.

Metric	XGB[a]	MNB[b]	DT[c]	AdaBoost[d]
Precision	0.97	0.94	0.95	0.94
Recall	0.97	0.93	0.95	0.94
F_1-score	0.97	0.93	0.95	0.94
CV[e] score	0.97	0.93	0.95	0.94

[a]XGB: extreme Gradient Boosting.

[b]MNB: multinomial naive Bayes.

[c]DT: decision tree.

[d]AdaBoost: adaptive boosting.

[e]CV: cross-validation.

Deep Learning Model Results

This section provides the experimental results of the 2 commonly used deep learning models, namely, BiLSTM and CNN. By leveraging the power of deep learning, these models are used to capture more complex patterns in the data and improve the detection of opioid misuse and overdose risks.

Table 6 compares the experimental results of deep learning models—CNN and BiLSTM—using 2 different word embeddings: FastText and GloVe, for the same binary classification task of identifying high- vs low-risk of opioid use. With FastText embeddings, both CNN and BiLSTM perform well, with BiLSTM slightly outperforming CNN across all metrics (precision, recall, F_1-score, and CV score) at 0.94 compared to CNN's consistent 0.92. When using GloVe embeddings, CNN demonstrates even stronger performance, achieving 0.95 across all metrics, making it the best-performing model among the deep learning approaches evaluated. However, BiLSTM with GloVe embeddings underperforms significantly, especially in precision (0.83) and F_1-score (0.85), suggesting that GloVe may not complement BiLSTM as effectively as FastText does for this specific task. Overall, the GloVe + CNN combination provides the highest performance, while the FastText + BiLSTM combination offers a strong alternative, indicating the critical influence of embedding choice on model effectiveness in opioid use risk classification.

XSL•FO
RenderX

Table 6. Results for deep learning models.

Models	Precision	Recall	F_1-score	CV[a] score
FastText				
CNN[b]	0.92	0.92	0.92	0.92
BiLSTM[c]	0.94	0.94	0.94	0.94
GloVe [d]				
CNN	0.95	0.95	0.95	0.95
BiLSTM	0.83	0.88	0.85	0.88

[a]CV: cross-validation.

[b]CNN: convolutional neural network.

[c]BiLSTM: bidirectional long short-term memory.

[d]GloVe: Global Vectors for Word Representation.

Transformer Model Results

This section provides the results of applying transformer-based models to the opioid risk detection task. These models are selected based on their superior performance in preliminary experiments and their widespread success in text classification tasks.

Table 7 provides the performance of transformer-based models—ELECTRA-base-discriminator, BioBERT, RoBERTa-base, and XLM-RoBERTa—each enhanced with a custom attention mechanism tailored for the binary classification of opioid use risk (high vs low). Among these, BioBERT demonstrates the highest performance, achieving near-perfect scores of 0.99 in precision, recall, F_1-score, and CV score, underscoring the synergy between domain-specific pretraining and task-specific attention modeling. The remaining models—ELECTRA, Robustly Optimized BERT Pretraining Approach, and XLM-RoBERTa—also perform robustly, each scoring 0.98 across all metrics, indicating the effectiveness of the custom attention mechanism in improving general transformer architectures for the health-related NLP task.

Table 7. Transformer results.

Model	Precision	Recall	F_1-score	CV[a] score
electra[b]-base-discriminator	0.98	0.98	0.98	0.98
biobert[c]-base-cased-v1.1	0.99	0.99	0.99	0.99
roberta[d]-base	0.98	0.98	0.98	0.98
xlm[e]-roberta	0.98	0.98	0.98	0.98
Train test split				
biobert-base-cased-v1.1	0.99	0.99	0.99	0.99
Without a custom attention mechanism				
biobert-base-cased-v1.1	0.96	0.96	0.96	0.96

[a]CV: cross-validation.

[b]ELECTRA: Efficiently Learning an Encoder that Classifies Token Replacements Accurately.

[c]BioBERT: bidirectional encoder representations from transformers for biomedical text mining.

[d]RoBERTa: Robustly Optimized BERT Pretraining Approach.

[e]XLM-RoBERTa: Cross-lingual Language Model–Robustly Optimized BERT Pretraining Approach.

To evaluate the model on unseen data, the top-performing model is trained on posts from 2022 to 2023, saved, and then tested on a held-out set of approximately 948 posts from 2024 to assess its generalization performance. BioBERT maintained a precision, recall, and F_1-score of 0.99 in this temporal evaluation, demonstrating that the model generalizes effectively to future, unseen data. These results indicate that the model's performance is robust and not an artifact of overfitting, confirming its reliability for real-world applications. Additionally, these results confirm that while transformer models are inherently powerful, incorporating a custom attention mechanism further refines their focus on risk-relevant textual cues, especially when combined with a domain-adapted model such as BioBERT.

It is noted that the proposed model without the custom attention mechanism achieved 0.96 across accuracy, precision, recall, and F_1-score, showing that it already captures much of the context in opioid-related text. Overall, by incorporating the custom attention layer, the performance improved to 0.99 on

XSL•FO
RenderX

all metrics, demonstrating that emphasizing domain-specific tokens and phrases allows the model to make more precise and interpretable risk-level predictions.

Error Analysis

Table 8, titled "top-performing models in each learning approach," compares the highest-performing models from 3 distinct machine learning paradigms—XGB from traditional machine learning, CNN with GloVe embeddings from deep learning, and BioBERT model enhanced with custom attention mechanisms from transformer-based NLP—for classifying high- and low-risk opioid substances. The performance is measured across 4 metrics: precision, recall, F_1-score, and CV score. The BioBERT model leads with a perfect score of 0.99 in all categories, showcasing its exceptional ability to capture deep semantic relationships within the data. XGB follows with solid performance, achieving 0.97 across the board, indicating that it remains a competitive option despite being a nondeep learning model. CNN (GloVe), while effective, ranks third with consistent scores of 0.95, suggesting that although it leverages word embeddings, it may lack the contextual understanding of transformers. This comparison clearly highlights the dominance of transformer models, such as Bidirectional Encoder Representations from Transformers, in handling complex, text-based substance risk classification tasks.

Table 8. Top-performing models in each learning approach.

Metric	XGB[a]	CNN[b] (GloVe[c])	BioBERT[d]
Precision	0.97	0.95	0.99
Recall	0.97	0.95	0.99
F_1-score	0.97	0.95	0.99
CV[e] score	0.97	0.95	0.99

[a]XGB: **e**xtreme Gradient Boosting.

[b]CNN: convolutional neural network.

[c]GloVe: Global Vectors for Word Representation.

[d]BioBERT: bidirectional encoder representations from transformers for biomedical text mining.

[e]CV: cross-validation.

The superior performance of the BioBERT model can be attributed not only to its bidirectional transformer architecture, which effectively captures deep contextual relationships in text data compared to traditional and shallow models such as XGB and CNN, but also to the integration of custom attention mechanisms that enable the model to focus on the most relevant parts of the input. This enhancement improves the detection of subtle, domain-specific patterns crucial for opioid risk classification.

Table 9 provides the performance of a classification model in distinguishing between "Low Risk" and "High Risk" classes using a 5-fold CV score. The model performs exceptionally well for both categories, achieving high precision, recall, and F_1-scores. For "Low Risk," precision is 0.98, recall is 0.99, and the F_1-score is 0.99, based on 1921 samples. Similarly, for "High Risk," precision is 0.99, recall is 0.99, and the F_1-score is 0.99, with 2601 samples. These results indicate a well-balanced model that accurately identifies both risk levels, with minimal misclassification. The confusion matrix for the proposed model based on CV is provided in Figure 3 (A, B, and C), providing a detailed view of true positives, true negatives, false positives, and false negatives. Figure 4 provides the confusion matrix obtained from a train-test split, demonstrating the model's performance on temporally separated, unseen data and highlighting its real-world generalization capability. Finally, Figure 5 shows the confusion matrix of our proposed BioBERT model with the traditional self-attention mechanism.

Table 9. Class-wise evaluation scores of the fine-tuned biobert-base-cased-v1.1 model used in our proposed methodology.

Class	Precision	Recall	F_1-score	Support
Low risk	0.98	0.99	0.99	1907
High risk	0.99	0.99	0.99	2568

XSL•FO
RenderX

Figure 3. Confusion matrix of top-performing models in each learning approach. BioBERT: bidirectional encoder representations from transformers for biomedical text mining; CNN: convolutional neural network; XGB: extreme Gradient Boosting.

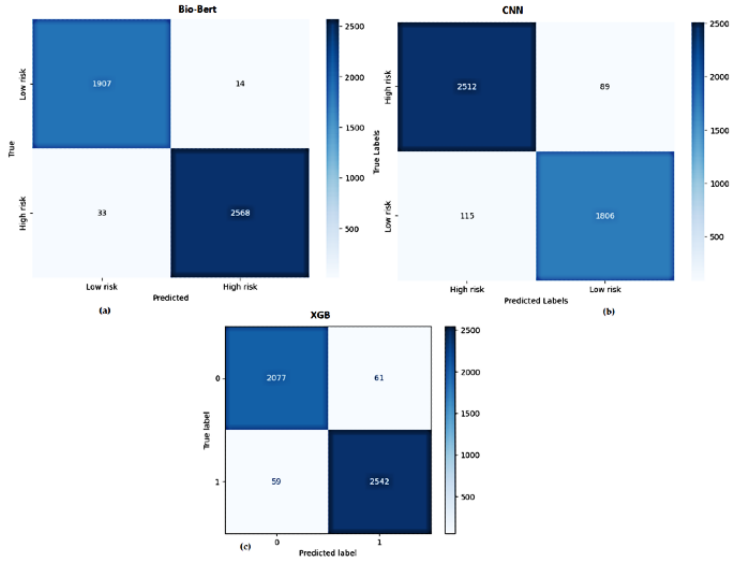
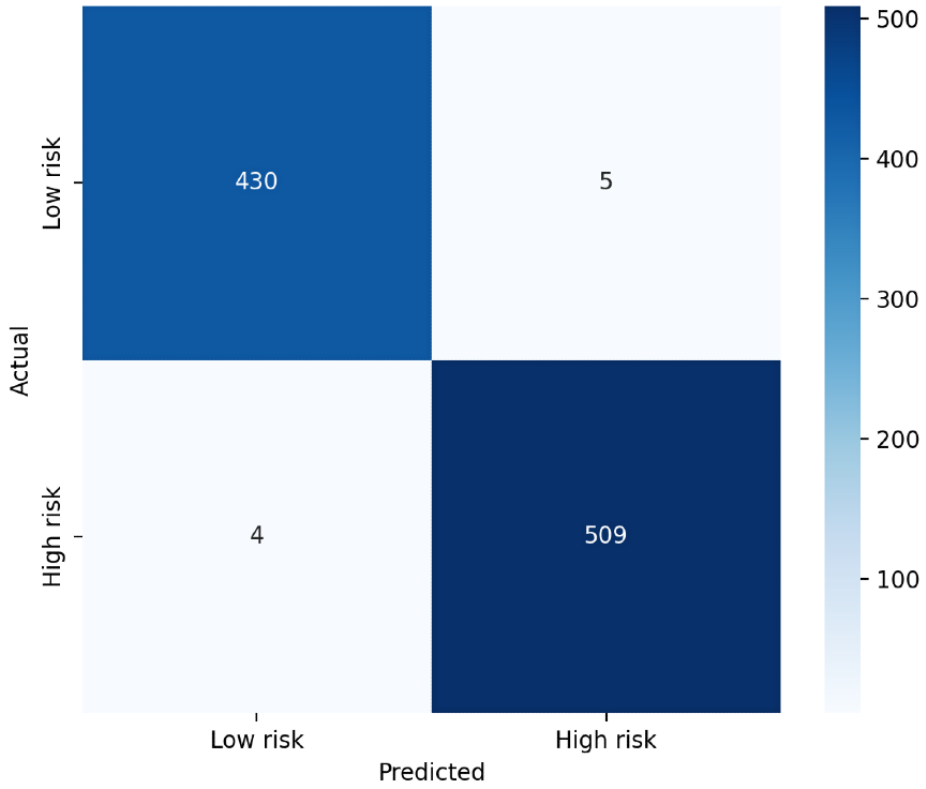
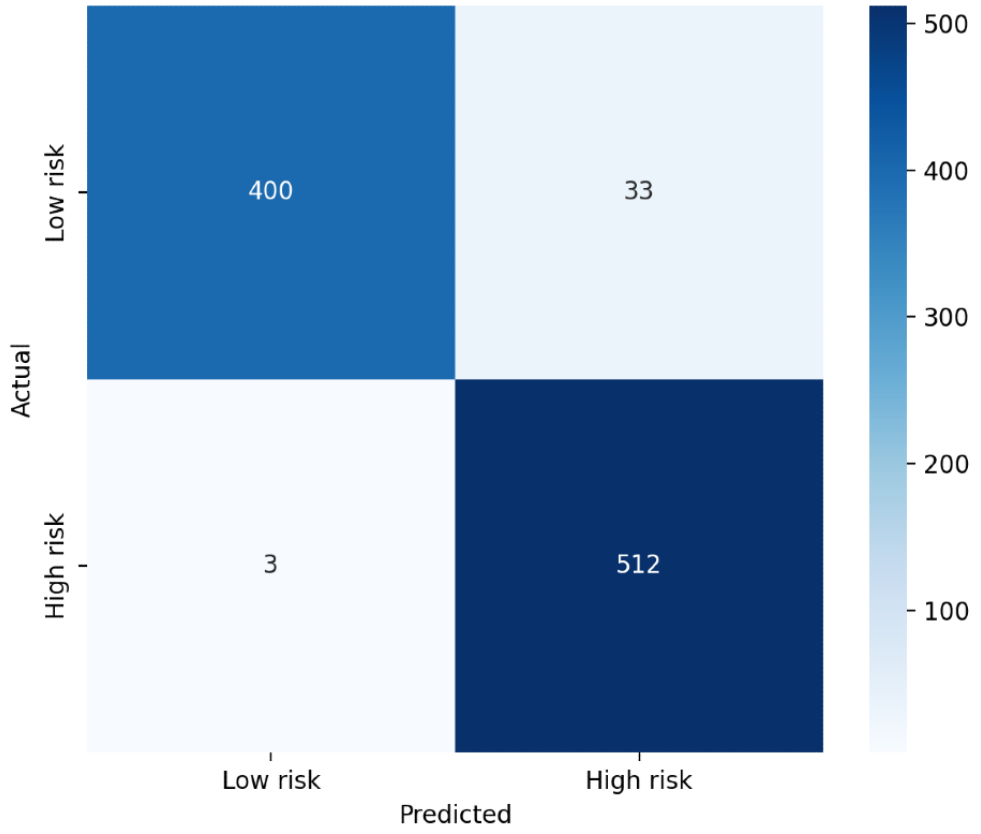


Figure 4. Confusion matrix of bidirectional encoder representations from transformers for biomedical text mining (BioBERT) using 2022-2023 train and 2024 test split.



XSL•FO
RenderX

Figure 5. Confusion matrix of bidirectional encoder representations from transformers for biomedical text mining (BioBERT) using the traditional self-attention mechanism.



Statistical Analysis

To determine whether BioBERT's superior performance over XGB is statistically significant across the 5 folds provided in Table 10 (F1 to F5), a paired t test was conducted using the F_1-scores from each fold. The paired *t* test yielded a test statistic of t_4=6.53 and a *P*=.003. Since the *P* value is well below the conventional threshold of .05, we reject the null hypothesis that there is no difference in performance between the 2 models. This result confirms that BioBERT's higher performance is statistically significant and unlikely to be due to random chance. Therefore, based on the per-fold F_1-scores, we conclude that BioBERT meaningfully and consistently outperforms XGB in this classification task.

Table 10. Comparison of accuracy across 5-fold cross-validation for bidirectional encoder representations from transformers for biomedical text mining (BioBERT) and XGB models.

Models	F1	F2	F3	F4	F5
XGB[a]	0.97	0.98	0.97	0.96	0.98
BioBERT[b]	0.99	0.99	0.99	0.98	0.99

[a]XGB: **e**xtreme Gradient Boosting.

[b]BioBERT: bidirectional encoder representations from transformers for biomedical text mining.

Discussion

Principal Findings

The results of this study underscore the significant potential of using advanced NLP techniques—particularly transformer-based models enhanced with custom attention mechanisms—for identifying opioid overdose risk levels from social media data. The exceptional performance of the proposed BioBERT-attention model, achieving an average CV score of 0.99, demonstrates its strong capability to capture nuanced semantic signals within Reddit posts that indicate high- or low-risk opioid use. This level of accuracy notably surpasses that of traditional and deep learning baselines, including the widely used XGB model, which, despite its robustness, achieved a lower score of 0.97. The statistically significant improvement, confirmed through a paired *t* test, further validates the effectiveness and reliability of the BioBERT-attention framework, affirming that the enhancements introduced by the custom attention layer contribute meaningful improvements rather than random performance fluctuations.

A key factor in this success is the domain-specific nature of BioBERT, which is pretrained on biomedical text and thus well-suited for understanding the health-related language and medical terminology frequently used in user-generated content

XSL•FO
RenderX

discussing opioid use. The addition of a custom attention mechanism further allowed the model to emphasize contextually important cues—such as symptom descriptions, drug names, dosages, and subjective experiences—that are critical for determining the associated risk level. This tailored architecture addresses the shortcomings of generic models that may overlook such task-specific signals.

Moreover, the manual annotation of the dataset, guided by clearly defined labeling criteria, provided a strong foundation for supervised learning by ensuring consistency and high-quality ground truth. This highlights the value of combining expert-driven dataset curation with advanced model architectures in health-related NLP tasks. While the model shows excellent performance on the curated Reddit dataset, future research should explore its generalizability to other social media platforms and real-time deployment settings, where informal language, slang, or multilingual code-switching might pose additional challenges.

Overall, this study contributes a novel and effective approach to automated opioid risk detection by leveraging the rich, self-reported data available on Reddit. The proposed BioBERT-attention model not only achieves state-of-the-art performance but also presents a scalable framework that can be adapted to other substance abuse contexts, potentially transforming how public health agencies monitor and respond to emerging drug use trends in real time.

Limitations

Despite the promising results demonstrated by our fine-tuned BioBERT model in detecting high-risk opioid substance use from Reddit data, several limitations must be acknowledged. First, our dataset is limited to Reddit posts, which may not capture the full diversity of opioid misuse discussions across different social media platforms or real-world contexts. The language, slang, and user demographics on Reddit may differ significantly from other online or offline populations, potentially limiting the generalizability of our findings. Second, although the dataset was manually annotated with detailed guidelines to ensure robustness, the subjective nature of annotation may introduce biases or inconsistencies. Additionally, some key contextual information, such as user history, temporal trends, or multimodal content (images and videos), was not considered, which could provide richer signals for more accurate risk detection. Third, the study focuses on supervised learning models with standard fine-tuning and does not explore more advanced domain adaptation techniques or novel architectures that might improve performance further, especially in handling imbalanced or noisy social media data.

Finally, the current evaluation is limited to computational experiments and statistical testing on CV folds, without real-world clinical validation or deployment. Future work should aim to validate these models in practical health care or intervention settings to assess their true impact and reliability in supporting harm reduction efforts.

Conclusion and Future Work

The findings of our study underscore the significant potential of social media data in identifying opioid misuse patterns and assessing overdose risks. By creating a manually annotated dataset from Reddit, we have developed a valuable resource for categorizing opioids into high- and low-risk groups based on real user experiences. Our carefully crafted annotation guidelines ensure both accuracy and consistency, providing a solid foundation for future research in this area. Moreover, our experimental analysis, using 5-fold CV, demonstrated that transfer learning models using a custom attention mechanism significantly outperform traditional machine learning models such as XGB, achieving a performance improvement of 2.06%. This reinforces the power of AI-driven approaches in detecting opioid-related risks early, facilitating timely interventions and more effective harm reduction strategies. Ultimately, our work contributes to ongoing efforts to combat the opioid crisis by leveraging automation and real-world social media insights to support public health initiatives.

While our model has demonstrated strong performance using Reddit data with 5-fold CV, we acknowledge the importance of further evaluation to ensure broader robustness and generalizability. In future work, we plan to extend our dataset by incorporating opioid-related discussions from additional social media platforms such as Twitter and Facebook. This platform expansion will allow us to test the model's adaptability and address platform shift challenges, ultimately improving cross-platform generalization. Additionally, we recognize the value of temporal evaluation and intend to conduct time-based splits to assess the model's ability to generalize over evolving opioid-related discourse, addressing temporal generalization concerns. We also plan to explore domain-adaptive pretraining techniques to better tailor transformer models to opioid-specific language and context, which is expected to further enhance domain-specific understanding and improve classification robustness.

By pursuing these directions, we aim to develop a more resilient and widely applicable opioid misuse detection framework that can support public health monitoring and intervention across diverse and dynamic social media environments.

Acknowledgments

This work was performed with partial support from the Mexican Government through grant A1-S-47854 of CONACYT, Mexico, and grants 20241816, 20241819, and 20240951 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank CONACYT for the computing resources provided through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico, and acknowledge support from Microsoft through the Microsoft Latin America PhD program.

XSL•FO
RenderX

Funding

No external financial support or grants were received for this work.

Data Availability

The datasets generated or analyzed during this study are not publicly available but are available from the first author upon reasonable request.

Authors' Contributions

The study was conceptualized by MA and AH, and the methodology was developed by RO, MA, and AB. Software development was carried out by MA and AB, and validation was performed by IB and MA. Formal analysis was conducted by IB and GS, while investigation was undertaken by IB and GS. Resources were provided by GS and IB, and data curation was managed by MA. The original draft was prepared by MA and RO, with review and editing conducted by MA and RO. Visualization was completed by MA and AB. Supervision and project administration were handled by GS.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Word cloud visualizing the most frequent terms extracted from the dataset.
[PNG File , 606 KB - infodemiology_v6i1e77783_app1.png]

Multimedia Appendix 2
Dataset statistics.
[PNG File , 113 KB - infodemiology_v6i1e77783_app2.png]

Multimedia Appendix 3
Dataset label distribution.
[PNG File , 53 KB - infodemiology_v6i1e77783_app3.png]

References

1. Collins FS, Koroshetz WJ, Volkow ND. Helping to end addiction over the long-term: the research plan for the NIH HEAL initiative. JAMA 2018;320(2):129-130 [FREE Full text] [doi: 10.1001/jama.2018.8826] [Medline: 29896636]
2. Coussens NP, Sittampalam GS, Jonson SG, Hall MD, Gorby HE, Tamiz AP, et al. The opioid crisis and the future of addiction and pain therapeutics. J Pharmacol Exp Ther 2019;371(2):396-408 [FREE Full text] [doi: 10.1124/jpet.119.259408] [Medline: 31481516]
3. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021;397(10289):2082-2097. [doi: 10.1016/S0140-6736(21)00393-7] [Medline: 34062143]
4. Gatchel RJ, McGeary DD, McGeary CA, Lippe B. Interdisciplinary chronic pain management: past, present, and future. Am Psychol 2014;69(2):119-130. [doi: 10.1037/a0035514] [Medline: 24547798]
5. Gallagher RM. Advancing the pain agenda in the veteran population. Anesthesiol Clin 2016;34(2):357-378. [doi: 10.1016/j.anclin.2016.01.003] [Medline: 27208715]
6. Chronic Pain (Primary and Secondary) in Over 16s: Assessment of all Chronic Pain and Management of Chronic Primary Pain. London: National Institute for Health and Care Excellence (NICE); 2021.
7. Bond C, Hannaford P. Issues related to monitoring the safety of over-the-counter (OTC) medicines. Drug Saf 2003;26(15):1065-1074. [doi: 10.2165/00002018-200326150-00001] [Medline: 14640771]
8. Duong M, Gulmez SE, Salvo F, Abouelfath A, Lassalle R, Droz C, et al. Usage patterns of paracetamol in France. Br J Clin Pharmacol 2016;82(2):498-503 [FREE Full text] [doi: 10.1111/bcp.12957] [Medline: 27059520]
9. Wolf MS, King J, Jacobson K, Di Francesco L, Bailey SC, Mullen R, et al. Risk of unintentional overdose with non-prescription acetaminophen products. J Gen Intern Med 2012;27(12):1587-1593 [FREE Full text] [doi: 10.1007/s11606-012-2096-3] [Medline: 22638604]
10. King JP, Davis TC, Bailey SC, Jacobson KL, Hedlund LA, Di Francesco L, et al. Developing consumer-centered, nonprescription drug labeling a study in acetaminophen. Am J Prev Med 2011;40(6):593-598. [doi: 10.1016/j.amepre.2011.02.016] [Medline: 21565649]
11. Forman RF, Woody GE, McLellan T, Lynch KG. The availability of web sites offering to sell opioid medications without prescriptions. Am J Psychiatry 2006;163(7):1233-1238. [doi: 10.1176/ajp.2006.163.7.1233] [Medline: 16816229]

XSL•FO
RenderX

12. Aubry L, Carr BT. Overdose, opioid treatment admissions and prescription opioid pain reliever relationships: United States, 2010-2019. Front Pain Res (Lausanne) 2022;3:884674 [FREE Full text] [doi: 10.3389/fpain.2022.884674] [Medline: 35992020]
13. Ahmad M, Sidorov G, Amjad M, Ameer I, Batyrshin I. Opioid crisis detection in social media discourse using deep learning approach. Information 2025;16(7):545. [doi: 10.3390/info16070545]
14. Bachhuber MA, Cunningham CO. Availability of buprenorphine on the internet for purchase without a prescription. Drug Alcohol Depend 2013;130(1-3):238-240 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.11.004] [Medline: 23201172]
15. Liang BA, Mackey T. Searching for safety: addressing search engine, website, and provider accountability for illicit online drug sales. Am J Law Med 2009;35(1):125-184. [doi: 10.1177/009885880903500104] [Medline: 19534258]
16. Bunting AM, Krawczyk N, Lippincott T, Gu Y, Arya S, Nagappala S, et al. Trends in fentanyl content on reddit substance use forums, 2013-2021. J Gen Intern Med 2023;38(15):3283-3287 [FREE Full text] [doi: 10.1007/s11606-023-08256-7] [Medline: 37296360]
17. Garg S, Taylor J, El Sherief M, Kasson E, Aledavood T, Riordan R, et al. Detecting risk level in individuals misusing fentanyl utilizing posts from an online community on Reddit. Internet Interv 2021;26:100467 [FREE Full text] [doi: 10.1016/j.invent.2021.100467] [Medline: 34804810]
18. List of most-visited websites. Wikipedia. URL: https://en.wikipedia.org/wiki/List_of_most-visited_websites [accessed 2025-02-01]
19. Sarker A, Nataraj N, Siu W, Li S, Jones CM, Sumner SA. Concerns among people who use opioids during the COVID-19 pandemic: a natural language processing analysis of social media posts. Subst Abuse Treat Prev Policy 2022;17(1):16 [FREE Full text] [doi: 10.1186/s13011-022-00442-w] [Medline: 35248103]
20. Chenworth M, Perrone J, Love JS, Graves R, Hogg-Bremer W, Sarker A. Methadone and suboxone® mentions on Twitter: thematic and sentiment analysis. Clin Toxicol (Phila) 2021;59(11):982-991 [FREE Full text] [doi: 10.1080/15563650.2021.1893742] [Medline: 33821724]
21. Stokes DC, Purtle J, Meisel ZF, Agarwal AK. State legislators' divergent social media response to the opioid epidemic from 2014 to 2019: longitudinal topic modeling analysis. J Gen Intern Med 2021;36(11):3373-3382 [FREE Full text] [doi: 10.1007/s11606-021-06678-9] [Medline: 33782896]
22. Ahmad M, Ullah F, Usman M, Habiba U, Batyrshin I, Sidorov G. Leveraging large language models for multi-class and multi-label detection of drug use and overdose symptoms on social media. arXiv. Preprint posted online on April 16, 2025 2025. [doi: 10.48550/arXiv.2504.12355]
23. Preiss A, Baumgartner P, Edlund MJ, Bobashev GV. Using named entity recognition to identify substances used in the self-medication of opioid withdrawal: natural language processing study of Reddit data. JMIR Form Res 2022;6(3):e33919 [FREE Full text] [doi: 10.2196/33919] [Medline: 35353047]
24. Lokala U, Phukan OC, Dastidar TG, Lamy F, Daniulaityte R, Sheth A. Detecting substance use disorder using social media data and the dark web: time- and knowledge-aware study. JMIRx Med 2024;5:e48519 [FREE Full text] [doi: 10.2196/48519] [Medline: 38717384]
25. Ahmad M, Farid H, Ameer I, Muzamil M, Jalal AHM, Batyrshin I, et al. Opioid named entity recognition (ONER-2025) from Reddit. arXiv. Preprint posted online on March 28, 2025 2025. [doi: 10.48550/arXiv.2504.00027]
26. Baird A, Xia Y, Cheng Y. Consumer perceptions of telehealth for mental health or substance abuse: a Twitter-based topic modeling analysis. JAMIA Open 2022;5(2):ooac028 [FREE Full text] [doi: 10.1093/jamiaopen/ooac028] [Medline: 35495736]
27. Ahmad M, Usman S, Farid H, Ameer I, Muzzamil M, Ameer H, et al. Hope speech detection using social media discourse (Posi-Vox-2024): a transfer learning approach. J Lang Teach Res 2024;10(4):31-43. [doi: 10.17323/jle.2024.22443]
28. Ullah F, Zamir MT, Ahmad M, Sidorov G, Gelbukh A. Hope: a multilingual approach to identifying positive communication in social media. 2024 Presented at: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS. Org; September 24, 2024; Valladolid, Spain URL: https://tinyurl.com/4z8a9djm
29. Chi Y, Chen HY. Investigating substance use via Reddit: systematic scoping review. J Med Internet Res 2023;25:e48905 [FREE Full text] [doi: 10.2196/48905] [Medline: 37878361]
30. Wanchoo K, Abrams M, Merchant RM, Ungar L, Guntuku SC. Reddit language indicates changes associated with diet, physical activity, substance use, and smoking during COVID-19. PLoS One 2023;18(2):e0280337 [FREE Full text] [doi: 10.1371/journal.pone.0280337] [Medline: 36735708]
31. Dunn J, Schifano F, Dudley E, Arillotta D, Guirguis A. From veterinary medicine to illicit drug supply: utilising social media to explore the rising emergence of veterinary medicines in human health. Brain Sci 2025;15(2):172 [FREE Full text] [doi: 10.3390/brainsci15020172] [Medline: 40002505]
32. Jha D, La Marca SR, Singh R. Identifying and characterizing opioid addiction states using social media posts. 2021 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 09-12, 2021; Houston, TX p. 913-918. [doi: 10.1109/bibm52615.2021.9669628]
33. Smith DA, Lavertu A, Salecha A, Hamamsy T, Humphreys K, Lembke A, et al. Monitoring the opioid epidemic via social media discussions. NPJ Digit Med 2025;8(1):284 [FREE Full text] [doi: 10.1038/s41746-025-01642-x] [Medline: 40374984]

XSL•FO
RenderX

34. Gandy LM, Ivanitskaya LV, Bacon LL, Bizri-Baryak R. Public health discussions on social media: evaluating automated sentiment analysis methods. JMIR Form Res 2025;9:e57395. [doi: 10.2196/57395] [Medline: 39773420]
35. Yang Z, Nguyen L, Jin F. Predicting opioid relapse using social media data. arXiv. Preprint posted online on November 14, 2018 2018. [doi: 10.48550/arXiv.1811.12169]
36. Li W, Hua Y, Zhou P, Zhou L, Xu X, Yang J. Characterizing public sentiments and drug interactions in the COVID-19 pandemic using social media: natural language processing and network analysis. J Med Internet Res 2025;27:e63755 [FREE Full text] [doi: 10.2196/63755] [Medline: 40053730]
37. Zhang X, Zhu J, Kenne DR, Jin R. Teenager substance use on Reddit: mixed methods computational analysis of frames and emotions. J Med Internet Res 2025;27:e59338 [FREE Full text] [doi: 10.2196/59338] [Medline: 39970428]
38. r/opiates. Reddit. URL: https://www.reddit.com/r/opiates/ [accessed 2026-02-11]
39. r/OpiatesRecovery. Reddit. URL: https://www.reddit.com/r/OpiatesRecovery/ [accessed 2026-02-11]
40. r/heroine. Reddit. URL: https://www.reddit.com/r/heroine/ [accessed 2026-02-11]
41. r/addiction. Reddit. URL: https://www.reddit.com/r/addiction/ [accessed 2026-02-11]

Abbreviations

AdaBoost: adaptive boosting.
BiLSTM: bidirectional long short-term memory
BioBERT: bidirectional encoder representations from transformers for biomedical text mining
ELECTRA: Efficiently Learning an Encoder that Classifies Token Replacements Accurately
CDC: Centers for Disease Control and Prevention
CNN: convolutional neural network
CV: cross-validation
DT: decision tree
GloVe: Global Vectors for Word Representation
IDF: inverse document frequency
MNB: multinomial naive Bayes
NLP: natural language processing
TF: term frequency
TF-IDF: term frequency–inverse document frequency
XGB: extreme Gradient Boosting
XLM-RoBERTa: Cross-lingual Language Model–Robustly Optimized BERT Pretraining Approach

Edited by W Ahmed; submitted 19.May.2025; peer-reviewed by L Cao, J Chen, Y Liu; comments to author 22.Jul.2025; revised version received 19.Oct.2025; accepted 13.Jan.2026; published 19.Feb.2026.

Please cite as:
Ahmad M, Orji R, Amjad M, Siddique A, Kubysheva N, Batyrshin I, Sidorov G
Automated Risk Assessment of Opioid Use: Analysis Using Pre-Trained Transformers on Social Media Data
JMIR Infodemiology 2026;6:e77783
URL: https://infodemiology.jmir.org/2026/1/e77783
doi:10.2196/77783
PMID:

©Muhammad Ahmad, Rita Orji, Maaz Amjad, Abubakar Siddique, Nailya Kubysheva, Ildar Batyrshin, Grigori Sidorov. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 19.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Social Support and Opinion Leaders in a Tuberculosis-Related Online Community in China: Content and Network Analyses

Xiaojun Fan[1], MA; Jueman Zhang[2], PhD; Xiuli Wang[1], PhD

[1]School of New Media, Peking University, Beijing, China

[2]Harrington School of Communication and Media, University of Rhode Island, Kingston, RI, United States

Corresponding Author:
Xiuli Wang, PhD
School of New Media
Peking University
No. 5 Yiheyuan Rd.
Haidian District
Beijing, 100871
China
Phone: 86 10 62766689
Email: xiuli.wang@pku.edu.cn

Abstract

Background: Tuberculosis (TB) remains one of the world's deadliest infectious diseases. Yet, despite the growing role of online health communities (OHCs) as key sources of social support, research on TB-related online communities remains scarce. Network analysis has been increasingly used to study OHCs and identify opinion leaders (OLs), offering a valuable approach to advancing knowledge about TB-related online communities.

Objective: This study examined the types of social support and the influence of OLs in a prominent TB-related online forum in China, with a particular focus on its curated subforum that served as a centralized space for user interaction. The subforum consisted of posts recommended by the forum's administrator and the corresponding user replies they generated.

Methods: The data consisted of all 438 administrator-recommended posts and the 150,570 associated user replies over 18 years, from the forum's launch in 2004 to 2021. The study used content analysis to examine the types of social support present in administrator-recommended posts, which are commonly considered high-quality. It then applied social network analysis to these posts and their associated user replies to identify OLs by using a Borda ranking method based on centrality measures and user tenure. Finally, semantic network analysis was used to explore topic clusters within each OL's posts and their associated user replies.

Results: The content analysis showed a high prevalence of informational and emotional support in the administrator-recommended posts. Of the 438 posts, 296 (67.5%) contained social support, with 150 containing informational support and 136 containing emotional support. Social support varied by post theme and whether the intent was to provide or seek it. Among disease knowledge posts, 74 out of 75 provided informational support. Emotional support was most frequently provided in nontreatment sharing posts (28/113) and most frequently sought in treatment experience posts (47/129). The social network analysis identified 10 OLs. The first was a former patient with TB, and the second was a pulmonary TB doctor. Together, they contributed 30.4% (133/438) of all the posts. Across the semantic network analyses of each OL's posts and their associated user replies, informational support was more prominent than emotional support.

Conclusions: The findings suggest that the examined TB-related online forum served as an important source of social support for people affected by TB in China, fostering an environment for both informational and emotional support. OLs played an important role by contributing posts and establishing a central position through reply interactions with users.

(JMIR Infodemiology 2026;6:e79140) doi:10.2196/79140

XSL•FO
RenderX

KEYWORDS

tuberculosis; TB; social support; online health community; OHC; social network analysis; semantic network analysis; opinion leader

Introduction

Background

Tuberculosis (TB), caused by a bacterium called *Mycobacterium tuberculosis*, can be transmitted among people through the air and usually affects the lungs [1]. As of 2023, TB has likely reclaimed its position as the world's leading cause of death from a single infectious agent after 3 years of COVID-19 leading, according to the World Health Organization [2]. Between 2014 and 2021, a total of 6,587,439 TB cases were reported in mainland China, with an average annual incidence rate of 59.17 per 100,000 people [3]. Although the TB incidence rate declined from 67.05 per 100,000 in 2014 to 46.40 per 100,000 in 2021 [3], China remains among the countries with the highest TB incidence rates, reporting about 734,400 new cases in 2023, ranking third globally and accounting for 6.8% of new cases worldwide [4].

TB treatment typically lasts at least 6 months. Throughout the treatment process, patients not only have to endure the physical pain caused by the disease and the side effects of medications but also experience negative impacts on their mental health [5-7]. Social support is considered a crucial factor in promoting treatment adherence, therapy completion, and the ability to navigate other challenges among patients with TB [8-10].

With the development of the internet and increased accessibility [11], online health communities (OHCs) have emerged as a leading source of social support for individuals facing health-related challenges, particularly through online forums [12,13] and social media groups [14-16]. The establishment and maintenance of a successful OHC requires key catalysts [17], often in the form of opinion leaders (OLs), who are defined as individuals in interpersonal communication networks that regularly provide information and opinions to others and exert influence on them [18]. Social network analysis has been increasingly used to identify influential users such as OLs and to reveal the community structure of online forums and social media groups [12-14].

Previous research has explored OHCs for various diseases [12-16]. However, the literature on TB-related OHCs remains scarce, which suggests a critical gap in understanding the state of online social support for people affected by TB. This study aimed to bridge this gap by exploring the TB forum in Baidu Post Bar (also known as Baidu Tieba, with "Bar" or "ba" meaning "forum"), a prominent TB-related OHC in China. Specifically, the study investigated the subforum within the TB forum that features posts recommended by the forum's administrator, which are commonly regarded as high-quality, along with the corresponding user replies they generated. The data, consisting of the administrator-recommended posts and the associated user replies in the subforum, were extracted over a span of 18 years, from the TB forum's launch in 2004 to 2021 [19].

The study has 3 main objectives. First, it uses content analysis to examine the types of social support provided and sought in the examined subforum. Second, it applies social network analysis to identify OLs using a Borda ranking method based on centrality measures and user tenure. Third, it uses semantic network analysis to uncover the topics discussed in the OLs' posts and their associated user replies, offering insights into the types of social support they provided to users. To the best of our knowledge, this is the first study to apply social network analysis and semantic network analysis to a TB-related OHC. Understanding these aspects could enhance knowledge about TB-related online social support, inform the design of more effective digital health interventions, and help address gaps in TB education and support networks.

Literature Review

Social Support in OHCs

The concept of social support involves providing information that makes individuals feel cared for, loved, valued, and connected within a network of reciprocal obligations [17]. It is postulated that social support positively impacts both physical and mental well-being [20], including alleviating stress and enhancing self-efficacy [17].

TB treatment generally requires at least 6 months, during which patients endure not only the physical discomfort caused by the disease and medication side effects, but also mental health challenges [5-7]. A study in China revealed that 59% of patients with TB experienced moderate to severe psychological stress [6]. A study from South Africa indicated that 60% of patients with TB exhibited symptoms of depression [7]. Social support is considered an important factor in helping patients with TB adhere to treatment, complete therapy, and overcome challenges [8-10]. A study from China demonstrated that social support from family members, friends, and other organizations contributed to improved medication adherence and alleviated symptoms of depression and anxiety [21].

Previous research outlines 4 primary types of social support: informational, emotional, instrumental, and appraisal support [22]. Another, more detailed classification includes informational, emotional, esteem, network, and tangible support [23,24]. A study on cancer-related OHCs identified opinion and personal narrative support as additional types in extending the detailed classification [12]. Furthermore, previous studies have revealed that informational and emotional support are the 2 most prevalent types of social support in OHCs for different health issues [25-27].

A scoping review of 49 studies worldwide about social support for people with TB and their households classified social support programs into 3 categories: financial intervention, food support, and community participation [8]. Among these, community participation fosters a supportive network and environment through activities such as providing educational resources and implementing educational activities [8]. However, limited access

XSL•FO
RenderX

to these support programs remains a major challenge for many individuals [8].

As internet access continues to expand [11], OHCs have become a prominent source of social support for individuals facing health-related challenges, particularly through online forums [12,13] and social media platforms [14-16]. However, research on TB-related online social support remains limited, underscoring the need for further investigation into digital platforms to better understand the state of social support for patients with TB.

As a type of online community focused on health, OHCs share common characteristics such as social relationships among users, specific organizational structures and discussion formats, sharing of historical content, community rituals, and a shared online discussion space [28], all of which promote user identity, nurture long-term connections, and encourage sustained commitment to community goals [29].

Online communities enable users to initiate discussions and engage with others' posts, fostering opportunities for interaction that help establish social networks [12]. In this context, online social support primarily refers to the support obtained through these interactions, often manifested in OHCs through the responses users receive from others [12]. Prior research has summarized that patients often turn to OHCs for social support due to limited access to adequate support in traditional networks, the convenience of computer-mediated communication, the need to cope with health-related stigma, and the perception of support providers as credible and similar [30].

According to Granovetter's Weak Ties Theory [31], the interactive relationships within online communities are considered weak ties, which, unlike strong ties such as those with family and friends, often provide a wider range of information and have a lower likelihood of conflict [32]. Previous studies have revealed that OHCs facilitate patient self-management through the sharing of health information and experiences [29,33].

OLs in OHCs

Lazarsfeld et al [18] introduced the concept of OLs in the 2-step flow of communication theory, which suggests that media messages do not directly influence the public. Instead, information flows first from mass media to OLs—individuals who actively consume, interpret, and filter media content based on their knowledge and expertise. The OLs then disseminate and discuss the information within their social networks, shaping public opinions and behaviors [34,35]. Rogers' [36] Diffusion of Innovation Theory also underscores the crucial role of OLs in spreading new information and influencing others. These individuals are typically influential figures within social networks, from whom others often seek information and advice [37].

Likewise, online OLs have been characterized as individuals who hold central positions in diffusing information within online communities and influencing public opinion [12,37]. A study on cancer-related OHCs revealed that online OLs actively shaped the agenda within OHCs by creating topics, primarily focusing on four themes: disease history and treatment, personal health and life updates, advocacy, and emotional expression [12].

In addition, research on cancer-related OHCs showed that online OLs provide various forms of social support through their replies to others, typically offering a combination of opinion support, emotional support, and network support [12]. However, a study on mental illness found that online OLs who post stereotypical and stigmatizing remarks on social media can reinforce public prejudices [38].

The influence of OLs in public health is evident not only in their role in disseminating health information but also in facilitating health promotion programs, including serving as role models for behavior change [39].

Within online communities, OLs have significantly more connections and a greater ability to foster shared attitudes among community members [12,40]. With the widespread use of mobile communication, online OLs can deliver information and advice to members within OHCs more promptly and effectively than those within traditional offline social networks [41].

Regarding the identity of OLs in OHCs, one study on cancer-related OHCs found that individuals with higher levels of cancer knowledge and a more optimistic outlook on life challenges were more likely to become OLs in these communities [42]. Another study on cancer-related OHCs identified the majority of OLs as patients with cancer, while the remaining were family or friend caregivers [12].

Research Questions

This study examined the subforum within the TB forum in Baidu Post Bar that features administrator-recommended posts, along with the replies those posts received, collected over 18 years from the TB forum's launch in 2004 to 2021. To understand the types of social support in the administrator-recommended posts, we used a 2-step content analysis. First, we classified each of the posts into 1 theme, drawing on existing classifications from the literature on OHCs [12,13], as well as emerging ones observed in the posts. The 6 resulting themes were disease knowledge, treatment experience, nontreatment sharing, community activities, community announcements, and questions and answers. Then, we identified posts containing social support and categorized them into 1 of 4 types based on the literature: informational, emotional, instrumental, or network [12,22]. We further distinguished whether each post provided or sought support. We posed the following questions:

- Research question 1 (RQ1): What types of social support are provided and sought in the administrator-recommended posts?
- Research question 2 (RQ2): How do different types of social support vary by post theme?

In addition, we used social network analysis to identify OLs, assessed by centrality and user-tenure metrics. We also applied semantic network analysis to uncover the topics discussed in the OLs' posts and their associated user replies. We posed the following questions:

- Research question 3 (RQ3): Who are the opinion leaders in the subforum based on centrality and user-tenure metrics?

XSL•FO
RenderX

- Research question 4 (RQ4): What topics emerge from opinion leaders' posts and their associated user replies?

Methods

Data Source

We examined the TB forum hosted on Baidu Post Bar, one of China's most prominent online communities related to TB. Baidu Post Bar, also known as Baidu Tieba, was launched in 2003 and represents one of the earliest and largest interest-based online platforms in China. Organized around thematic "bars" (forums) dedicated to specific topics, Baidu Post Bar enables registered users to initiate discussion threads or participate in existing ones through replies. At its peak around 2015, Baidu Post Bar encompassed over 22 million forums and attracted approximately 300 million monthly active users. Although subsequent competition from emerging social media platforms has reduced its dominance [19,43], Baidu Post Bar remains a valuable case for studying Chinese internet culture and community dynamics, owing to its vast historical user base and distinctive user-generated content ecosystem. The TB forum, as one of the platform's earliest and most active disease-specific communities, offers a unique dataset for investigating the core mechanisms driving health-related online communities—namely, social support and opinion leadership.

We focused our analysis on the subforum of administrator-recommended posts within the TB forum. In Baidu Post Bar, an administrator holds the highest authority and is responsible for forum governance, policy enforcement, and user management. Each forum includes a prominent navigation tab leading to a centralized subforum for all administrator-recommended posts, which are widely perceived as high-quality content. This subforum fosters focused user engagement through replies, which are the sole interaction mechanism. In contrast, the broader forum contains numerous low-engagement threads, yielding fragmented reply networks unsuitable for robust social network analysis. By restricting the dataset to the subforum containing administrator-recommended posts and their associated user replies, we ensured a cohesive, high-interaction dataset that was optimal for examining social support and opinion leadership dynamics.

Data collection spanned from 2004 to December 31, 2021, encompassing the TB forum's entire lifecycle—from its inception in 2004 through Baidu Post Bar's peak popularity and its subsequent decline amid competition from newer social media. This longitudinal window captures the forum's rise, maturity, and waning activity, culminating in the near cessation of administrator-recommended posts by late 2021 (only 2 posts in 2021) and a registered user base of approximately 113,000. Using Scrapy [44], an open-source web scraping framework for Python (Guido van Rossum), we programmatically retrieved 438 original administrator-recommended posts and 150,570 associated user replies, with the earliest post dated May 27, 2004. Additional publicly available metadata included post and reply timestamps, anonymized user identities, gender, registration date, and lifetime post counts.

Content Analysis to Examine Themes and Social Support in the Posts

Content analysis has often been used to examine social support in OHCs [24,25,45]. The unit of analysis is an original post in the administrator-recommended subforum. Each post consists of a user identity, post date, post title, and post body. We used a 2-step content analysis to manually code each post's theme and social support based on its title and body. In step 1, the researchers classified each of the 438 posts into 1 of 6 themes, drawing on existing classifications from the literature on OHCs [12,13], as well as emerging ones observed in the posts. The first theme, disease knowledge, includes information about TB treatment, research progress, and advice on returning to work or school. The second theme, treatment experience, consists of patients sharing their experience with TB treatment. The third theme, nontreatment sharing, covers aspects of daily life or travel experience during or after TB treatment, as well as personal talents and skills. The fourth theme, community activities, contains information about group activities initiated within the OHC, including offline events. The fifth theme, community announcements, consists of updates shared by the management team of the TB OHC. The final theme, questions and answers, covers consultations on TB-related issues and responses provided by doctors or experienced patients.

In the second step, the researchers identified 296 of the 438 posts as containing social support. These 296 posts were then classified into 1 of 4 types of social support—informational, emotional, instrumental, and network support—based on prior literature [12,22]. In addition, each post was coded according to whether it provided or sought social support. Informational support includes advice, knowledge, and resources related to treatment options, medication use, hospital information, research findings, and other relevant topics to help individuals navigate TB-related challenges and make informed decisions. Emotional support involves expressing empathy, concern, compassion, and encouragement to help individuals cope with TB-related struggles and difficult emotions. Instrumental support refers to practical resources and assistance to help individuals address their needs. Network support focuses on fostering a sense of belonging within a community through actions such as offering greetings, acknowledging other group members, and engaging in mutual information exchange.

For example, a post titled "Is there anyone familiar with bronchial TB? What medications are you taking?" (including its body text) was coded under the disease knowledge theme as a request for informational support. Another post titled "I'm not afraid of anything, except for making my parents worry" (including its body text) was coded under the treatment experience theme as seeking emotional support. Two researchers independently coded 50 posts randomly selected from the 438 posts to test intercoder reliability. Krippendorff's α was 0.92 for classifying post themes, 0.89 for classifying the type of social support a post provides, and 0.90 for classifying the type of social support a post seeks. Each researcher then coded half of the remaining posts independently.

XSL•FO
RenderX

Social Network Analysis to Examine Community Structure and Identify OLs

This study used social network analysis to examine the community structure and to identify OLs within the subforum featuring administrator-recommended posts and their associated user replies. Social network analysis involves analyzing the structure of a social network, which consists of a set of actors connected by 1 or more relationships. In this context, nodes represent users, and edges represent their connections [46]. We modeled the subforum as a directed interaction network, where each node represents a user and a directed edge from user *i* to user *j* indicates that user *i* replied to user *j*, which is the only form of interaction available in the subforum.

We identified the top 10 OLs using a Borda ranking method based on 4 centrality metrics and a user-tenure metric [47,48]. User tenure was included as a separate indicator in the Borda aggregation to capture cumulative participation in the forum, rather than to normalize or adjust the structural centrality measures. The 4 centrality metrics are (1) normalized in-degree centrality (the proportion of distinct users who replied to a given user); (2) normalized out-degree centrality (the proportion of distinct users a given user replied to); (3) betweenness centrality (the extent to which a user lies on shortest reply paths between other users); and (4) PageRank (a weighted measure that reflects replies from other highly influential users) [40]. User tenure was operationalized as the time since registration. Each metric was first converted into a separate rank, and these ranks were then aggregated using a Borda count to produce an overall influence score as follows.



where $r_m(i)$ is the rank of user *i* on metric *m* among *n* users. The final Borda rank was derived from ordering the users by S_i.

Gephi 0.10.1 (Mathieu Bastian, Sébastien Heymann, and Mathieu Jacomy) was used to conduct the social network analysis [49], using the Yifan Hu proportional layout algorithm.

Semantic Network Analysis of OLs' Posts and All User Replies

This study used semantic network analysis to investigate the topics discussed in OLs' posts and user replies. Semantic network analysis is frequently used to analyze large volumes of text and identify the main topics. Unlike social network analysis, which treats actors as nodes and relationships as edges, semantic network analysis treats words as nodes and their co-occurrences as edges [46]. In a semantic network, words that convey the same topic tend to appear together in the same cluster. The higher the frequency with which a specific group of words appears in the semantic network, the more prominent the corresponding topic.

Since the corpus used in this study was in Chinese, we used ROST-CM6 (Wuhan University), a commonly used Chinese content mining and analysis software. Initially, all texts were tokenized, and high-frequency words were extracted based on a frequency ranking of the top 200 words. Subsequently, high-frequency words with similar meanings were merged. Word pairs exhibiting co-occurrence relationships within the set of high-frequency words were then used to construct a co-occurrence matrix. Finally, the semantic networks of high-frequency words were visualized using Gephi (Mathieu Bastian, Sébastien Heymann, and Mathieu Jacomy) [47], and the words were translated from Chinese into English for presentation.

Ethical Considerations

This study analyzes publicly available, anonymized data obtained from an online platform. No private, personally identifiable, or sensitive information was accessed or collected. All data were deidentified before analysis. The research did not involve direct interaction with human participants. In accordance with items (1) and (2) of Article 32 of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects, jointly issued by the National Health Commission, the Ministry of Education, the Ministry of Science and Technology, and the National Administration of Traditional Chinese Medicine of the People's Republic of China, the study was exempt from institutional review board review for the following reasons. The research is conducted using legally obtained public online data generated through observation without interfering with public behaviors; all data have been anonymized; the study causes no harm to the human body and does not involve sensitive personal information or commercial interests [50].

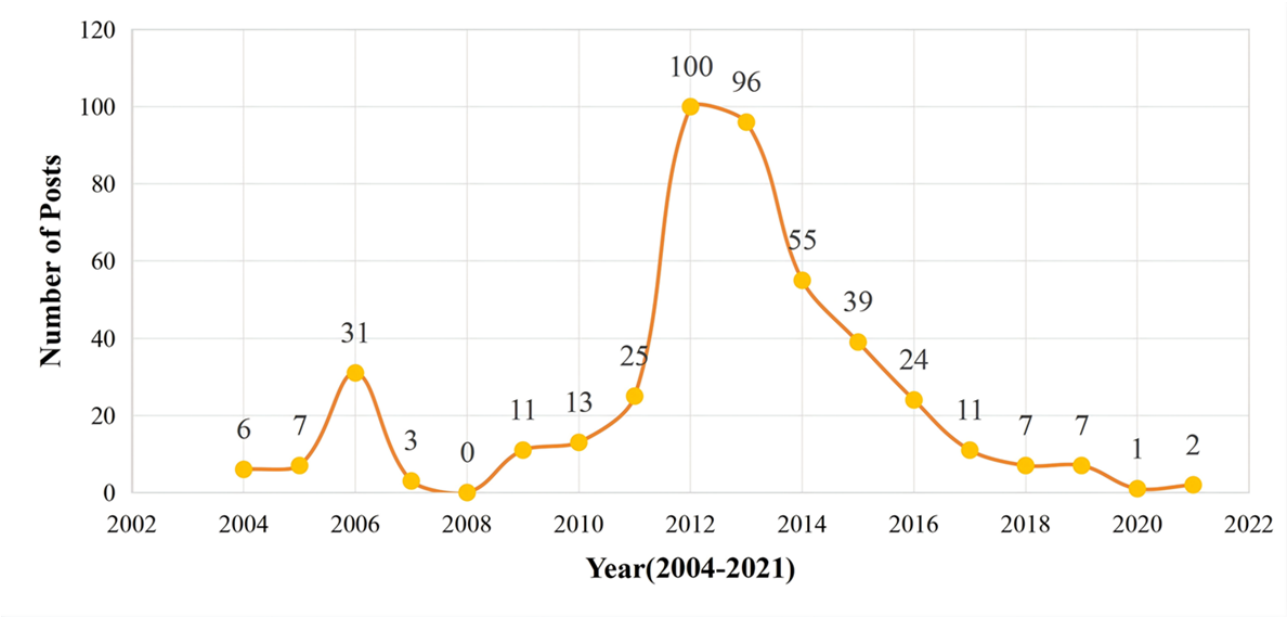
Results

Original Posts by Year

Figure 1 illustrates the annual number of original posts in the TB forum from its launch in 2004 through 2021. The peak occurred in 2012, followed closely by 2013. This surge can be attributed primarily to 2 factors. First, the 2012-2013 period coincided with the height of user activity on Baidu Post Bar, which drew a large number of patients with TB to the forum for interaction and support. Second, the production of high-quality posts depended heavily on active moderation and recommendation by the forum administrator. In 2012, a new administrator took over the role. Compared to the previous one, the new administrator exhibited greater engagement, stronger organizational skills, and more substantial influence within the community. He authored 24 posts in 2012 and 15 in 2013, while recommending 140 posts over those 2 years.

XSL•FO
RenderX

Figure 1. The number of administrator-recommended posts from 2004 to 2021.



Subsequently, the conditions supporting sustained content production eroded. Core users gradually left the forum upon recovery, and more significantly, the emergence of new social media platforms such as Weibo (known as China's Twitter; Sina Corporation) and WeChat (a super app developed by Tencent and widely used for messaging, social networking, mobile payments, and various online services) prompted a broader migration away from Baidu Post Bar. As a result, the volume of administrator-recommended posts declined steadily over the year. By 2020 and 2021, the number of administrator-recommended posts had fallen to just 1 and 2 posts, respectively.

Themes and Social Support in Administrator-Recommended Posts

Table 1 presents the classification of themes and social support in the 438 original posts. Among the 6 themes, the most prevalent was treatment experience, where patients shared their experience with TB treatment, appearing in 29.5% (129/438) of the posts. This was followed by nontreatment sharing at 25.8% (113/438), disease knowledge at 17.1% (75/438), community activities at 16% (10/438), community announcements at 7.5% (33/438), and questions and answers at 4.1% (18/438).

XSL•FO
RenderX

Table 1. Classification of themes and social support in administrator-recommended posts (N=438). One example post title (labeled a-i) is provided for each category containing ≥10 posts. Coding was based on both the post title and the body text.

Theme	Posts (N=438), n (%)	Social support (n=296), n (%)	Types of social support							
			Informational (n=150), n (%)		Emotional (n=136), n (%)		Instrumental (n=4), n (%)		Network (n=6), n (%)	
			Provide	Seek	Provide	Seek	Provide	Seek	Provide	Seek
Disease knowledge	75 (17.1)	75 (25.3)	74[a] (16.9)	0 (0)	1 (0.2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Treatment experience	129 (29.5)	115 (38.9)	25[b] (5.7)	12[c] (2.7)	30[d] (6.8)	47[e] (10.7)	0 (0)	1 (0.2)	0 (0)	0 (0)
Nontreatment sharing	113 (25.8)	41 (13.9)	3 (0.7)	1 (0.2)	28[f] (6.4)	8 (1.8)	0 (0)	0 (0)	1 (0.2)	0 (0)
Community activities	70 (16)	29 (9.8)	3 (0.7)	2 (0.5)	18[g] (4.1)	4 (0.9)	1 (0.2)	1 (0.2)	0 (0)	0 (0)
Community announcements	33 (7.5)	18 (6.1)	12[h] (2.7)	0 (0)	0 (0)	0 (0)	1 (0.2)	0 (0)	5 (1.1)	0 (0)
Questions and answers	18 (4.1)	18 (6.1)	12[i] (2.7)	6 (1.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Total	438 (100)	296 (67.5)	129 (29.5)	21 (4.8)	77 (17.6)	59 (13.5)	2 (0.5)	2 (0.5)	6 (1.4)	0 (0)

[a]Example post title: is there anyone familiar with bronchial tuberculosis? What medications are you taking?

[b]Example post title: process record: treatment of tuberculosis in Canada.

[c]Example post title: record my first day on medication.

[d]Example post title: to those who have recovered and finished medication, check in here and cheer on the newcomers.

[e]Example post title: I am not afraid of anything, except for making my parents worry.

[f]Example post title: [positive energy] teenager with pulmonary tuberculosis dances the shuffle and shows that patients with tuberculosis are no different from others!

[g]Example post title: dear friends, whether we know each other or not, let's comfort and encourage one another!

[h]Example post title: hello everyone! To help you better communicate, interact, and seek advice in this forum, please read this post carefully. It will answer common questions and explain the forum's rules. We hope this guide is helpful to you!

[i]Example post title: [back-to-school-season] questions and answers on updated conditions for suspension and resumption of studies—all students welcome!

RQ1 examined the types of social support provided and sought in the subforum featuring administrator-recommended posts. As Table 1 shows, about 67.5% (296/438) of the posts contained social support. Among the 4 types of social support, informational support was the most prevalent, appearing in 150 posts, with 86% (129/150) providing it and 14% (21/150) seeking it. This was closely followed by emotional support, present in 136 posts, with 56.6% (77/136) providing it and 43.4% (59/136) seeking it. Instrumental support and network support were minimal, found in 4 and 6 posts, respectively.

RQ2 explored how different types of social support varied across post themes. Analysis of the distribution of informational and emotional support across the 3 most common themes—treatment experience, nontreatment sharing, and disease knowledge—revealed distinct patterns. Among the 129 posts about treatment experience, 89.1% (115/129) contained social support. Seeking emotional support was the most prevalent, appearing in 40.9% (47/115) of the posts, followed by 26.1% (30/115) posts providing emotional support, 21.7% (25/115) providing informational support, and 10.4% (12/115) seeking informational support. In contrast, only 36.3% (41/113) of posts about nontreatment sharing included social support, with providing emotional support being the most dominant type (68.3%, 28/41 of the posts). For the 75 posts about disease knowledge, providing informational support was overwhelmingly dominant, appearing in 98.7% (74/75) of posts. Table 2 presents the top 10 most-replied-to posts that contain social support.

XSL•FO
RenderX

Table 2. Top 10 most-replied-to posts containing social support.

Reply rank	Reply count	Post translated from Mandarin[a] (date posted)	Theme coded in the content analysis	Social support type was coded in the content analysis
1	37,474	You're welcome to ask the doctor about TB[b] (Feb 6, 2006)[c].	Treatment experience	Informational (provide)
2	30,075	Record my first day on medication (Sept 17, 2011).	Treatment experience	Informational (seek)
3	7844	A hub for comforting communication (Oct 4, 2011).	Community activities	Emotional (seek)
4	7395	I'm not afraid of anything, except for making my parents worry (Aug 23, 2011).	Treatment experience	Emotional (seek)
5	7217	I'll live well (Nov 6, 2012).	Non-treatment sharing	Emotional (seek)
6	7057	July 2015: Girls who love to smile have good luck (Feb 24, 2016).	Non-treatment sharing	Emotional (provide)
7	6436	Anti-TB Diary (Sept 2, 2014).	Treatment experience	Informational (provide)
8	6332	Process record: treatment of TB in Canada (Aug 15, 2014).	Treatment experience	Informational (provide)
9	6217	Recording daily medication with photos and text—Stay persistent and keep going! (July 10, 2013)	Treatment experience	Emotional (provide)
10	5779	Process record: treatment of TB in Canada (Apr 5, 2015).	Treatment experience	Informational (provide)

[a]Each original post was manually coded for theme and social support based on its title and body. Due to space limitations, the table presents only the post titles rather than the full post content.

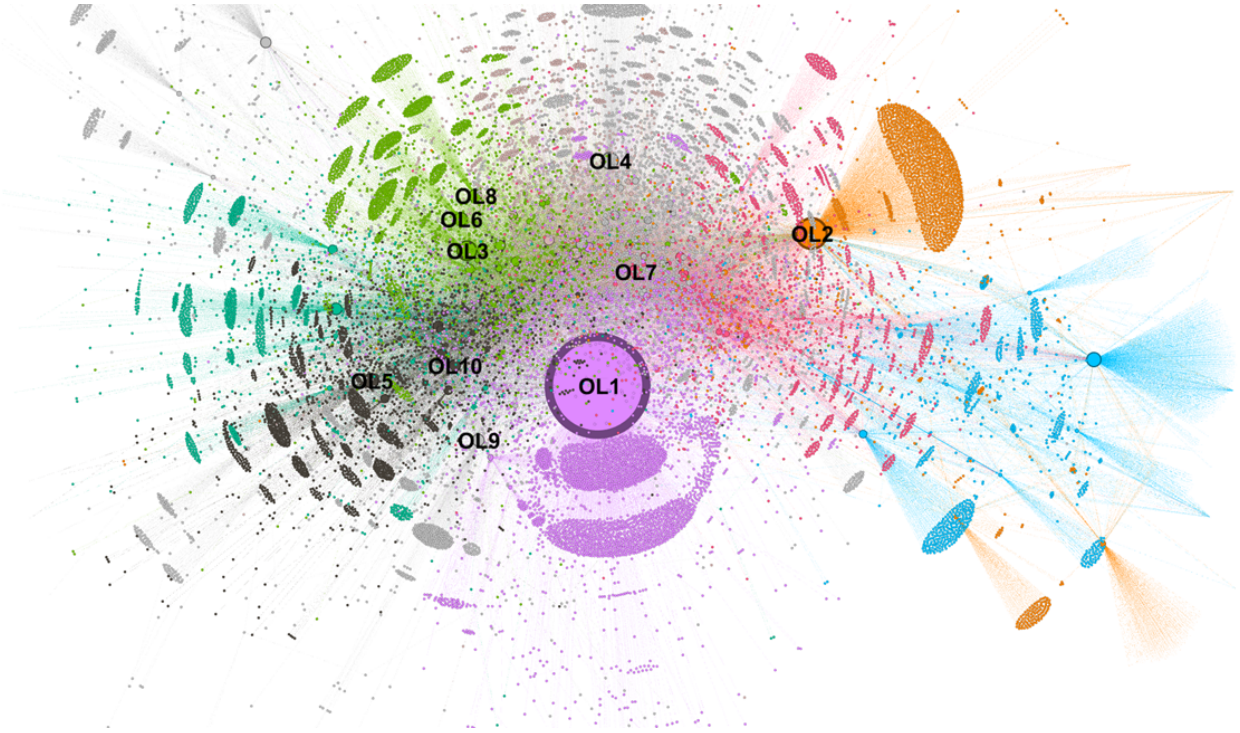
[b]TB: tuberculosis.

[c]The user who posted the post is a doctor specializing in pulmonary tuberculosis.

Identifying OLs

Figure 2 depicts the overall reply-based social network within the TB subforum of administrator-recommended posts, clustered using the modularity module via the Louvain algorithm and visualized with the Yifan Hu proportional layout algorithm in Gephi [47]. The network reveals a highly centralized structure dominated by the top 10 OLs (labeled OL1-OL10), each surrounded by dense clusters of varying colors representing distinct interaction communities. OL1 (purple) commands the largest and most cohesive cluster, indicating paramount influence and extensive reply engagement. Other leaders—such as OL2 (orange), OL3 (green), and OL7 (pink)—anchor substantial but smaller subnetworks, while peripheral clusters exhibit sparser connections. This configuration underscores the gatekeeping role of OLs, channeling the majority of discourse and support exchange, consistent with a hierarchical, star-like topology characteristic of health-related online communities where influential users mediate information flow and relational ties.

Figure 2. Opinion leaders (OLs) and subcommunities of the administrator-recommended posts and user replies.



XSL•FO
RenderX

Table 3 presents the top 10 OLs identified using the Borda ranking method, which aggregates scores across 4 centrality metrics and a user-tenure metric.

Table 3. Top 10 opinion leaders identified using the Borda ranking method.

User ID[a]	Rank[b]	Indegree centrality (normalized)	Outdegree centrality (normalized)	Betweenness centrality	PageRank	User tenure (years)
OL[c]1	1	0.2215	0.1262	0.2368	0.0599	14.4
OL2	2	0.0920	0.0103	0.0441	0.0206	16.9
OL3	3	0.0311	0.0301	0.0275	0.0072	6.7
OL4	4	0.0249	0.0195	0.0201	0.0073	5.8
OL5	5 (tie)	0.0473	0.0048	0.0136	0.0097	7.9
OL6	5 (tie)	0.0252	0.0137	0.0140	0.0061	9.3
OL7	7	0.0256	0.0126	0.0132	0.0066	9.3
OL8	8	0.0179	0.0168	0.0174	0.0061	8.4
OL9	9	0.0214	0.0180	0.0153	0.0050	7.2
OL10	10	0.0181	0.0164	0.0122	0.0042	7.8

[a]ID: identity.

[b]For the Borda ranking, each metric was converted into ranks separately. These ranks were then combined using a Borda count aggregation method, as explained in the Methods section.

[c]OL: opinion leader.

The top-ranked OL led all 4 centrality measures: receiving replies from the largest proportion of distinct users (highest normalized in-degree centrality), replying to the largest proportion of distinct users (highest normalized out-degree centrality), occupying the most structurally important positions in reply paths (highest betweenness centrality), and receiving replies from other highly influential users (highest PageRank). Although this user did not have the longest tenure in the forum—the second-ranked OL had been active for more years—the consistently superior centrality scores outweighed the tenure difference in the Borda aggregation, resulting in this user's overall first-place ranking.

The first-ranked OL served as the administrator of the TB forum at the time of data collection and is a former patient with TB who had completed treatment and fully recovered. The second-ranked OL, a medical doctor specializing in pulmonary TB, is the founder and initial administrator of the TB forum.

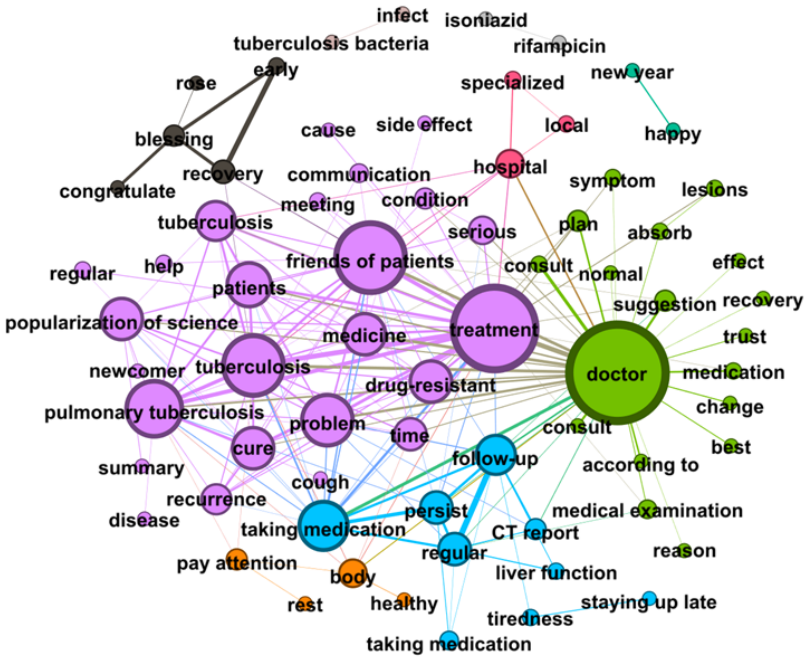
Out of the 438 administrators-recommended posts, the first OL contributed 71 posts, accounting for 16.2% of the total. The second OL contributed 62 posts, accounting for 14.2% of the total. Combined, the 2 OLs contributed 30.4% (133/438) of the posts.

Semantic Network Visualization of Topics in OLs' Posts and User Replies

RQ4 examined the topics that emerged from OLs' posts and the replies these posts received from users. Semantic network visualization, achieved using the modularity module via the Louvain algorithm, was used to uncover the topics emerging from each of the top 10 OLs' posts and their associated replies. Figures 3-12 present the resulting semantic networks, with Figure 3 corresponding to OL 1, Figure 4 to OL 2, and so forth to Figure 12. In these visualizations, each node represents a high-frequency word translated from Chinese, with node size proportional to its frequency of co-occurrence with other high-frequency words. Edge thickness signifies the strength of co-occurrence between word pairs. Colors delineate distinct topic clusters but may not reflect cluster magnitude in terms of the number of words or nodes within each cluster.

XSL•FO
RenderX

Figure 3. Semantic network of opinion leader 1 (OL1) posts and associated user replies.



As illustrated in Figure 3, the semantic network constructed from the first OL's posts and the associated user replies identifies seven distinct topic clusters, each denoted by a unique color and characterized by its high-frequency words. The largest cluster (purple, n=25 words) focuses on TB treatment information, including transmission routes, therapeutic options for contacts, pulmonary TB, drug resistance, cure rates, recurrence risks, and treatment duration. The second-largest cluster (green, n=18 words) revolves around the keyword "doctor," emphasizing the importance of professional medical consultation and treatment. The third cluster (blue, n=9 words) addresses treatment adherence and lifestyle guidelines, including regular check-ups, medication compliance, liver protection, and avoidance of overexertion. The fourth cluster (black, n=5 words) conveys blessings and wishes for speedy recovery. The fifth (orange, n=4 words) highlights rest and overall health maintenance. The sixth (gray, n=4 words) pertains to antibiotic regimens for TB. Finally, the smallest cluster (teal, n=2 words) reflects New Year greetings.

The first OL exemplifies both informational and emotional support in interactions with forum users. For instance, when a user inquired whether joint pain during medication treatment was normal, the first OL responded: "Joint pain is considered normal; many patients experience it after taking the medication. Check your liver and kidney function—if it's caused by high uric acid, your attending doctor can provide appropriate treatment." This exchange illustrates informational support through practical medical guidance and reassurance based on clinical knowledge.

Emotional support is equally evident. When responding to a distressed minor patient, the first OL wrote: "Tuberculosis is no longer a terminal illness. It's normal to feel shocked and panicked at first due to a lack of understanding. With proper treatment and adherence to medication, it can be cured. Many patients in this forum have already recovered, and you will too." Such replies offer empathy, normalization of fear, and hopeful encouragement, fostering resilience amid stigma and uncertainty.

Notably, the first OL's most engaged posts—generating 3693 replies—explicitly called on recovered patients with TB to share encouragement with newly diagnosed individuals, reinforcing the forum's role as a mutual-aid network driven by peer expertise and solidarity. These examples underscore how OLs blend authoritative knowledge with compassionate outreach, enhancing both cognitive understanding and affective coping in the community.

Figure 4 depicts the semantic network derived from the second OL's posts and the associated user replies. The visualization identifies 4 topic clusters, each delineated by a distinct color and characterized by its high-frequency words. The largest cluster (purple, n=25 words) centers on TB symptoms, diagnosis, and treatment, with particular emphasis on pulmonary TB. The second-largest cluster (green, n=17 words) revolves around the keyword "doctor," underscoring the importance of seeking advice and treatment from doctors. The third cluster (orange, n=15 words), anchored by the keyword "TB," highlights patient symptoms and clinical characteristics, featuring words such as "bronchus," "pathological changes," "lesions," and "lymph nodes." The smallest cluster (blue, n=2 words) comprises "medicine" and "response," reflecting discussions of pharmacological reactions or treatment responses.

XSL•FO
RenderX

Figure 4. Semantic networks of opinion leader 2 (OL2) posts and associated user replies.

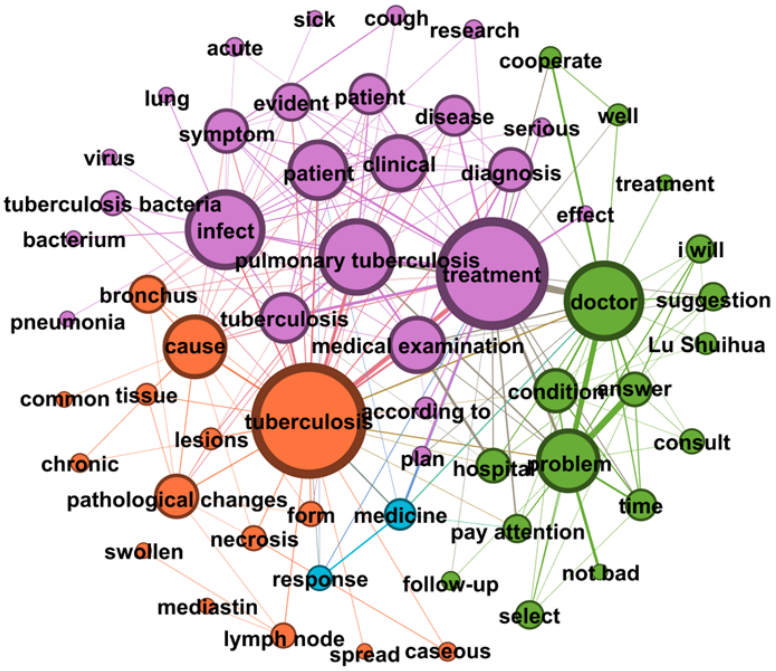
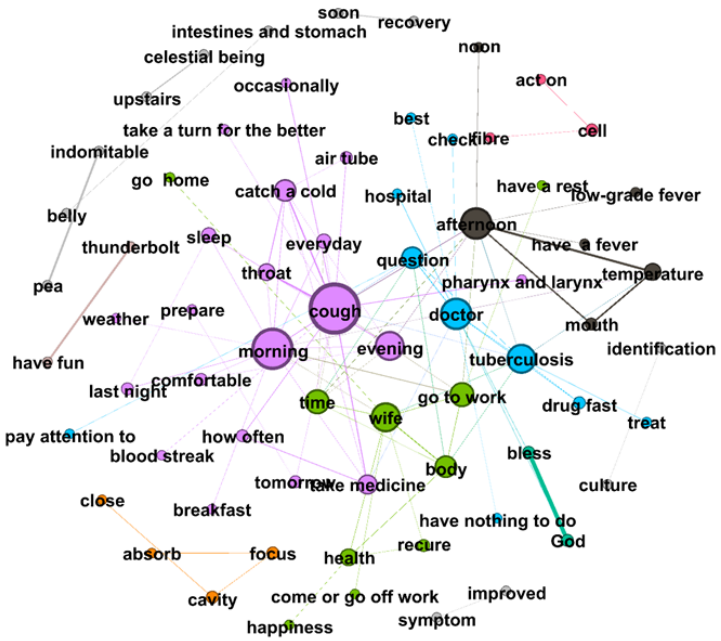


Figure 5. Semantic networks of opinion leader 3 (OL3) posts and associated user replies.



XSL•FO
RenderX

Figure 6. Semantic networks of opinion leader 4 (OL4) posts and associated user replies.

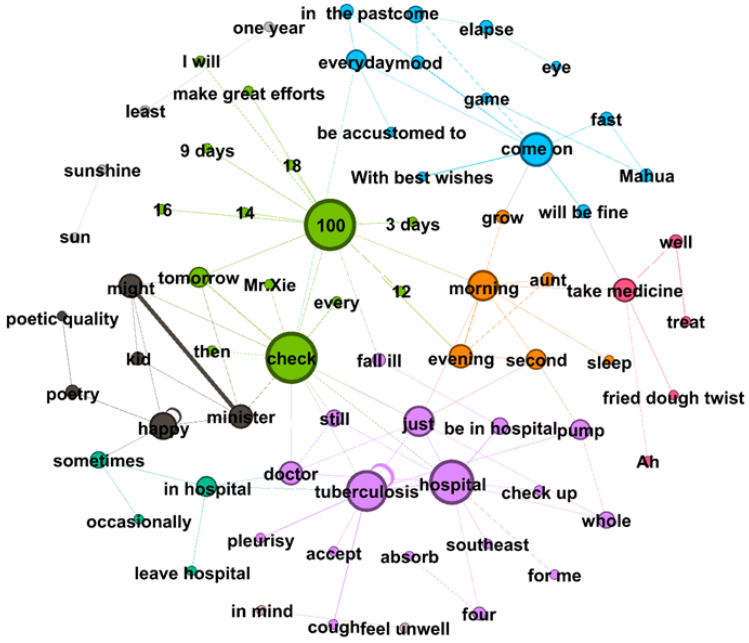
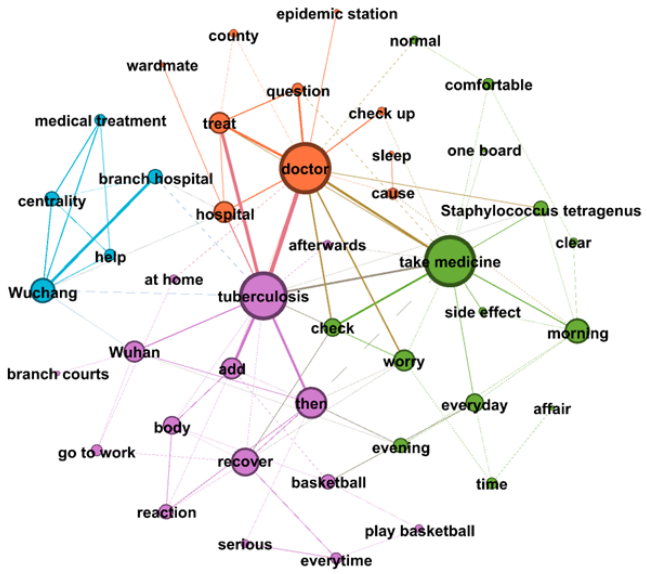


Figure 7. Semantic networks of opinion leader 5 (OL5) posts and associated user replies.



XSL•FO
RenderX

Figure 8. Semantic networks of opinion leader 6 (OL6) posts and associated user replies.

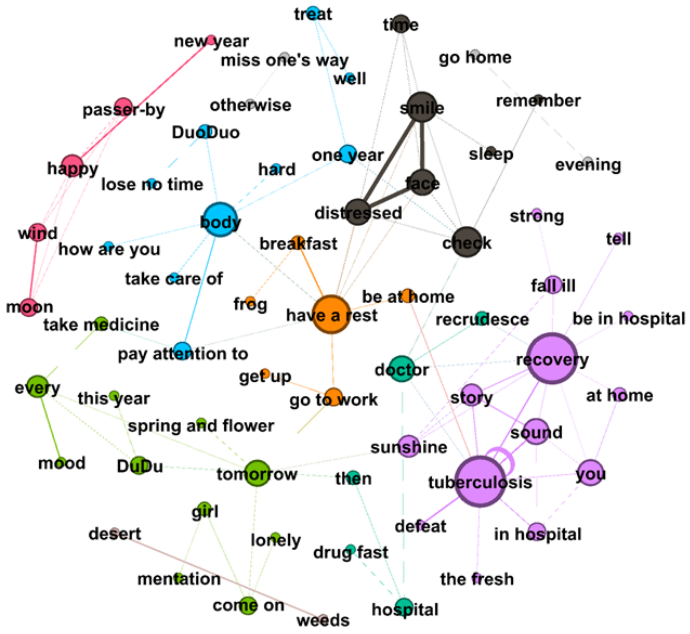
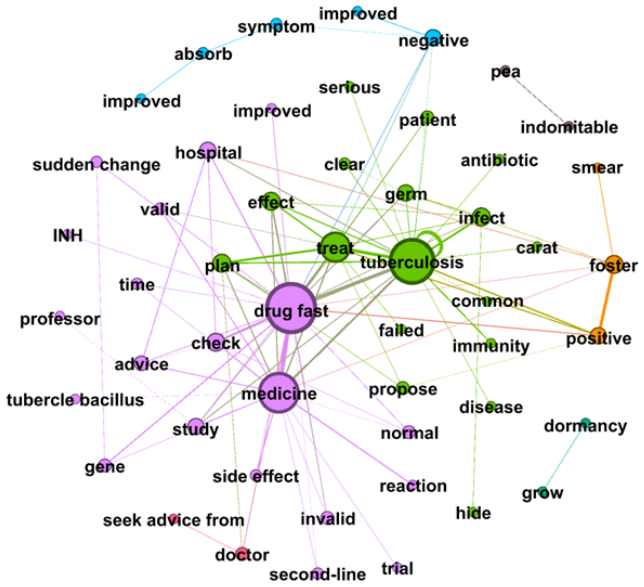


Figure 9. Semantic networks of opinion leader 7 (OL7) posts and associated user replies.



XSL•FO
RenderX

Figure 10. Semantic networks of opinion leader 8 (OL8) posts and associated user replies.

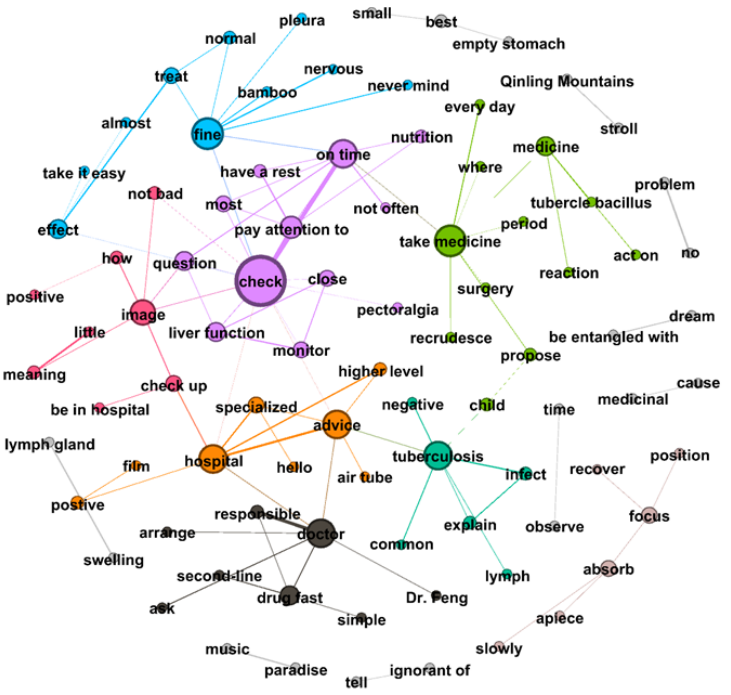
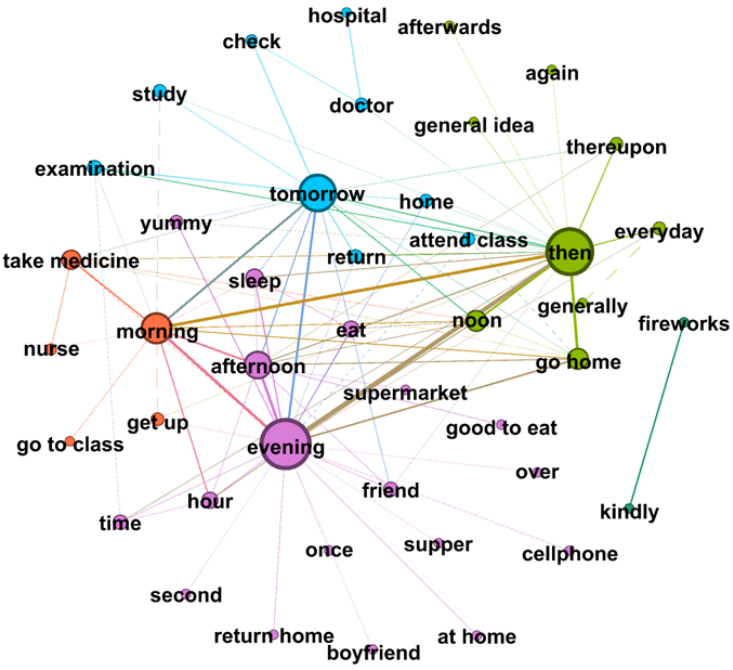


Figure 11. Semantic networks of opinion leader 9 (OL9) posts and associated user replies.

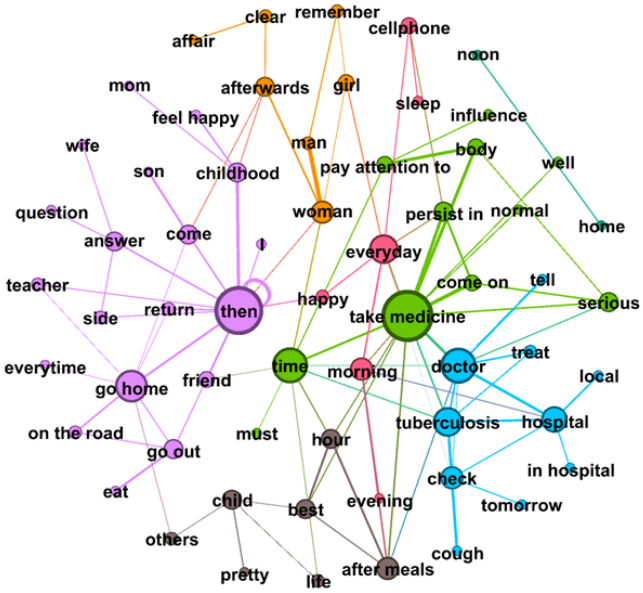


Notably, the second OL's 2006 post titled "Ask the Doctor" received an extraordinary 37,474 replies, making it the most-replied-to post in the subforum. For instance, 1 user asked whether rifampin and other anti-TB drugs could be taken after meals. The second OL responded: "Rifampin should be taken on an empty stomach, while other medications do not have such strict requirements. This is determined by the drug's mechanism of action." Such interactions highlight the second OL's pivotal role in the community's formative years, delivering professional consultations that bridged critical knowledge gaps for patients.

Across the semantic networks of the top 10 OLs' posts and their associated replies (Figures 3-12), informational support consistently predominates, manifesting as clusters centered on treatment protocols, diagnostic guidance, and clinical expertise. Emotional support, while present, remains secondary. This pattern aligns with the platform's asynchronous, text-based nature, which favors knowledge dissemination over affective exchange, and highlights the expert-driven character of influence in TB-related OHCs.

XSL•FO
RenderX

Figure 12. Semantic networks of opinion leader 10 (OL10) posts and associated user replies.



Discussion

Despite the growing importance of OHCs, empirical research on TB-related OHCs remains limited, leaving a significant gap in understanding the nature of social support available to affected individuals amid increasing internet access in China. This study helps bridge this gap by examining the TB forum on Baidu Post Bar, a prominent TB-related online community in China. Focusing on its subforum of administrator-recommended posts and associated replies, we applied content analysis, social network analysis, and semantic analysis to a longitudinal dataset spanning 18 years from the TB forum's inception in 2004 to 2021.

Principal Findings

The literature suggests OHCs as an important source of social support for health issues [12,13], which is confirmed by this study in the context of TB. In the leading TB online forum in China examined in this study, 67.5% (296/438) of the administrator-recommended posts contained social support. Consistent with previous studies on OHCs [25-27], informational and emotional support were the most common types. Moreover, there were more posts providing support than those seeking it, fostering a supportive online environment. Out of the 296 posts containing social support, informational support was the most prevalent, appearing in 150 posts—129 providing it and 21 seeking it. Emotional support was also significant, present in 136 posts, with 77 providing it and 59 seeking it. While access to traditional social support programs remains a major challenge for people affected by TB and their households [8], increased internet access has provided a channel for social support, which could be further leveraged for effective digital health interventions.

An examination of the 3 most common themes—treatment experience, nontreatment sharing, and disease knowledge—revealed distinct patterns of social support. Among the 129 posts on treatment experience, 115 contained social support, primarily seeking emotional support (47 posts). Of the 113 nontreatment sharing posts, 41 included social support, primarily providing emotional support (28 posts). For the 75 posts about disease knowledge, providing informational support was overwhelmingly dominant, appearing in 74 posts.

The social network analysis identified 10 leading OLs based on a Borda ranking method, which reflects overall influence derived from 4 centrality measures and user tenure as a temporal factor. While centrality measures capture users' positions and interaction patterns within the network, user tenure reflects the longevity of participation and is interpreted in this study as a complementary, rather than controlling, dimension of influence.

Specifically, the first-ranked OL not only achieved the highest combined in-degree and out-degree centrality but also topped each individually, indicating the highest levels of replies received and sent. The second-ranked OL placed second in combined centrality. Among the 438 posts, the first OL contributed 71 (16.2%) posts, and the second contributed 62 (14.2%) posts, accounting for a combined 30.4% (133/438) of all posts.

Consistent with previous research identifying OLs in OHCs primarily as patients [12], the top-ranked OL in this study was a former patient with TB. This individual's experiences of completing treatment and achieving full recovery likely provided substantial disease knowledge and a positive outlook, both of which have been shown to facilitate opinion leadership in OHCs [42]. The second-ranked OL was a pulmonary TB doctor with professional expertise. In addition, the first OL's role as the TB forum's administrator and the second OL's role as the forum's initial administrator highlight the significance of administrative positions in shaping influence through higher posting volume and denser reply-based interactions.

XSL•FO
RenderX

Importantly, tenure was not used to normalize or adjust the centrality measures. Rather, it was included as a distinct indicator in the Borda aggregation to reflect cumulative presence in the community. Accordingly, the Borda ranking should be interpreted as capturing overall influence, which may derive from both structural embeddedness and long-term participation. Structural centrality and tenure are therefore analytically distinguished in this study, and high tenure alone does not imply opinion leadership in the absence of strong network centrality.

Finally, the semantic network analysis uncovered the topic clusters in each OL's posts and their associated user replies. Across these semantic networks, informational support was more dominant than emotional support. Between the 2 highest-ranked OLs who served as administrators of the TB forum, their semantic networks shared common themes, including treatment information for TB and pulmonary TB, as well as the importance of consulting doctors. However, their themes diverged in scope. As a former patient with TB, the first OL addressed a broader range of topics and provided more emotional support, as reflected by 1 theme conveying well wishes for recovery and another theme containing holiday greetings. The first OL's most-replied-to post fostered social support within the forum by urging patients with TB who had completed medication treatment to respond and encourage newly diagnosed patients with TB. As a pulmonary TB doctor, the second OL provided more informational support, drawing on professional expertise. One theme focused on clinical symptoms and pathology, using more medically specific language. In the second OL's most-replied-to post, titled "Ask the Doctor," professional information and advice were shared.

These findings demonstrate that OLs predominantly provide informational and emotional support while anchoring the dominant topical clusters. This pattern offers empirical validation for an adapted 2-step flow model within OHCs, wherein influential users act as gatekeepers, selectively filtering and amplifying health information. The triangulation of methods further reveals interdependent mechanisms: structural positions (captured through social network analysis) systematically shape discursive content (manifest in semantic clusters) and support provision (coded via content analysis). Collectively, these insights yield a multilevel framework of OHC functioning that transcends descriptive enumeration, theorizing how relational architecture mediates both knowledge dissemination and affective sustenance in stigmatized illness contexts.

Limitations

This study has several limitations that warrant consideration. First, our analysis was confined to the subforum that features administrator-recommended posts, which are widely regarded by forum users as high-quality content and generate dense interaction patterns. While this focus ensured a cohesive and analytically tractable dataset suitable for network analysis, it excluded low-engagement threads from the broader forum. Importantly, restricting the network to threads curated by administrators may introduce a circular validation effect, whereby administrators are more likely to appear structurally central due to their role in initiating or curating these discussions.

We acknowledge that a sensitivity analysis—such as reconstructing the network after excluding administrator-authored initial posts to examine whether administrators remain central based solely on reply-based interactions—would provide a more rigorous test of this potential bias. However, such analyses were not conducted in this study. Consequently, the identified opinion leadership, particularly among administrators, should be interpreted as conditional on this curated subforum context rather than as a definitive representation of influence across the entire TB forum. Future research should address this limitation by incorporating the full forum dataset and conducting sensitivity analyses to assess the robustness of centrality findings under alternative sampling and network construction strategies.

Second, the cross-sectional aggregation of data across 2004-2021 obscures temporal evolution. Longitudinal disaggregation—for example, annual or periodic analysis of support types, OL prominence, and topical shifts—could illuminate how social support and leadership structures adapted to changing platform vitality and external health events.

Finally, Baidu Post Bar's declining usage in recent years amid competition from newer platforms limits the generalizability of findings to the contemporary Chinese OHC landscapes. Comparative studies of TB-related online communities on emerging social media platforms would enrich the literature and reveal platform-specific mechanisms of support provision and influence diffusion.

Conclusions

This study demonstrates that the TB forum on Baidu Post Bar served as an important source of social support for people affected by TB in China, fostering an environment rich in both informational and emotional resources. Of the 438 analyzed posts, 67.5% (296/438) contained social support, with informational support appearing in 150 posts and emotional support in 136. Thematic distribution revealed distinct patterns: disease knowledge posts overwhelmingly provided informational support (74/75), while treatment experience posts most frequently sought emotional support (47/129), and nontreatment sharing posts most commonly offered emotional support (28/113).

Using a Borda ranking method that integrates 4 centrality measures and user tenure, we identified 10 OLs. Semantic network analysis of each OL's posts and their associated user replies revealed that informational support predominated over emotional support. The 2 highest-ranked OLs—a former patient with TB and a pulmonary TB doctor—exerted the strongest influence through reply interactions. While both emphasized TB treatment and the importance of medical consultation, the patient OL covered a wider range of topics and provided more emotional support, whereas the doctor OL focused predominantly on authoritative informational content.

These findings suggest that OLs play a central role in filtering and amplifying health information, consistent with an adapted 2-step flow model in OHCs. The integration of content analysis, social network analysis, and semantic network analysis not only maps the structure of support provision but also reveals how

XSL•FO
RenderX

structural positions (centrality), discursive content (topical clusters), and support types interact to sustain community functioning.

Funding

No external financial support or grants were received for this work.

Conflicts of Interest

None declared.

References

1. Zaman K. Tuberculosis: a global health problem. J Health Popul Nutr 2010;28(2):111-113 [FREE Full text] [doi: 10.3329/jhpn.v28i2.4879] [Medline: 20411672]
2. Tuberculosis. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/tuberculosis [accessed 2025-03-01]
3. Deng LL, Zhao F, Li Z, Zhang WW, He GX, Ren X. Epidemiological characteristics of tuberculosis incidence and its macro-influence factors in Chinese mainland during 2014-2021. Infect Dis Poverty 2024;13(1):34 [FREE Full text] [doi: 10.1186/s40249-024-01203-6] [Medline: 38773558]
4. Global tuberculosis report 2024. World Health Organization. URL: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024 [accessed 2025-10-30]
5. Thomas BE, Shanmugam P, Malaisamy M, Ovung S, Suresh C, Subbaraman R, et al. Psycho-socio-economic issues challenging multidrug resistant tuberculosis patients: a systematic review. PLoS One 2016;11(1):e0147397 [FREE Full text] [doi: 10.1371/journal.pone.0147397] [Medline: 26807933]
6. Feng D, Xu L. The relationship between perceived discrimination and psychological distress among Chinese pulmonary tuberculosis patients: the moderating role of self-esteem. Psychol Health Med 2015;20(2):177-185. [doi: 10.1080/13548506.2014.958505] [Medline: 25243362]
7. Naidoo P, Mwaba K. Helplessness, depression, and social support among people being treated for tuberculosis in South Africa. Soc Behav Pers 2010;38(10):1323-1333 [FREE Full text] [doi: 10.2224/sbp.2010.38.10.1323]
8. Todd H, Hudson M, Grolmusova N, Kazibwe J, Pearman J, Skender K, et al. Social protection interventions for TB-affected households: a scoping review. Am J Trop Med Hyg 2023;108(4):650-659 [FREE Full text] [doi: 10.4269/ajtmh.22-0470] [Medline: 36806490]
9. Paz-Soldán VA, Alban RE, Jones CD, Oberhelman RA. The provision of and need for social support among adult and pediatric patients with tuberculosis in Lima, Peru: a qualitative study. BMC Health Serv Res 2013;13:290 [FREE Full text] [doi: 10.1186/1472-6963-13-290] [Medline: 23899353]
10. Nirmal A, Kuzmik A, Sznajder K, Lengerich E, Fredrick N, Chen M, et al. 'If not for this support, I would have left the treatment!': qualitative study exploring the role of social support on medication adherence among pulmonary tuberculosis patients in Western India. Glob Public Health 2022;17(9):1945-1957 [FREE Full text] [doi: 10.1080/17441692.2021.1965182] [Medline: 34459366]
11. LaCoursiere SP. A theory of online social support. ANS Adv Nurs Sci 2001;24(1):60-77. [doi: 10.1097/00012272-200109000-00008] [Medline: 11554534]
12. Zhu Y, Guan M, Donovan E. Elaborating cancer opinion leaders' communication behaviors within online health communities: network and content analyses. Social Media + Society 2020;6(2) [FREE Full text] [doi: 10.1177/2056305120909473]
13. Lu Y, Luo S, Liu X. Development of social support networks by patients with depression through online health communities: social network analysis. JMIR Med Inform 2021;9(1):e24618 [FREE Full text] [doi: 10.2196/24618] [Medline: 33279878]
14. Ramamoorthy T, Karmegam D, Mappillairaju B. Use of social media data for disease based social network analysis and network modeling: a systematic review. Inform Health Soc Care 2021;46(4):443-454. [doi: 10.1080/17538157.2021.1905642] [Medline: 33877944]
15. Cheng Z, Hao H, Tsofliou F, Katz M, Zhang Y. Effects of online support and social media communities on gestational diabetes: a systematic review. Int J Med Inform 2023;180:105263. [doi: 10.1016/j.ijmedinf.2023.105263] [Medline: 37907014]
16. Bender JL, Jimenez-Marroquin M, Jadad AR. Seeking support on Facebook: a content analysis of breast cancer groups. J Med Internet Res 2011;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
17. Cobb S. Social support as a moderator of life stress. Psychosom Med 1976;38(5):300-314 [FREE Full text] [doi: 10.1097/00006842-197609000-00003] [Medline: 981490]
18. Lazarsfeld PF, Berelson B, Gaudet H. The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign. New York, NY: Columbia University Press; 1944.
19. Baidu Tieba users lost 90%: former king, unable to return to its peak. Jiemian News. URL: https://m.jiemian.com/article/6796047.html [accessed 2023-01-30]

XSL•FO
RenderX

20. Thoits PA. Mechanisms linking social ties and support to physical and mental health. J Health Soc Behav 2011;52(2):145-161. [doi: 10.1177/0022146510395592] [Medline: 21673143]
21. Gu L, Wu S, Zhao S, Zhou H, Zhang S, Gao M, et al. Association of social support and medication adherence in Chinese patients with type 2 diabetes mellitus. Int J Environ Res Public Health 2017;14(12):1522 [FREE Full text] [doi: 10.3390/ijerph14121522] [Medline: 29211039]
22. House JS. Work Stress and Social Support. Boston, MA: Addison-Wesley Publishing Company; 1981.
23. Cutrona CE, Suhr JA. Controllability of stressful events and satisfaction with spouse support behaviors. Communication Research 1992;19(2):154-174 [FREE Full text] [doi: 10.1177/009365092019002002]
24. Liu Y, Zhu Y, Xia Y. Support-seeking strategies and social support provided in Chinese online health communities related to COVID-19. Front Psychol 2021;12:783135 [FREE Full text] [doi: 10.3389/fpsyg.2021.783135] [Medline: 34867695]
25. van Gastel D, Antheunis M, Tenfelde K, van de Graaf DL, Geerts M, Nieboer T, et al. Social support among women with potential essure-related complaints: analysis of facebook group content. JMIR Form Res 2023;7:e32592 [FREE Full text] [doi: 10.2196/32592] [Medline: 37535412]
26. Yip J. Evaluating the communication of online social support: a mixed-methods analysis of structure and content. Health Commun 2020;35(10):1210-1218. [doi: 10.1080/10410236.2019.1623643] [Medline: 31154856]
27. Maestre JF, Herring SC, Min A, Connelly CL, Shih PC. Where and how to look for help matters: analysis of support exchange in online health communities for people living with HIV. Information 2018;9(10):259 [FREE Full text] [doi: 10.3390/info9100259]
28. Stanoevska-Slabeva K, Schmid BF. A typology of online communities and community supporting platforms. 2001 Presented at: Proceedings of the 34th Annual Hawaii International Conference on System Sciences; January 06-06, 2001; Maui, HI URL: https://doi.org/10.1109/HICSS.2001.927041 [doi: 10.1109/hicss.2001.927041]
29. Chen L, Straub D. The impact of virtually crowdsourced social support on individual health: analyzing big datasets for underlying causalities. 2015 Presented at: Americas Conference on Information Systems (AMCIS 2015); August 13-15, 2015; Fajardo, Puerto Rico URL: https://aisel.aisnet.org/amcis2015/HealthIS/GeneralPresentations/18
30. Wright KB. Communication in health-related online social support groups/communities: a review of research on predictors of participation, applications of social support theory, and health outcomes. Rev Commun Res 2016;4:65-87 [FREE Full text] [doi: 10.12840/issn.2255-4165.2016.04.01.010]
31. Granovetter MS. The strength of weak ties: a network theory revisited. In: Marsden PV, Lin N, editors. Social Structure and Network Analysis. Thousands Oak: Sage Publication; 1982:105-130.
32. Wright KB, Miller CH. A measure of weak-tie/strong-tie support network preference. Communication Monographs 2010;77(4):500-517 [FREE Full text] [doi: 10.1080/03637751.2010.502538]
33. Willis E. The making of expert patients: the role of online health communities in arthritis self-management. J Health Psychol 2014;19(12):1613-1625. [doi: 10.1177/1359105313496446] [Medline: 23988679]
34. Kang M, Liang T, Sun B, Mao HY. Detection of opinion leaders: static vs. dynamic evaluation in online learning communities. Heliyon 2023;9(4):e14844 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e14844] [Medline: 37089294]
35. Katz E, Lazersfeld PF. Personal Influence: The Part Played by People in the Flow of Mass Communications. New York, NY: Free Press; 1955.
36. Rogers EM. Diffusion of Innovations. New York, NY: Free Press; 1955.
37. Huffaker D. Dimensions of leadership and social influence in online communities. Hum Commun Res 2010;36(4):593-617 [FREE Full text] [doi: 10.1111/j.1468-2958.2010.01390.x]
38. Wang W, Liu Y. Discussing mental illness in Chinese social media: the impact of influential sources on stigmatization and support among their followers. Health Commun 2016;31(3):355-363. [doi: 10.1080/10410236.2014.957376] [Medline: 26360792]
39. Valente P, Pumpuang P. Identifying opinion leaders to promote behavior change. Health Educ Behav 2007;34(6):881-896. [doi: 10.1177/1090198106297855] [Medline: 17602096]
40. Knoke D, Yang S. Social Network Analysis. Thousand Oaks, CA: SAGE Publications Ltd; 2019.
41. Turcotte J, York C, Irving J, Scholl R, Pingree R. News recommendations from social media opinion leaders: effects on media trust and information seeking. J Comput-Mediat Comm 2015;20(5):520-535 [FREE Full text] [doi: 10.1111/jcc4.12127]
42. Kim E, Scheufele DA, Han JY, Shah D. Opinion leaders in online cancer support groups: an investigation of their antecedents and consequences. Health Commun 2017;32(2):142-151 [FREE Full text] [doi: 10.1080/10410236.2015.1110005] [Medline: 27192376]
43. Tieba. Baidu. URL: https://baike.baidu.com/item/%E8%B4%B4%E5%90%A7/122101 [accessed 2025-03-01]
44. Scrapy. GitHub. URL: https://github.com/scrapy/scrapy [accessed 2026-03-05]
45. Qian Y, Gui W, Ma F, Dong Q. Exploring features of social support in a Chinese online smoking cessation community: a multidimensional content analysis of user interaction data. Health Informatics J 2021;27(2):14604582211021472 [FREE Full text] [doi: 10.1177/14604582211021472] [Medline: 34082598]
46. Wellman B, Berkowitz SD. Social Structures: A Network Approach. Cambridge, United Kingdom: Cambridge University Press; 1988.

XSL•FO
RenderX

47. Madotto A, Liu J. Super-spreader identification using meta-centrality. Sci Rep 2016;6:38994 [FREE Full text] [doi: 10.1038/srep38994] [Medline: 28008949]
48. Pethes R, Bodor-Eranus E, Takács K, Kovács L. The core might change anyhow we define it: the instability of key actors in longitudinal social network data. Complexity 2024;2024:1-22 [FREE Full text] [doi: 10.1155/2024/3956877]
49. Gephi. URL: https://gephi.org/ [accessed 2024-03-05]
50. Measures for ethical review of life science and medical research involving human subjects. National Health Commission, Ministry of Education, Ministry of Science and Technology, State Administration of Traditional Chinese Medicine. URL: https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm [accessed 2026-03-25]

Abbreviations

OHC: online health community
OL: opinion leader
TB: tuberculosis

Edited by M Haupt; submitted 16.Jun.2025; peer-reviewed by K Han, A Staffini; comments to author 03.Oct.2025; revised version received 24.Feb.2026; accepted 24.Feb.2026; published 31.Mar.2026.

Please cite as:
Fan X, Zhang J, Wang X
Understanding Social Support and Opinion Leaders in a Tuberculosis-Related Online Community in China: Content and Network Analyses
JMIR Infodemiology 2026;6:e79140
URL: https://infodemiology.jmir.org/2026/1/e79140
doi:10.2196/79140
PMID:41915425

©Xiaojun Fan, Jueman Zhang, Xiuli Wang. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 31.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating Public Sentiment on Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder Compared With Other Mental Health Disorders From Posts on X (Formerly Known as Twitter): Longitudinal Analysis

Vitor Tiosso Batistetti[1*], MD; Facundo G Sanchez[2*], MD; Andrea Varaona[3], MS; Francisco Lara-Abelenda[4], PhD; Mariana Pinto da Costa[5,6], MSc, MD, PhD; Juan Pablo Chart-Pascual[6,7,8,9], MD; Alberto Rodriguez-Quiroga[10,11], MD, PhD; Javier Quintero[10,11*], MD, PhD; Miguel Angel Alvarez-Mon[10,11,12,13*], MD, PhD

[1]Universidad Abierta Interamericana, Buenos Aires, Argentina

[2]Department of Psychiatry, Lincoln Medical Center, New York, NY, United States

[3]Department of Medicine and Medical Specialties, Faculty of Medicine and Health Sciences, University of Alcala, Alcalá de Henares, Spain

[4]Department of Signal Theory and Communications, Rey Juan Carlos University, Fuenlabrada, Spain

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

[6]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[7]BioAraba Research Institute, Vitoria, Spain

[8]Department of Psychiatry, Hospital Universitario de Álava, Vitoria, Spain

[9]University of the Basque Country, Vitoria, Spain

[10]Department of Psychiatry and Mental Health, Hospital Universitario Infanta Leonor, Madrid, Spain

[11]Department of Legal and Psychiatry, Universidad Complutense de Madrid, Madrid, Spain

[12]CIBERSAM-ISCIII, Biomedical Research Networking Centre in Mental Health, Instituto de Salud Carlos III, Madrid, Spain

[13]Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain

[*]these authors contributed equally

Corresponding Author:
Facundo G Sanchez, MD
Department of Psychiatry
Lincoln Medical Center
234 E 149th St
7B-121
New York, NY, 10451
United States
Phone: 1 9712553638
Email: facufgs@gmail.com

Abstract

Background: Neurodevelopmental disorders, especially attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD), have seen a marked rise in public attention, yet research on public opinion remains limited. Social media analysis offers real-time, unfiltered insights into public perceptions, enabling empirical examination of public attitudes and opinions.

Objective: This study aimed to assess the evolution of public opinion on ADHD and ASD between 2009 and 2023 by analyzing posts from X (formerly known as Twitter; X Corp), comparing perceptions across English and Spanish languages and against other mental health conditions.

Methods: Posts mentioning keywords related to ADHD and ASD and control conditions (eg, depression, anxiety, insomnia, bipolar disorder, schizophrenia, suicide, and substance use disorders) were collected from X between 2009 and 2023. The dataset included posts in both English and Spanish. Machine learning algorithms were then applied to classify post content into predefined categories, including volume of posts, engagement, personal experiences, trivialization, perceived causes, and perceived treatability. Parametric and nonparametric tests were used to assess for differences by language. Descriptive statistics were presented using tables and graphical representations.

XSL•FO
RenderX

Results: A total of 852,990 posts were analyzed, including 511,510 (59.97%) in English and 341,480 (40.03%) in Spanish. Overall, post volume on mental health conditions increased across the study period. In English, posts about ADHD (97,084/511,510, 18.98%) and ASD (74,619/511,510, 14.59%) were among the most frequent, while of the 341,480 Spanish posts, there were 49,475 (14.49%) ASD posts, significantly outnumbering ADHD posts (n=18,223, 5.34%; chi-square test $P<.001$). Engagement analysis indicated a notable increase in likes and reposts per post over time, particularly after 2019, with ADHD-related posts in English experiencing peak engagement during the COVID-19 pandemic. However, ASD posts had comparatively lower engagement across languages. Posts sharing personal experiences were more polarized in Spanish, with higher proportions of negative and positive experiences compared with English posts. Trivialization of mental illnesses was less common in Spanish posts than in English posts, particularly for ADHD (17,053/18,223, 93.59%; chi-square test $P<.001$) and ASD (41,933/49,475, 84.73%; chi-square test $P<.001$). User-perceived causes included multifactorial factors, biological or genetic factors, substance use, psychological susceptibility, acute psychosocial stressors, and COVID-19. Perceived treatability varied by language but consistently included high perceived incurability, limited improvement despite professional help, and low perceived self-manageability except for anxiety.

Conclusions: Analysis of social media discourse showed that ADHD attracted higher post volumes, particularly during the COVID-19 pandemic, often described with multifactorial causes including substance use and genetics. ASD consistently received lower engagement. Both language groups showed low trivialization, awareness of the chronicity of the illness, and limited support for the self-management of mental health conditions. These findings underscore social media's value for capturing direct public perceptions to guide future educational and intervention efforts.

(JMIR Infodemiology 2026;6:e74440) doi:10.2196/74440

KEYWORDS

neurodevelopmental disorders; social media; public opinion; mental health; trivialization; Twitter, X; cross-cultural comparison; English and Spanish analysis

Introduction

Neurodevelopmental disorders, such as attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD), are complex conditions affecting communication, attention, and behavior [1]. Public interest in these disorders has surged in recent years, leading to more frequent diagnoses and widespread discussion about their nature [2]. Despite their increased visibility, public understanding remains limited and misconceptions persist [3-7]. Traditional survey methods, including structured questionnaires and qualitative interviews, while valuable, often fail to capture these nuanced perceptions [8,9]. ADHD is common with pooled global estimates of approximately 5% in children and recent US estimates of approximately 1 in 10 among school-aged children. ASD is also common, with global estimates of approximately 1% and recent US surveillance estimates of approximately 1 in 36 among children [10,11]. The existing literature suggests that the rise in ADHD and ASD diagnoses may be attributed to changes in diagnostic criteria and greater clinical recognition [2]. However, this may not translate into a broader public understanding of these conditions [3,4,7,12]. Traditional research methods, such as surveys, interviews, and focus groups, may not capture the full spectrum of public beliefs and discourse [8,9]. These limitations of survey-based approaches informed our research questions about how ADHD and ASD are discussed in more naturalistic, publicly expressed discourse on social media and whether patterns differ by condition and language. X (formerly known as Twitter; X Corp) provides a large, more diverse, and spontaneous sample in real time, offering insights without the constraints of structured methodologies [13]. This platform facilitates anonymous communication by eliminating the complexities of face-to-face interactions, including nonverbal cues [14]. By using advanced machine learning methodologies to analyze posts, it is possible to quantify and characterize trends in public discourse on the platform, providing an alternative approach to traditional survey-based assessments of public perception [6].

In this study, we addressed the existing gap in understanding platform-level public discourse on neurodevelopmental disorders by analyzing a large sample of posts collected over more than a decade. We compared English- and Spanish-language posts to examine whether discourse patterns differ across large language communities. We hypothesized that, despite higher post volumes and engagement, public discourse on ADHD and ASD remains mixed, varying by condition and language. Our analysis sought to elucidate current opinions and attitudes, thereby informing future strategies for education and interventions.

Methods

Data Collection

Posts were gathered using a search engine called Tweet Binder (SocialBro SL) [15] that has access to the totality of public posts, covering the period from 2009 to 2023. Posts were analyzed over this same period, reflecting the full time range available in the exported dataset for our query at the time of retrieval. The start year corresponds to the earliest period with consistent availability of posts and engagement metadata in the dataset [16]. The data collection focused on posts that mentioned the following keywords related to mental illness in both English and Spanish: *ASD/autism*, *ADHD*, *depression*, *bipolardisorder*, *anxiety*, *substanceuse*, *schizophrenia*, *insomnia*, *suicide*. The dataset included posts in English and Spanish from all locations. Metadata associated with each post, including the date, time, user location (if available), and user engagement metrics such

XSL•FO
RenderX

as likes and reposts, were also collected. Engagement metrics included likes and reposts as provided by Tweet Binder; reply or comment counts were not available in the exported dataset and were not analyzed. The collected posts underwent filtering to remove duplicates, non-English or non-Spanish posts, and posts lacking substantial content (eg, posts with only hashtags, emojis, or single words). For clarity, we use the term "post" to refer to a unit of content on X (formerly called a "tweet" on Twitter).

Ethical Considerations

This study was approved by the Research Ethics Committee of the Universidad de Alcalá (approval code: OE 14_2020). The study was conducted in compliance with the ethical principles of the World Medical Association Declaration of Helsinki (7th revision, 2013). As the study exclusively analyzed publicly available, anonymized posts from X with no direct interaction with or identification of individual users, the ethics committee determined that individual informed consent was not required and granted a waiver of informed consent. No compensation was provided, as no human participants were directly enrolled in this study. All data were anonymized.

Machine Learning

Machine learning is essential for analyzing large datasets that are impractical to evaluate manually. As a subset of artificial intelligence, machine learning includes 3 main approaches: supervised, unsupervised, and semisupervised learning [17]. In this research, we used semisupervised learning, which combines elements of both supervised and unsupervised methods using both labeled and unlabeled data. This method extends traditional manual analysis, aiming to develop a model that replicates expert evaluations for classifying millions of posts.

The machine learning application begins with a preprocessing step where the posts are normalized by splitting negative contractions, removing special characters and repetitions, and converting emojis into text. Next, 2 manually classified datasets (1 in English and 1 in Spanish), each consisting of 1500 posts, were labeled by 2 study team annotators using a prespecified codebook; Spanish-language labels were completed by native Spanish annotators, and disagreements were resolved by consensus. We used 75% of each labeled dataset for training and 25% for testing, which were randomly divided into a 75% training subset (1125 posts) and a 25% testing subset (375 posts). The training subset was used to train a machine learning model for each classification category: *classification* (distinguishing posts relevant to mental health discussions from nonrelevant content); *trivialization* (identifying posts containing terms or expressions that minimize, dismiss, or undermine the seriousness, impact, or legitimacy of mental health conditions); *personalexperience* (capturing firsthand accounts or self-reported experiences related to a mental health condition); *user-perceived treatability* (assessing whether a condition is viewed as incurable, treatable with professional help, or self-manageable, reflecting whether a post suggested that symptoms can improve with treatment including medication and psychosocial interventions); and *causes* (detecting attributions regarding the origins of mental health conditions, whether multifactorial, biological or genetic, substance related, or psychosocial). The testing subset was then used to validate the models' performance.

Despite the availability of various pretrained models for text classification, we opted for the Transformer-Based Language Model Pretrained on English Posts (BERTweet) model [18] (VinAI Research) for the English dataset and Transformer-Based Language Model Pretrained on Spanish Text (BETO; Universidad de Chile) [19] for the Spanish dataset. BERTweet, trained on 80 GB of text containing more than 860 million English posts, was selected due to its extensive use in the literature [20,21] and its specific training on English posts similar to those we evaluated. For the Spanish dataset, we chose BETO, a Bidirectional Encoder Representations from Transformers model trained on a Spanish corpus, which is also popularly used in the literature [22,23].

To ensure these models accurately replicate expert analyses, fine-tuning was conducted for each category. This process involves adjusting the parameters of the pretrained models using data specific to the new task, leveraging the general knowledge acquired during pretraining on large, unlabeled datasets to adapt them for more specialized tasks. The Spanish dataset was used to fine-tune BETO, while the English dataset was used to fine-tune BERTweet. A common challenge during fine-tuning is the imbalance of options within each category. To address this, we used the Easy Data Augmentation (EDA) [24] pipeline to generate additional posts, ensuring balanced representation across categories. EDA generates new posts by replacing words with synonyms, randomly removing some words, and rearranging word positions. EDA was used only to augment the training data during model fine-tuning; synthetic EDA-generated posts were not included in the final analytic dataset or in results.

The performance of the fine-tuned models was evaluated using the test datasets, with the F_1-score used to measure accuracy across all categories. The English dataset models achieved the following F_1-scores: classification (0.78), trivialization (0.70), personal experience (0.69), user-perceived treatability (0.78), and causes (0.65). Similarly, the Spanish dataset models presented an equivalent performance: classification (0.79), trivialization (0.78), personal experience (0.69), user-perceived treatability (0.70), and causes (0.66). After confirming the models' effectiveness, they were deployed to classify the remaining posts. First, the classification category was applied to both datasets, and only the posts classified as relevant were further analyzed across the other categories, whereas those classified as irrelevant were discarded. There is no universally accepted F_1-score cutoff across tasks and class distributions; we report F_1-scores to allow comparison across our models and interpret them in the context of class imbalance and the complexity of each label. We used supervised machine learning to classify posts into predefined content categories. Within the personal experience category, posts were additionally categorized by valence (negative, neutral, or positive). Negative personal experience posts included experiences such as stigma or discrimination or distress related to the condition.

XSL•FO
RenderX

Descriptive Analysis

Descriptive statistics were presented both in tables and in different forms of graphical representation, including line graphs for the temporal evolution of post frequencies and clustered bar charts for comparing different categories of posts by language. These analyses were performed using the Python programming language (version 3.10.12; Python Software Foundation).

To assess the impact of ADHD and ASD on X, we analyzed the number of posts from 2009 to 2023 compared to other common psychiatric diagnoses. Specifically, we included anxiety disorders, schizophrenia, bipolar disorder, depressive disorders, substance use disorder (SUD), insomnia, and suicide.

Posts describing personal experiences were treated as one component of the broader public discourse and analyzed separately from general discussion themes. In this study, *public sentiment* refers to what people publicly say about these conditions on the platform, including the themes and attitudes expressed in posts and how widely posts are liked or shared.

Statistical Analysis

Statistical analysis of the data was performed using both parametric and nonparametric tests to assess between-group differences in categories associated with mental illness on X by language. Initially, we verified the completeness of the data by confirming the absence of missing data on relevant variables.

To test for statistical assumptions, the Shapiro-Wilk test was used to assess the normality of the residuals, and the Levene test evaluated the homogeneity of variances between groups. As the assumptions required for ANOVA were not fully met, alternative methods were chosen. A simplified ANOVA was performed to confirm the initial findings, and the Kruskal-Wallis test was applied as a nonparametric alternative. The results were subsequently adjusted using the Benjamini-Hochberg method to control for the false discovery rate and the Holm method to adjust for the type 1 error rate. To further explore differences between specific groups within each category, the Dunn test was used for pairwise comparison.

Additionally, the chi-square test was used to determine whether the observed differences between the frequencies of categories associated with different illnesses by language and the expected frequencies were statistically significant.

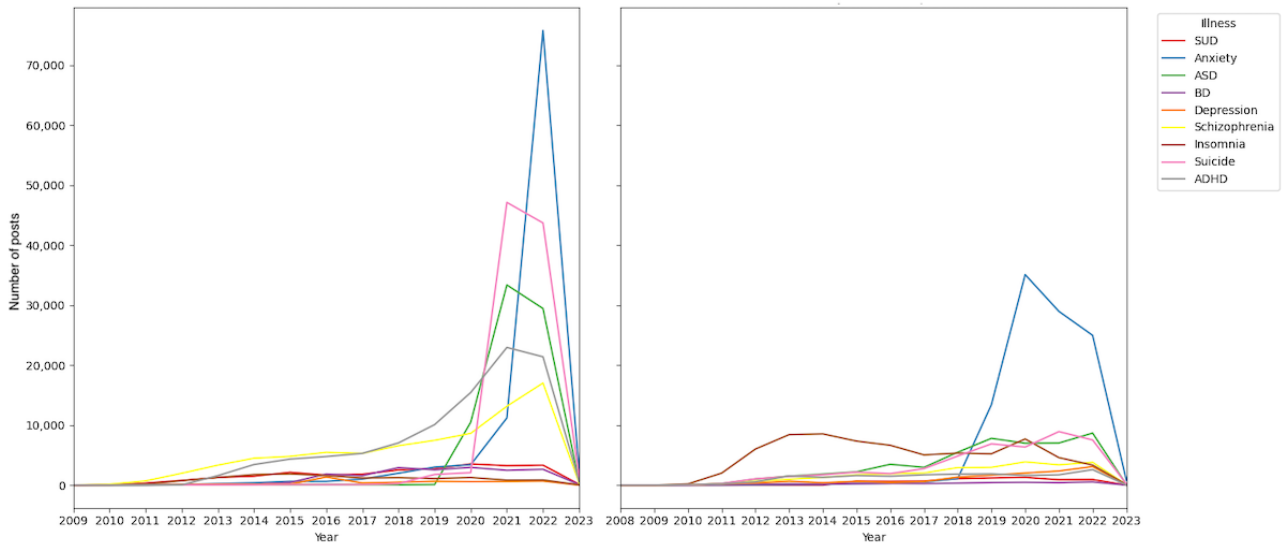
Results

Number of Posts Related to ADHD, ASD, and Other Mental Illnesses

From the total dataset of posts across all conditions, 852,990 were deemed classifiable for analysis. Among these classifiable posts, 511,510 of 852,990 (59.97%) posts were in English, and 341,480 (40.03%) were in Spanish. Overall, the number of posts about mental health conditions increased over the years (Figure 1). In English, 97,084 of 511,510 (18.98%) posts were related to ADHD, and 74,619 (14.59%) posts were related to ASD, with ADHD posts significantly more frequent than ASD posts (chi-square test $P<.001$). In Spanish, ADHD-related posts totaled 18,223 of 341,480 (5.34%) posts, while ASD-related posts totaled 49,475 (14.49%), with ASD posts significantly outnumbering ADHD posts (chi-square test $P<.001$).

In English-language posts, the most mentioned conditions were anxiety (100,723/511,510, 19.69% of posts); ADHD (97,084/511,510, 18.98% of posts); and suicide (96,817/511,510, 18.93% of posts). In Spanish-language posts, the most mentioned conditions were anxiety (105,780/341,480, 30.98% of posts); insomnia (70,561/341,480, 20.66% of posts); and ASD (49,475/341,480, 14.49% of posts).

Figure 1. Annual number of posts by mental health condition. Left: posts in English. Right: posts in Spanish. ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; BD: bipolar disorder; SUD: substance use disorder.



Engagement Analysis

Figure 2 presents line graphs illustrating the number of likes per post over time for various mental health conditions in both English- and Spanish-language posts. Overall, an increasing trend in likes per post was observed across conditions, particularly after 2019. Likes per post differed significantly across conditions in both English- and Spanish-language posts

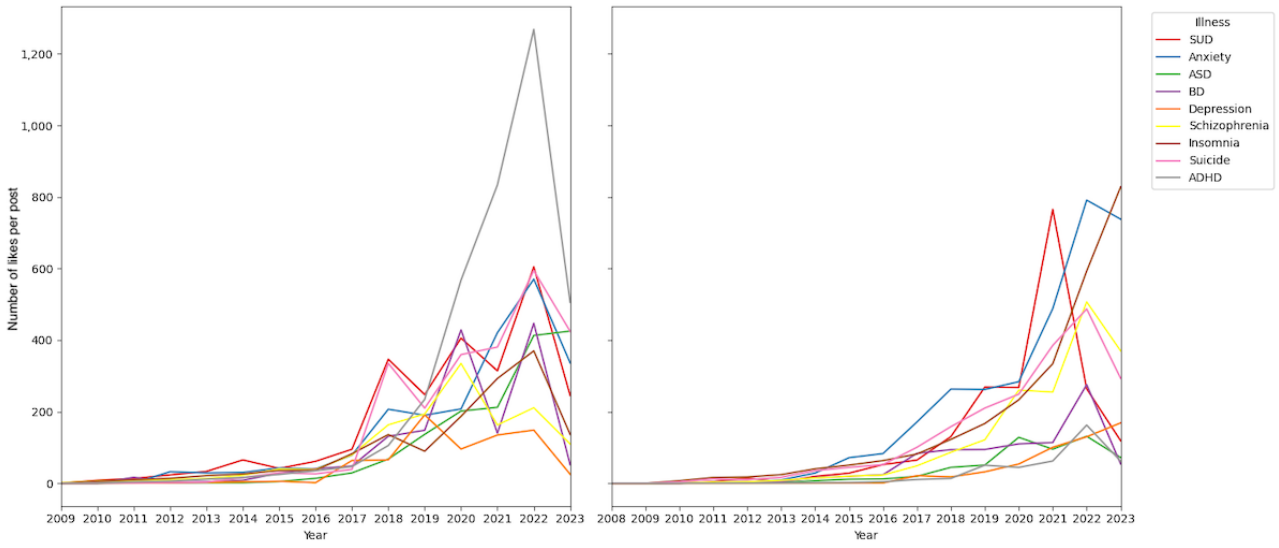
XSL•FO
RenderX

(*P*<.001 [Kruskal-Wallis test]). In English-language posts, ADHD-related posts experienced a significant increase in engagement, peaking around 2021, with likes per post reaching more than 1200. This period coincided with the COVID-19 pandemic. However, ASD did not receive the same level of attention as ADHD, with engagement levels remaining relatively low compared with other conditions. Across the full study period, ADHD posts had higher likes per post than ASD posts in both English and Spanish (both *P*<.001 [Kruskal-Wallis test]). Among the other conditions, anxiety, SUD, and suicide maintained steady engagement over the years, with several conditions surpassing 500 likes per post at their peaks. In contrast, engagement with ADHD-related posts in Spanish was much lower than in English and only slightly exceeded that of ASD at its peak in 2022. ADHD likes per post differed significantly between English- and Spanish-language posts (*P*<.001 [Kruskal-Wallis test]). Anxiety, SUD, and insomnia showed significant engagement, peaking during the pandemic. Depression, schizophrenia, and suicide followed similar trends with an increase in likes around 2020 to 2021.

Figure 3 presents line graphs illustrating the number of reposts per post over time for various mental health conditions. Overall, there was a noticeable increase in reposts per post over time. Reposts per post differed significantly across conditions in both English and Spanish posts (*P*<.001 [Kruskal-Wallis test]). In English-language posts, ADHD-related posts showed notable engagement, with a pronounced increase between 2018 and 2022. ASD-related posts did not receive the same level of attention, with engagement levels remaining relatively low. Suicide-related posts had the highest peak in engagement, with the number of reposts per post nearing 200 around 2018 and another significant but smaller peak around 2022. Anxiety-related posts also experienced a significant increase in engagement, with peaks around 2012, 2018, and 2022. SUD-related posts maintained steady engagement over the years, with noticeable peaks in 2018, 2020, and 2022.

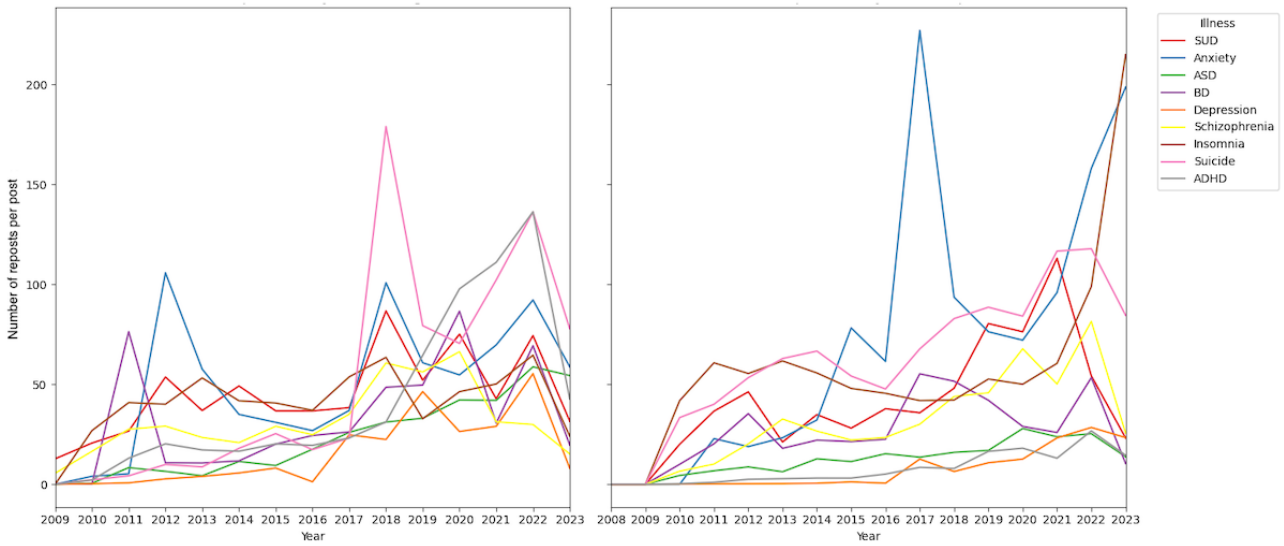
In Spanish-language posts, ADHD-related posts showed increased engagement but did not reach the high levels observed in English. ASD-related posts remained less prominent in terms of engagement, similar to the pattern observed in English-language posts. In Spanish-language posts, anxiety-related posts showed the most significant increase in engagement, with reposts per post peaking in 2017 and again in 2022. A similar pattern was observed with insomnia-related posts. SUD-related posts also showed notable engagement, with a peak around 2021. Across the full study period, ADHD posts had higher reposts per post than ASD posts in both English and Spanish (both *P*<.001 [Kruskal-Wallis test]), and ADHD reposts per post differed significantly between English- and Spanish-language posts (*P*<.001 [Kruskal-Wallis test]).

Figure 2. Line graphs illustrating the number of likes per post over time for various mental health conditions in both English (left) and Spanish (right). ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; BD: bipolar disorder; SUD: substance use disorder.



XSL•FO
RenderX

Figure 3. Number of reposts per post over time for various mental health conditions in both English (left) and Spanish (right). ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; BD: bipolar disorder; SUD: substance use disorder.



Types of Personal Experiences in Posts About ADHD and ASD Compared to Other Mental Illnesses

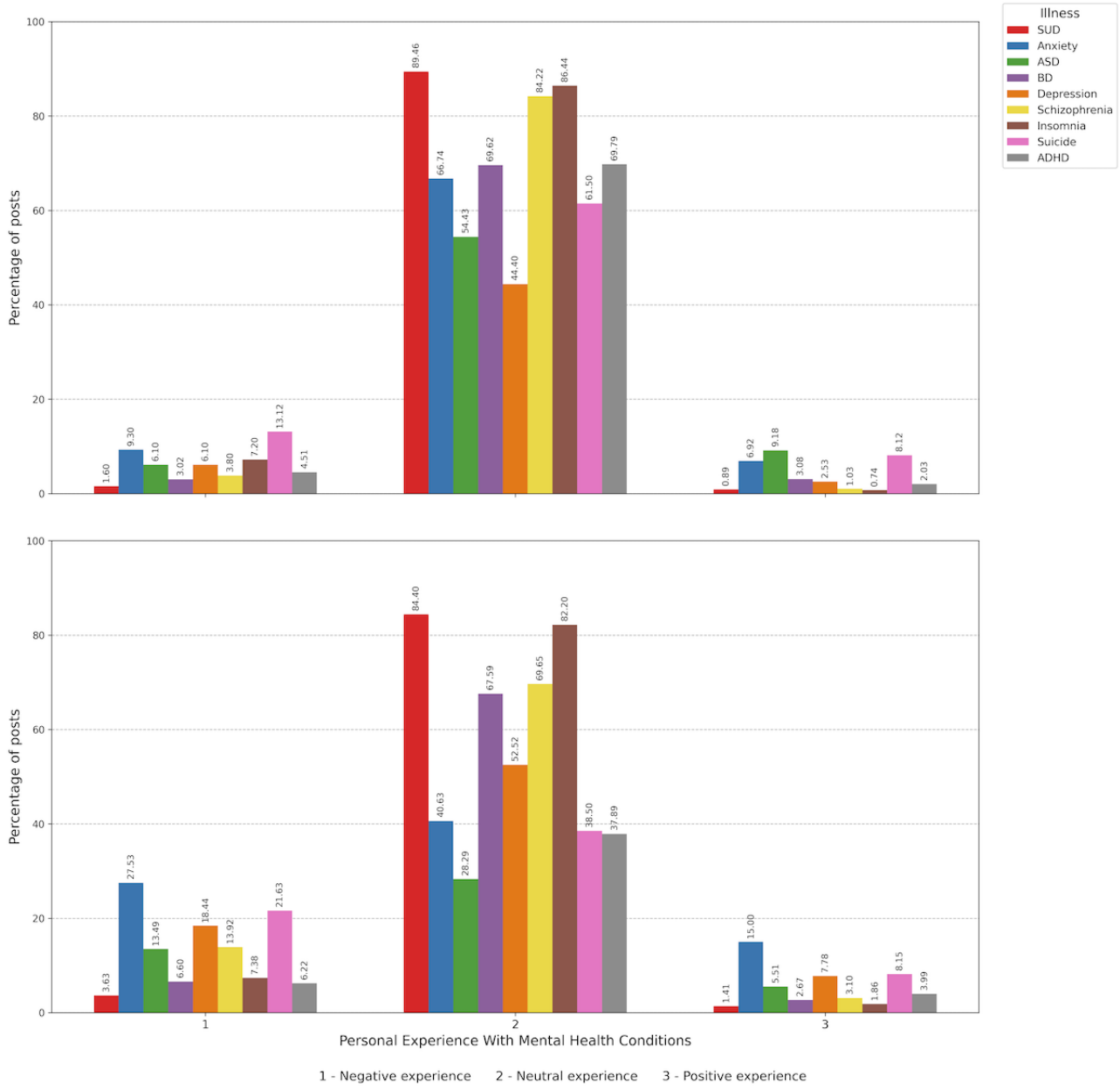
Figure 4 illustrates the distribution of posts sharing personal experiences. Among ADHD-related posts in English, 4378 of 97,084 (4.51%) posts described negative experiences, including stigmatization; 67,755 (69.79%) were neutral; and 1971 (2.03%) conveyed positive experiences. For ASD, 4552 of 74,619 (6.10%) posts indicated negative experiences; 40,615 (54.43%) were neutral; and 6850 (9.18%) were positive.

Among other mental illnesses, including anxiety, bipolar disorder, depression, insomnia, SUD, schizophrenia, and suicide, the distribution followed a similar trend, with an average of 6.30% of negative experiences (SD 3.99%, range 1.60%-13.13%); 71.77% of neutral experiences (SD 16.17%, range 44.38%-89.46%); and 3.33% of positive experiences (SD 3.01%, range 0.74%-8.12%).

When comparing languages, Spanish-language posts about ADHD showed a higher percentage of both negative (1133/18,223, 6.22% vs 4378/97,084, 4.51%) and positive (727/18,223, 3.99% vs 1971/97,084, 2.03%) experiences but fewer neutral experiences (37.89% vs 69.79%) than English posts. Similarly, Spanish-language posts about ASD contained more negative (6674/49,475, 13.49% vs 4552/74,619, 6.1%) experiences but fewer positive (2726/49,475, 5.51% vs 6850/74,619, 9.18%) and neutral (13,996/49,475, 28.29% vs 40,615/74,619, 54.43%) experiences than their English counterparts. Spanish posts were more polarized, showing more negative and positive experiences, while English posts were more often neutral. Language differences in the distribution of personal experience categories were statistically significant for both ADHD and ASD (both chi-square test $P<.001$). This pattern was consistent across various mental illnesses.

XSL•FO
RenderX

Figure 4. Distribution of posts that published personal experiences with mental health conditions. Top panel: posts in English. Bottom panel: posts in Spanish. ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; BD: bipolar disorder; SUD: substance use disorder.



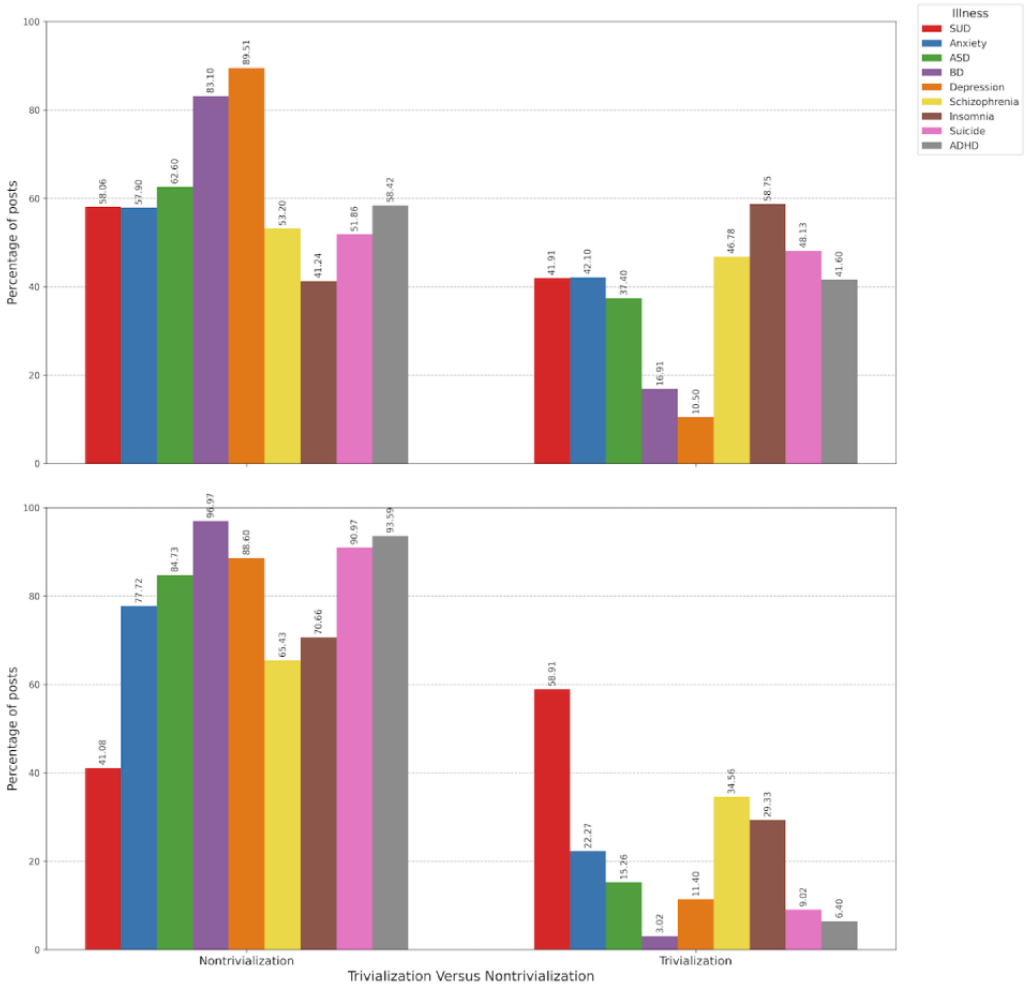
Trivialization of ADHD and ASD in Posts Compared to Other Mental Illnesses

Figure 5 shows that nontrivializing posts were more common overall, with this trend being even more pronounced in Spanish-language posts. In English-language posts, nontrivializing posts accounted for an average of 61.77% (SD 15.24%, range 41.24%-89.51%). ADHD- and ASD-related posts followed this pattern, with 56,716 of 97,084 (58.42%) posts and 46,711 of 74,619 (62.60%) posts being nontrivializing, respectively.

In Spanish-language posts, the percentage of nontrivializing posts was higher (mean 78.86%, SD 17.67%, range 41.08%-96.97%). For ADHD- and ASD-related posts, nontrivializing posts accounted for 17,055 of 18,223 (93.59%) posts and 41,933 of 49,475 (84.73%) posts, respectively, with significant differences between English- and Spanish-language posts for both conditions (both chi-square test $P<.001$).

XSL•FO
RenderX

Figure 5. Comparison of nontrivializing vs trivializing posts about mental health conditions. Top panel: posts in English. Bottom panel: posts in Spanish. ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; BD: bipolar disorder; SUD: substance use disorder.



User-Perceived Causes of Mental Illness Analysis

Figure 6 illustrates user-perceived causes of mental illness. For ADHD in English-language posts, the most commonly attributed causes were multifactorial factors (60,124/97,084, 61.93%), followed by substance use (15,281/97,084, 15.74%) and psychological susceptibility (12,572/97,084, 12.95%). For ASD-related posts, the leading causes were multifactorial factors (45,144/74,619, 60.5%); acute psychosocial stressors (13,461/74,619, 18.04%); and psychological susceptibility (7551/74,619, 10.12%). Notably, substance use was associated with ADHD in 15,281 of 97,084 (15.74%) English-language posts, whereas in Spanish-language posts, this association was much lower (805/18,223, 4.42%). For other psychiatric conditions in English-language posts, multifactorial causes were the most frequently cited for anxiety (32,775/100,723, 32.54%); bipolar disorder (7517/18,061, 41.62%); depression (2055/5473, 37.55%); insomnia (10,880/14,082, 77.26%); SUD (8930/25,190, 35.45%); schizophrenia (31,791/79,478, 40%); and suicide (36,326/96,817, 37.52%). Psychological susceptibility was the most common perceived cause for anxiety (37,590/100,723, 37.32%) and was the second most common cause for schizophrenia (19,130/79,478, 24.07%) and suicide (30,313/96,817, 31.31%). Unsurprisingly, substance use was the most frequently cited cause for SUD (12,215/25,190, 48.49%), while it was also commonly mentioned in bipolar disorder–related posts (6827/18,061, 37.8%). Overall, the distribution of perceived-cause categories differed significantly between ADHD and ASD in English-language posts (chi-square test $P<.001$).

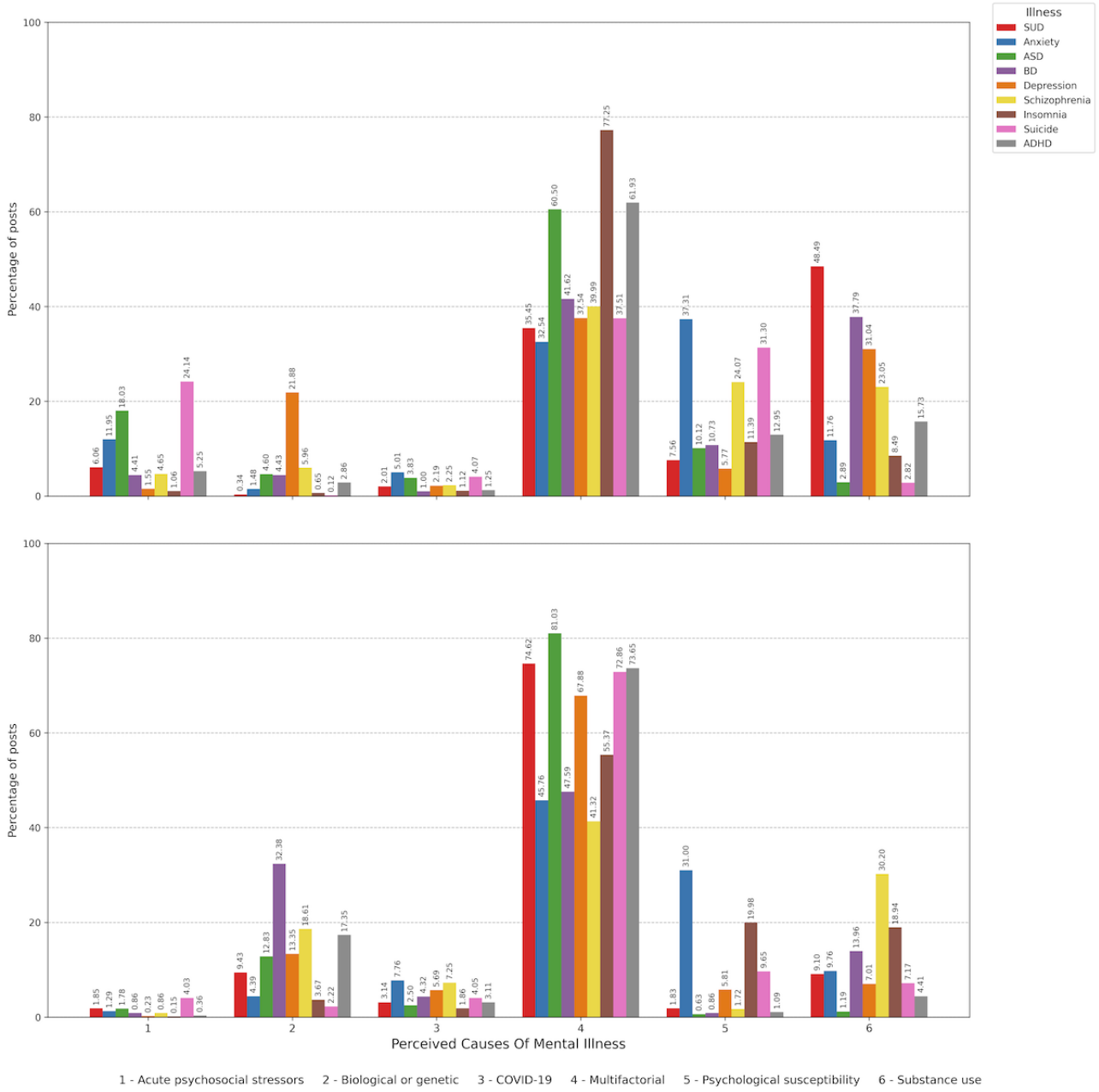
In Spanish-language posts, there were some differences in perceived causes. For ADHD, the most attributed causes were multifactorial (13,421/18,223, 73.65%), followed by biological or genetic factors (3162/18,223, 17.35%) and substance use (805/18,223, 4.42%). For ASD, the leading causes were multifactorial (40,095/49,475, 81.04%), biological or genetic factors (6353/49,475, 12.84%), and COVID-19 (1237/49,475, 2.5%). For other psychiatric conditions in Spanish-language posts, multifactorial causes were the most frequently cited, with higher percentages than in English-language posts for SUD (5606/7512, 74.63%); suicide (33,670/46,205, 72.87%); depression (9517/14,019, 67.89%); insomnia (39,077/70,561, 55.38%); bipolar disorder (1763/3703, 47.6%); anxiety (48,416/105,780, 45.77%); and schizophrenia (10,751/26,019, 41.32%). Biological or genetic factors were the second most frequently cited cause for bipolar disorder (1199/3703, 32.39%) and SUD (709/7512, 9.44%), while substance use was the second most cited cause for schizophrenia (7860/26,019, 30.21%). Psychological susceptibility was most frequently cited for anxiety (32,802/105,780, 31.01%) but was attributed less frequently to other conditions. Overall, the distribution of

XSL•FO
RenderX

perceived-cause categories differed significantly between ADHD and ASD in Spanish-language posts (chi-square test $P<.001$).

When comparing languages, the distribution of perceived-cause categories differed significantly between English- and Spanish-language posts for both ADHD and ASD (both chi-square test $P<.001$).

Figure 6. User-perceived causes of mental illness. Top panel: posts in English. Bottom panel: posts in Spanish. ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; BD: bipolar disorder; SUD: substance use disorder.



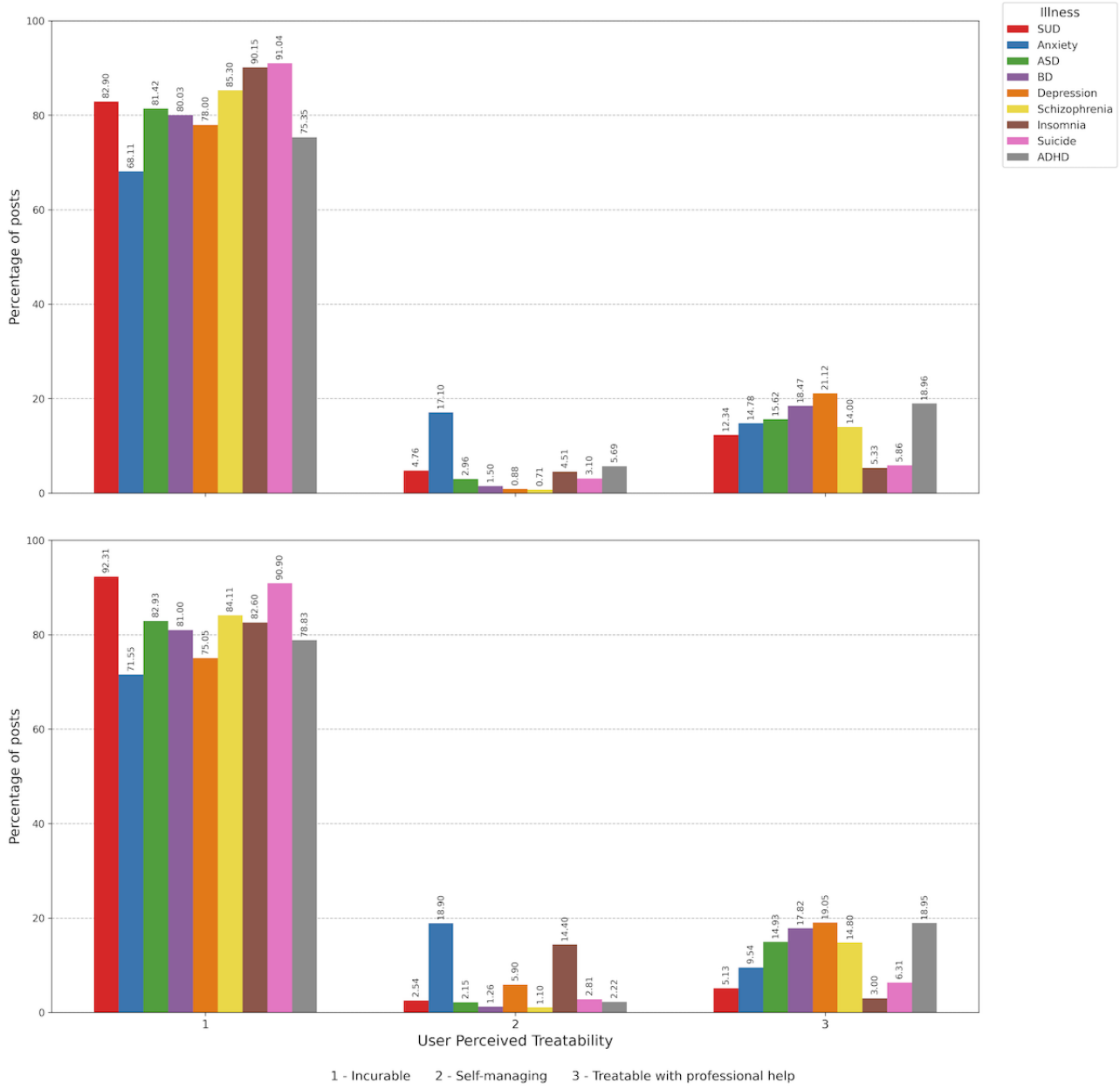
Percentage of User-Perceived Treatability of Mental Illnesses in Posts About ADHD and ASD vs Other Mental Illnesses

Figure 7 illustrates the user-perceived treatability of mental illnesses. Among English-language posts about ADHD, 73,153 of 97,084 (75.35%) users view it as incurable; 18,407 (18.96%) describe it as treatable with professional help; and 5524 (5.69%) mention that it can be self-managed. For ASD, the perception of incurability was even higher at 60,755 of 74,619 (81.42%) posts, with 11,655 (15.62%) mentioning it is treatable with professional help and 2209 (2.96%) referring to it as self-manageable. The distribution of treatability categories differed between ADHD and ASD posts in English (chi-square test $P<.001$). For other psychiatric conditions, English-language posts reflect a high perception of incurability (mean 82.22%, SD 7.88%, range 68.12%-91.04%). The perception that these conditions are treatable with professional help averages 13.13% (SD 5.92%, range 5.33%-21.12%), while self-manageability averages 4.65% (SD 5.73%, range 0.71%-17.10%). Anxiety stands out, with 17,224 of 100,723 (17.10%) users stating it can be self-managed, significantly higher than the 3.01% average for other conditions when anxiety was excluded (chi-square test $P<.001$).

XSL•FO
RenderX

In Spanish-language posts, some differences emerged. For ADHD, 14,365 of 18,223 (78.83%) users perceived it as incurable, slightly higher than the sentiment in English-language posts, while 3453 of 18,223 (18.95%) users believed it is treatable with professional help. The perception of self-manageability was lower (405/18,223, 2.22%). For ASD, the perception of incurability was even higher (41,030/49,475, 82.93%), while 7387 (14.93%) believed it was treatable with professional help, and only 1064 (2.15%) thought it can be self-managed. The distribution of treatability categories differed between ADHD and ASD posts in the Spanish language (chi-square test $P<.001$). For other psychiatric conditions, Spanish-language posts indicated a similarly high perception of incurability at 82.49% (SD 7.60%, range 71.56%-92.32%). However, the perception of treatability with professional help was lower at 10.81% (SD 6.43%, range 2.99%-19.05%), while the perception of self-manageability was slightly higher at 6.70% (SD 7.10%, range 1.10%-18.90%). When comparing languages, the distribution of treatability categories differed significantly between English- and Spanish-language posts for both ADHD and ASD (both chi-square test $P<.001$).

Figure 7. User-perceived treatability of mental illnesses. Top panel: posts in English. Bottom panel: posts in Spanish. ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; BD: bipolar disorder; SUD: substance use disorder.



Discussion

Principal Findings

The study used machine learning techniques to analyze data from X on public perceptions of ADHD and ASD, with other mental illnesses included as comparators. Our analysis of 852,990 posts revealed a substantial increase in post volume and engagement across the observation period. Notably, ADHD-related posts in English experienced a marked surge in engagement (likes and reposts) during the COVID-19 pandemic, with likes per post exceeding 1200 between 2020 and 2023. In

XSL•FO
RenderX

contrast, ASD consistently received lower engagement in both English- and Spanish-language posts. Personal experiences differed by language, with ADHD posts being predominantly negative in both languages and ASD posts being mostly positive in English but largely negative in Spanish. Overall, posts conveyed a nontrivializing attitude toward neurodevelopmental disorders, particularly in Spanish. Perceived causes also varied, with English-language posts frequently citing substance use and psychological factors, while Spanish-language posts emphasized biological or genetic causes, with substance use cited as a secondary factor for ADHD and COVID-19 cited as a secondary factor for ASD. Users generally viewed mental health conditions as chronic with limited optimism for treatment; fewer than 7% believed in self-management, except for anxiety. Professional help was rarely endorsed, although ADHD was viewed as more treatable than ASD. These insights can inform clinicians and researchers about public attitudes and experiences regarding neurodevelopmental disorders, particularly ADHD and ASD, underscoring the value of social media analysis in providing real-time, unfiltered, and empirically derived data in areas where traditional research methods may encounter limitations.

The findings suggest growing public interest in ADHD during the analyzed period, while ASD received comparatively less attention [25]. Diagnostic criteria changes in *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (2013) [26], including older age of onset, fewer symptom requirements for adults, and addition of age-appropriate clinical examples, likely contributed to the increase in ADHD awareness and prevalence [27,28]. Social media proliferation of ADHD content may have further boosted public engagement [29]. Furthermore, the COVID-19 pandemic may have exacerbated ADHD symptoms, increasing recognition and diagnosis [30,31]. Butt et al [2] found that ADHD-related visits in Ontario increased by 32% compared with visit rates before the pandemic, especially among female individuals aged 5 to 9 and 20 to 55 years. Conversely, ASD awareness has grown but not to the extent of ADHD awareness, possibly due to individuals with autism favoring limited social interactions during the pandemic [32]. Pandemic lockdowns may have reduced social demands, easing certain stressors for individuals with ASD [33]. Thorell et al [34] reported that distance learning negatively impacted children with ADHD more than those with ASD. In the same study, families affected by ASD even reported some positive effects, such as reduced social pressure [34]. Additionally, it is possible that individuals with ASD participate more in private online communities that are not captured by mainstream social media [35,36].

While awareness of ADHD and ASD has increased, experiences shared online reveal persistent challenges. ADHD was associated predominantly with negative experiences in both languages, consistent with previous findings [7,37,38]. Mueller et al [7] noted significant ADHD-related stigma, including bullying and social withdrawal. ASD elicited mostly positive experiences in English-language posts but predominantly negative experiences in Spanish-language posts. Morgan et al [39] highlighted systemic barriers, including language, cultural obstacles, and issues with translated screening tools, limiting timely ASD diagnosis for Spanish speakers and increasing negative experiences. Griffiths et al [38] reported that adults with autism experienced higher rates of negative life events, such as abuse and financial exploitation, leading to increased anxiety, depression, and reduced life satisfaction. Similarly, Cooper et al [37] found that individuals with autism had lower self-esteem and higher depression and anxiety compared with peers without autism. Families also reported stigma and discrimination, negatively impacting their mental health and ability to provide support [40].

It is unclear whether increased awareness of ADHD and ASD has significantly reduced stigma and trivialization [3,4,7]. Social media has likely enhanced visibility, empathy, and acceptance through authentic portrayals and advocacy [29,41]. Educational initiatives and the more recent neurodiversity movement also reinforced positive attitudes [42]. Dillenburger et al [12] found ASD awareness among school peers increased from 50% in those aged 11 years to 80% in those aged 16 years; half knew someone with autism, while about 3% self-identified as autistic. However, Alsehemi et al [43] found that although 88% of respondents had heard of ASD, 41% admitted limited understanding.

Although most users endorsed multifactorial causes, the perceived contributing factors for ADHD and ASD varied by language. English-language posts often linked mental illness to substance use and psychological vulnerability, while Spanish-language posts emphasized biological and genetic explanations. ADHD was frequently described as multifactorial in English-language posts, with substance use and psychological factors cited, aligning with the literature on ADHD's complex genetic and environmental etiology [44]. In contrast, Spanish-language posts primarily attributed both ADHD and ASD to biological or genetic causes, with secondary mentions of substance use for ADHD and COVID-19 for ASD. These findings align with research on ADHD's complex etiology, involving genetic and environmental influences [44]. Other studies have also noted frequent user emphasis on behavioral outcomes, such as self-medication and mental health struggles [45]. The perception of ASD as multifactorial echoes findings by Mitchell and Locke [46], who reported that the public often cites genetic and neurological causes for this condition.

Our study found the public typically views mental health conditions as chronic, with limited optimism about treatment outcomes. However, ADHD is uniquely seen as more treatable, comparable to depression, likely reflecting its documented responsiveness and generally improved prognosis into adulthood [47]. Conversely, ASD is perceived as less amenable to significant improvement. The limited interest in professional support expressed in posts contrasts with other findings, such as Angermeyer et al [48], who reported high public regard for professional help. This discrepancy may result from survey biases, including social desirability—respondents providing socially acceptable answers—and selection bias, as more engaged individuals typically participate. Anxiety was an exception regarding perceived self-manageability. This may be due to its high lifetime prevalence of 28.8% [49] and broad public awareness [50]. The perception may also be influenced by the widespread availability of self-help resources, including books, websites, and apps [51], as well as by the demonstrated

XSL•FO
RenderX

effectiveness of nonpharmacological interventions such as cognitive behavioral therapy, which can often be self-administered [52,53].

Using X offers advantages over traditional surveys, which may suffer from response distortion and nonresponse bias [8,9]. X provides real-time, unsolicited opinions from a broad user base, potentially yielding more naturalistic and representative insights into public views on mental health. This study has limitations. X users may not represent the general population, as they skew younger and more tech-savvy [54]. Only English- and Spanish-language posts were analyzed, excluding other language groups. Language is an imperfect proxy for geography and health systems; we interpreted differences as discourse differences rather than as country-level differences. Keyword-based collection may have missed relevant posts, and despite data cleaning, misclassification and noise remain risks. Our keyword strategy did not include legacy-diagnostic terms (eg, Asperger); posts using only these terms may be undercaptured, particularly in earlier years of the study period. Public sentiment here refers to publicly available discourse on the platform. We did not reliably separate individual users from professional or institutional accounts, which may influence observed discourse patterns. Analyses were conducted at the post level rather than the individual user level; therefore, comention of multiple conditions reflects discourse overlap and cannot be interpreted as clinical comorbidity. Individual experiences based solely on text may miss emotional nuance, and the study's observational design limits causal interpretations. Future research should expand to additional social media platforms and incorporate a broader range of languages to capture more diverse perspectives. Incorporating advanced sentiment analysis techniques, such as context-aware and emotion-specific models, could further deepen our understanding of public attitudes toward mental health.

Conclusions

This study applied machine learning to data obtained from X to examine public perceptions of ADHD and ASD. ADHD-related content saw a surge in engagement during the COVID-19 pandemic, while ASD-related content received less attention. Although most posts were neutral, negative experiences—particularly for ADHD—were more common. Users generally viewed mental health conditions as chronic and difficult to treat, with fewer than 7% expressing belief in self-management, except for anxiety. In English-language posts, perceived causes of ADHD and ASD were largely multifactorial, often citing substance use and psychological factors. Spanish-language posts followed similar patterns in tone but emphasized biological and genetic factor explanations.

These insights can inform clinicians and researchers about public attitudes and experiences regarding neurodevelopmental disorders, particularly ADHD and ASD, underscoring the value of social media analysis in capturing real-time, unfiltered, and empirically derived data—especially in areas where traditional research methods may be limited by response biases or restricted reach.

Acknowledgments

JPCP is supported by the Alicia Koplowitz Foundation.

Funding

This study (FIS-PI22/00653) was supported by the Instituto de Salud Carlos III (Plan Estatal de I+D+I 2020-2027) and co-financed by ProA Capital and Halekulani S.L. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

JQ has served as a speaker and participated in scientific advisory boards for Takeda, Sincrolab, Janssen, Bial, and Lilly. He has also received research funding from the Instituto de Salud Carlos III. JPCP is supported by the Alicia Koplowitz Foundation. All other authors declare no conflicts of interest.

References

1. Thapar A, Cooper M, Rutter M. Neurodevelopmental disorders. Lancet Psychiatry 2017 Apr;4(4):339-346. [doi: 10.1016/s2215-0366(16)30376-5]
2. Butt DA, Stephenson E, Kalia S, Moineddin R, Tu K. Patient visits and prescriptions for attention-deficit/hyperactivity disorder from 2017-2021: impacts of COVID-19 pandemic in primary care. PLoS One 2023 Mar 13;18(3):e0281307 [FREE Full text] [doi: 10.1371/journal.pone.0281307] [Medline: 36913355]
3. Stuart H, Arboleda-Flórez J. A public health perspective on the stigmatization of mental illnesses. Public Health Rev 2012 Dec 4;34:12. [doi: 10.1007/bf03391680]
4. Hinshaw SP, Cicchetti D. Stigma and mental disorder: conceptions of illness, public attitudes, personal disclosure, and social policy. Dev Psychopathol 2000 Jan 16;12(4):555-598. [doi: 10.1017/s0954579400004028] [Medline: 11202034]
5. Gerber JS, Offit PA. Vaccines and autism: a tale of shifting hypotheses. Clin Infect Dis 2009 Feb 15;48(4):456-461 [FREE Full text] [doi: 10.1086/596476] [Medline: 19128068]

XSL•FO
RenderX

6. Robinson P, Turk D, Jilka S, Cella M. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation. Soc Psychiatry Psychiatr Epidemiol 2019 Jan 1;54(1):51-58 [FREE Full text] [doi: 10.1007/s00127-018-1571-5] [Medline: 30069754]
7. Mueller AK, Fuermaier AB, Koerts J, Tucha L. Stigma in attention deficit hyperactivity disorder. Atten Defic Hyperact Disord 2012 Sep;4(3):101-114 [FREE Full text] [doi: 10.1007/s12402-012-0085-3] [Medline: 22773377]
8. Kamakura WA. Common methods bias. In: Wiley International Encyclopedia of Marketing. Hoboken, NJ: John Wiley & Sons; 2010.
9. Gove WR, Geerken MR. Response bias in surveys of mental health: an empirical investigation. AJS 1977 May;82(6):1289-1317. [doi: 10.1086/226466] [Medline: 889001]
10. Maenner MJ, Warren Z, Williams AR, Amoakohene E, Bakian AV, Bilder DA, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. MMWR Surveill Summ 2023 Mar 24;72(2):1-14 [FREE Full text] [doi: 10.15585/mmwr.ss7202a1] [Medline: 36952288]
11. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry 2007 Jun;164(6):942-948. [doi: 10.1176/ajp.2007.164.6.942] [Medline: 17541055]
12. Dillenburger K, Jordan JA, McKerr L, Lloyd K, Schubotz D. Autism awareness in children and young people: surveys of two populations. J Intellect Disabil Res 2017 Aug 07;61(8):766-777. [doi: 10.1111/jir.12389] [Medline: 28593714]
13. Lachmar EM, Wittenborn AK, Bogen KW, McCauley HL. #MyDepressionLooksLike: examining public discourse about depression on Twitter. JMIR Ment Health 2017 Oct 18;4(4):e43 [FREE Full text] [doi: 10.2196/mental.8141] [Medline: 29046270]
14. Pinchevski A, Peters JD. Autism and new media: disability between technology and society. New Media Soc 2016 Jul 09;18(11):2507-2523. [doi: 10.1177/1461444815594441]
15. Tweet Binder - X (Twitter) hashtag analytics. Tweet Binder. URL: https://www.tweetbinder.com/ [accessed 2025-03-30]
16. Retweet limited rollout. X Corp. URL: https://blog.x.com/en_us/a/2009/retweet-limited-rollout?utm [accessed 2026-01-04]
17. Berry MW, Mohamed A, Yap BW. Supervised and Unsupervised Learning for Data Science. Cham, Switzerland: Springer; 2019.
18. Nguyen DQ, Vu T, Nguyen AT. BERTweet: a pre-trained language model for English tweets. In: Proceedings of the 2020 EMNLP. 2020 Presented at: EMNLP 2020; November 16-20, 2020; Online. [doi: 10.18653/v1/2020.emnlp-demos.2]
19. Cañete J, Chaperon G, Fuentes R, Ho JH, Kang H, Pérez J. Spanish pre-trained BERT model and evaluation data. ArXiv Preprint posted online on August 6, 2023 [FREE Full text]
20. Kumar A, Singh JP, Singh AK. COVID-19 fake news detection using ensemble-based deep learning model. IT Prof 2022 Mar 1;24(2):32-37. [doi: 10.1109/mitp.2022.3149140]
21. Kim MG, Kim M, Kim JH, Kim K. Fine-tuning BERT models to classify misinformation on garlic and COVID-19 on Twitter. Int J Environ Res Public Health 2022 Apr 22;19(9):5126 [FREE Full text] [doi: 10.3390/ijerph19095126] [Medline: 35564518]
22. Huang J, Han Z, Zhu R, Guo M, Sun K. Conspiracy theory text classification based on CT-BERT and BETO models. In: Proceedings of the Conference and Labs of the Evaluation Forum. 2024 Presented at: CLEF 2024; September 09-12, 2024; Grenoble, France URL: https://ceur-ws.org/Vol-3740/paper-242.pdf
23. de Arriba A, Oriol M, Franch X. Applying sentiment analysis on Spanish tweets using BETO. In: Proceedings of the Iberian Languages Evaluation Forum. 2021 Presented at: IberLEF 2021; September 21, 2021; Málaga, Spain URL: https://ceur-ws.org/Vol-2943/emoeval_paper9.pdf
24. Wei J, Zou K. EDA: easy data augmentation techniques for boosting performance on text classification tasks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019 Presented at: EMNLP-IJCNLP 2019; January 31, 2019; Hong Kong, China. [doi: 10.18653/v1/d19-1670]
25. Abdelnour E, Jansen M, Gold J. ADHD diagnostic trends: increased recognition or overdiagnosis? Mo Med 2022;119(5):467-473 [FREE Full text] [Medline: 36337990]
26. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
27. Sibley MH, Waxmonsky JG, Robb JA, Pelham WE. Implications of changes for the field: ADHD. J Learn Disabil 2013 Nov 05;46(1):34-42. [doi: 10.1177/0022219412464350] [Medline: 23128456]
28. Matte B, Anselmi L, Salum GA, Kieling C, Gonçalves H, Menezes A, et al. ADHD in DSM-5: a field trial in a large, representative sample of 18- to 19-year-old adults. Psychol Med 2015 Jan;45(2):361-373. [doi: 10.1017/S0033291714001470] [Medline: 25066615]
29. 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 [FREE Full text] [doi: 10.1177/07067437221082854] [Medline: 35196157]

XSL•FO
RenderX

30. Rogers MA, MacLean J. ADHD symptoms increased during the Covid-19 pandemic: a meta-analysis. J Atten Disord 2023 Jun 06;27(8):800-811 [FREE Full text] [doi: 10.1177/10870547231158750] [Medline: 36879524]
31. Auro K, Holopainen I, Perola M, Havulinna AS, Raevuori A. Attention-deficit/hyperactivity disorder diagnoses in Finland during the COVID-19 pandemic. JAMA Netw Open 2024 Jun 03;7(6):e2418204 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.18204] [Medline: 38935377]
32. Kalantari N, Payandeh A, Zampieri M, Motti V. Understanding the language of ADHD and autism communities on social media. In: Proceedings of the IEEE International Conference on Big Data. 2023 Presented at: BigData 2023; December 15-18, 2023; Sorrento, Italy URL: https://ieeexplore.ieee.org/document/10386833 [doi: 10.1109/bigdata59044.2023.10386833]
33. Lugo-Marín J, Gisbert-Gustemps L, Setien-Ramos I, Español-Martín G, Ibañez-Jimenez P, Forner-Puntonet M, et al. COVID-19 pandemic effects in people with Autism Spectrum Disorder and their caregivers: evaluation of social distancing and lockdown impact on mental health and general status. Res Autism Spectr Disord 2021 May;83:101757 [FREE Full text] [doi: 10.1016/j.rasd.2021.101757] [Medline: 33649707]
34. Thorell LB, Fuermaier AB, Christiansen H, Steinmayr R, Baeyens D, de la Peña AG, et al. Distance learning during the COVID-19 pandemic for children with ADHD and/or ASD: a European multi-center study examining the role of executive function deficits and age. Child Adolesc Psychiatry Ment Health 2022 Dec 13;16(1):101 [FREE Full text] [doi: 10.1186/s13034-022-00540-4] [Medline: 36514179]
35. Sánchez Amate JJ, Luque de la Rosa A. Transformation of the mental health of the autism spectrum community: contemporary challenges in the post-pandemic era. Brain Sci 2025 Feb 11;15(2):178 [FREE Full text] [doi: 10.3390/brainsci15020178] [Medline: 40002510]
36. Bakombo S, Ewalefo P, Konkle AT. The influence of social media on the perception of autism spectrum disorders: content analysis of public discourse on YouTube videos. Int J Environ Res Public Health 2023 Feb 13;20(4):3246 [FREE Full text] [doi: 10.3390/ijerph20043246] [Medline: 36833941]
37. Cooper K, Smith LG, Russell A. Social identity, self‐esteem, and mental health in autism. Eur J Soc Psychol 2017 Oct 02;47(7):844-854. [doi: 10.1002/ejsp.2297]
38. Griffiths S, Allison C, Kenny R, Holt R, Smith P, Baron-Cohen S. The Vulnerability Experiences Quotient (VEQ): a study of vulnerability, mental health and life satisfaction in autistic adults. Autism Res 2019 Oct 05;12(10):1516-1528 [FREE Full text] [doi: 10.1002/aur.2162] [Medline: 31274233]
39. Morgan CH, Abdul-Chani M, Sherrod GM, King KP, Guest KC. The influence of Spanish language on autism spectrum disorder screening: a systematic review. Int J Dev Disabil 2024 Apr 10;70(8):1343-1366 [FREE Full text] [doi: 10.1080/20473869.2023.2193490] [Medline: 39713504]
40. Yin M, Li Z, Zhou C. Experience of stigma among family members of people with severe mental illness: a qualitative systematic review. Int J Ment Health Nurs 2020 Apr 24;29(2):141-160. [doi: 10.1111/inm.12668] [Medline: 31648408]
41. Tint A, Weiss JA. Family wellbeing of individuals with autism spectrum disorder: a scoping review. Autism 2016 Apr 06;20(3):262-275. [doi: 10.1177/1362361315580442] [Medline: 25948599]
42. Leadbitter K, Buckle KL, Ellis C, Dekker M. Autistic self-advocacy and the neurodiversity movement: implications for autism early intervention research and practice. Front Psychol 2021 Apr 12;12:635690 [FREE Full text] [doi: 10.3389/fpsyg.2021.635690] [Medline: 33912110]
43. Alsehemi MA, Abousaadah MM, Sairafi RA, Jan MM. Public awareness of autism spectrum disorder. Neurosciences (Riyadh) 2017 Jul 13;22(3):213-215 [FREE Full text] [doi: 10.17712/nsj.2017.3.20160525] [Medline: 28678216]
44. da Silva BS, Grevet EH, Silva LC, Ramos JK, Rovaris DL, Bau CH. An overview on neurobiology and therapeutics of attention-deficit/hyperactivity disorder. Discov Ment Health 2023 Jan 05;3(1):2 [FREE Full text] [doi: 10.1007/s44192-022-00030-1] [Medline: 37861876]
45. Chen L, Jeong J, Simpkins B, Ferrara E. Exploring the behavior of users with attention-deficit/hyperactivity disorder on Twitter: comparative analysis of tweet content and user interactions. J Med Internet Res 2023 May 17;25:e43439 [FREE Full text] [doi: 10.2196/43439] [Medline: 37195757]
46. Mitchell GE, Locke KD. Lay beliefs about autism spectrum disorder among the general public and childcare providers. Autism 2015 Jul 22;19(5):553-561. [doi: 10.1177/1362361314533839] [Medline: 24852751]
47. Cortese S, Adamo N, Del Giovane C, Mohr-Jensen C, Hayes AJ, Carucci S, et al. Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. Lancet Psychiatry 2018 Sep;5(9):727-738 [FREE Full text] [doi: 10.1016/S2215-0366(18)30269-4] [Medline: 30097390]
48. Angermeyer MC, van der Auwera S, Carta MG, Schomerus G. Public attitudes towards psychiatry and psychiatric treatment at the beginning of the 21st century: a systematic review and meta-analysis of population surveys. World Psychiatry 2017 Feb 26;16(1):50-61 [FREE Full text] [doi: 10.1002/wps.20383] [Medline: 28127931]
49. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2015 Sep;17(3):327-335 [FREE Full text] [doi: 10.31887/DCNS.2015.17.3/bbandelow] [Medline: 26487813]
50. Coles ME, Coleman SL. Barriers to treatment seeking for anxiety disorders: initial data on the role of mental health literacy. Depress Anxiety 2010 Jan;27(1):63-71. [doi: 10.1002/da.20620] [Medline: 19960488]

XSL•FO
RenderX

51. Haug T, Nordgreen T, Öst LG, Havik OE. Self-help treatment of anxiety disorders: a meta-analysis and meta-regression of effects and potential moderators. Clin Psychol Rev 2012 Jul;32(5):425-445. [doi: 10.1016/j.cpr.2012.04.002] [Medline: 22681915]
52. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015 Mar 05(3):CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
53. Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Huibers MJ. How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. World Psychiatry 2016 Oct 22;15(3):245-258 [FREE Full text] [doi: 10.1002/wps.20346] [Medline: 27717254]
54. Sloan L, Morgan J, Burnap P, Williams M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. PLoS One 2015;10(3):e0115545 [FREE Full text] [doi: 10.1371/journal.pone.0115545] [Medline: 25729900]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ASD: autism spectrum disorder
BERTweet: Transformer-Based Language Model Pretrained on English Posts
BETO: Transformer-Based Language Model Pretrained on Spanish Text
EDA: Easy Data Augmentation
SUD: substance use disorder

Edited by R Cuomo; submitted 31.Mar.2025; peer-reviewed by G Mayer, KW Tay; comments to author 12.Aug.2025; revised version received 04.Feb.2026; accepted 09.Feb.2026; published 09.Apr.2026.

Please cite as:
Tiosso Batistetti V, Sanchez FG, Varaona A, Lara-Abelenda F, Pinto da Costa M, Chart-Pascual JP, Rodriguez-Quiroga A, Quintero J, Alvarez-Mon MA
Evaluating Public Sentiment on Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder Compared With Other Mental Health Disorders From Posts on X (Formerly Known as Twitter): Longitudinal Analysis
JMIR Infodemiology 2026;6:e74440
URL: https://infodemiology.jmir.org/2026/1/e74440
doi:10.2196/74440
PMID:

©Vitor Tiosso Batistetti, Facundo G Sanchez, Andrea Varaona, Francisco Lara-Abelenda, Mariana Pinto da Costa, Juan Pablo Chart-Pascual, Alberto Rodriguez-Quiroga, Javier Quintero, Miguel Angel Alvarez-Mon. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 09.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Early Depression Detection in Social Media: Monitoring of Individual Nighttime Dynamics and Large Language Model Analysis

Bicheng Yu[1], BE; Zhichang Zhang[1], PhD; Lulu Ma[1], BE; Jiongfu Cai[1], BE; Yuanyuan Zhang[1], BE

College of Artificial Intelligence and Computer Science, Northwest Normal University, Lanzhou, Gansu, China

Corresponding Author:
Zhichang Zhang, PhD
College of Artificial Intelligence and Computer Science
Northwest Normal University
967 Anning E. Road
Lanzhou, Gansu, 730070
China
Phone: 86 13038769329
Email: zzc@nwnu.edu.cn

Abstract

Background: Depression has become a major global public health challenge, and early intervention is critical for improving patient outcomes. Current depression detection techniques based on social media data (traditional risk detection) rely heavily on users' complete historical information, which cannot meet the timeliness requirements of early intervention. This underscores the need for early risk detection (ERD) methods emphasizing early-stage, real-time warning. However, existing ERD studies have notable limitations such as (1) they overlook temporal activity patterns hidden in posting time stamps, missing vital warning signals; and (2) they depend on static templates or resource-intensive sequence models, resulting in limited interpretability and inefficient use of early data, ultimately constraining their clinical applicability for early intervention.

Objective: This study aims to develop an efficient, reliable, and interpretable ERD model. The core objectives are to extract temporal activity patterns features from posting time stamps, thereby enriching feature dimensions for risk detection; to leverage large language models (LLMs) for improved text filtering precision and depression-related factor analysis; and to achieve accurate early detection of depression to support clinical intervention.

Methods: We propose the Monitoring of Individual Nighttime Dynamics (MIND) and LLM analysis model, which integrates two key innovations: (1) circadian activity dynamics: posting time stamps are transformed into temporal activity patterns, analyzing fluctuations in posting frequency and timing to derive sleep-related features, thereby compensating for the limitations of text-only approaches, and (2) LLM depression profiler: LLMs are used for dynamic text filtering, automatically removing irrelevant noise and focusing on potential depression-related cues. Based on LLM semantic understanding, latent depression risk factors are identified, enhancing interpretability for clinical treatment and robustness to noise.

Results: Experiments on the eRisk2017 (data source: Reddit [Reddit Inc]) benchmark dataset demonstrated that MIND significantly outperformed existing baseline models in early detection sensitivity, specificity, and accuracy. By combining sleep-related features with text analysis, the model achieved interpretable, traceable predictions that can support clinical treatment. ALL relevant experimental code is publicly available.

Conclusions: The MIND model combines temporal activity pattern features with LLM-based text analysis, addressing the challenges of poor interpretability and inefficient use of early-stage data in existing ERD methods. It significantly enhances early detection performance, offering a new paradigm for applying social media data in ERD, thereby enabling earlier intervention and reducing the public health burden of depression.

(JMIR Infodemiology 2026;6:e87138) doi:10.2196/87138

KEYWORDS

depression; early risk detection; sleep analysis; large language model; circadian rhythm

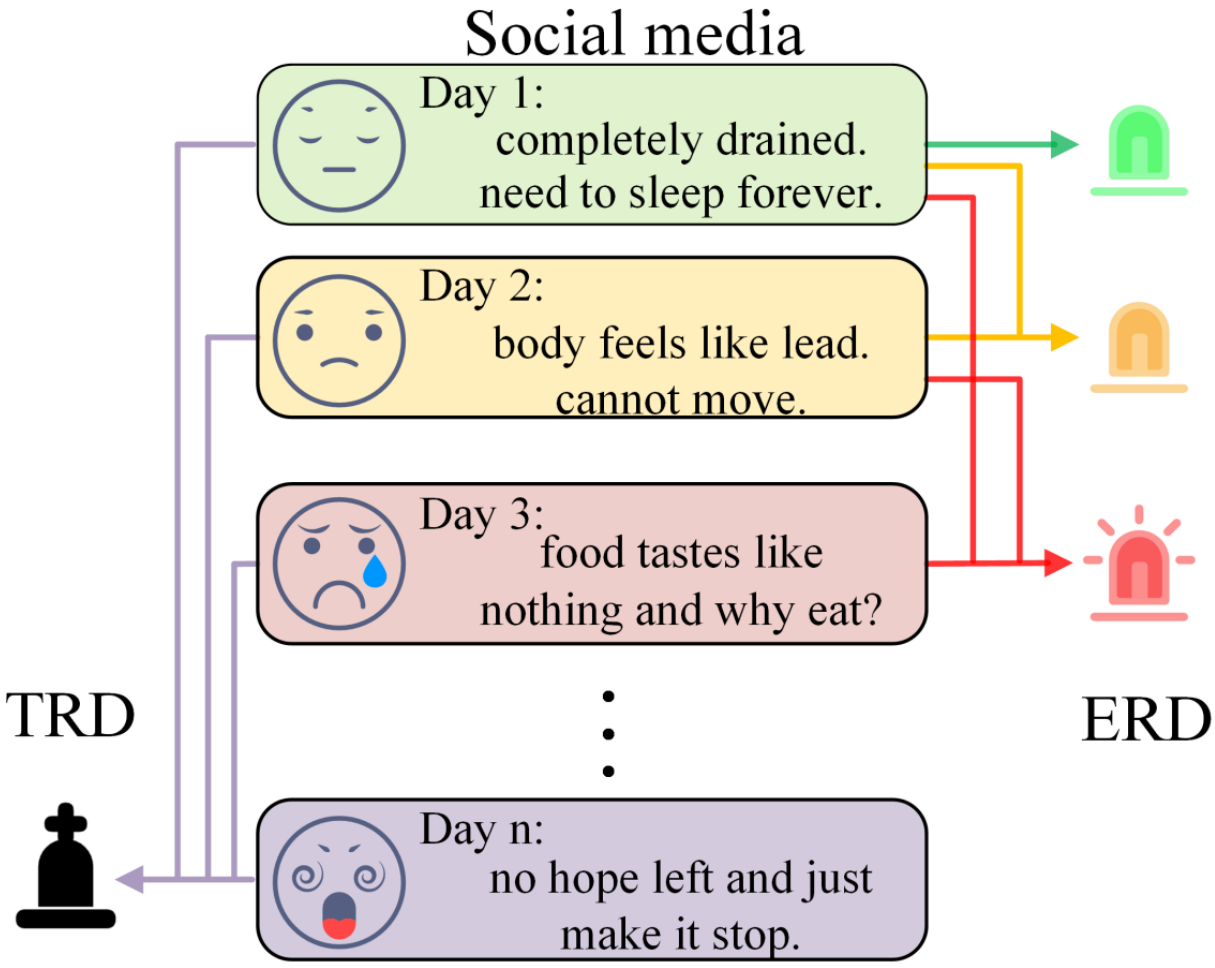
XSL•FO
RenderX

Introduction

Background

Depression affects approximately 260 million people worldwide [1-3]. This mental disorder leads to prolonged negative emotions, feelings of loss, and, in severe cases, life-threatening risks, making it a major public health concern. As social media increasingly becomes a platform for individuals to record their daily experiences and express emotions, user-generated content often reflects genuine psychological states [4]. This has enabled the academic community to explore social media data for depression detection. Over the years, the eRisk workshop has explored various contexts for identifying psychological disorders and predatory conversations [5-7], contributing valuable insights to the broader field of mental health monitoring through digital traces. Traditional risk detection (TRD) identifies depression from users' sufficient, long-term social media history and has achieved significant success. However, TRD does not align with the clinical principle of "early detection and early treatment" [8], as it relies on sufficient, long-term user data. To bridge this gap, early risk detection (ERD) has been proposed. ERD aims to predict depression risks based on limited, early-stage user data, allowing timely warnings and interventions. Unlike TRD, ERD prioritizes high precision under constrained data availability, with evaluation criteria focusing not only on final accuracy but also on the earliness of prediction, as illustrated in Figure 1.

Figure 1. Task schematic diagram. The left side illustrates the traditional risk detection (TRD) task, which predicts a user's depression status using all available user information. In contrast, the right side represents the early risk detection (ERD) task, which aims to predict a user's depression status by using early-stage information.



Previous Work

ERD research generally follows two strategies:

1. Long-sequence processing methods: recurrent neural networks [9] and their variants, along with CNNs [10] and variants of the attention mechanism [11], have been used to process full user histories [12-14]. While timestamp data has occasionally been considered, analyses have been limited to simple statistical summaries, failing to capture rhythmic features. Although comprehensive, such methods are resource-intensive and often ineffective at detecting sparse but meaningful emotional signals. Moreover, they cannot identify specific time points of psychological state changes, limiting their clinical value.
2. Filtering and denoising methods: template matching and sliding windows have been used to filter high-information texts [15,16], reducing noise complexity. However, these static methods lack adaptability to individual differences and developmental trajectories. While effective in

XSL•FO
RenderX

identifying overt depressive content, they struggle with latent signals due to limitations of earlier NLP methods, offering insufficient support for intervention.

Studies in TRD also continuously provide new insights for the advancement of ERD research. Early TRD approaches relied on text feature analysis using traditional machine learning models (eg, support vector machines and decision trees) and linguistic inquiry and word count–based features [17], but were limited by noise sensitivity and feature unimodality. Multimodal fusion of text, audio, and images [18,19] improved robustness, while more recent studies used large language models (LLMs) for data augmentation and graph structures for multimodal fusion, advancing clinical depression dialogue prediction [20]. Nonetheless, TRD methods remain constrained by their reliance on complete histories, making them unsuitable for early intervention.

In summary, current ERD methods face the following three limitations: (1) neglect of temporal information in time stamps, particularly sleep-related patterns critical to depression; (2) static, template-driven filtering with weak interpretability; and (3) resource-heavy, long-sequence methods that fail to capture sparse but crucial early cues.

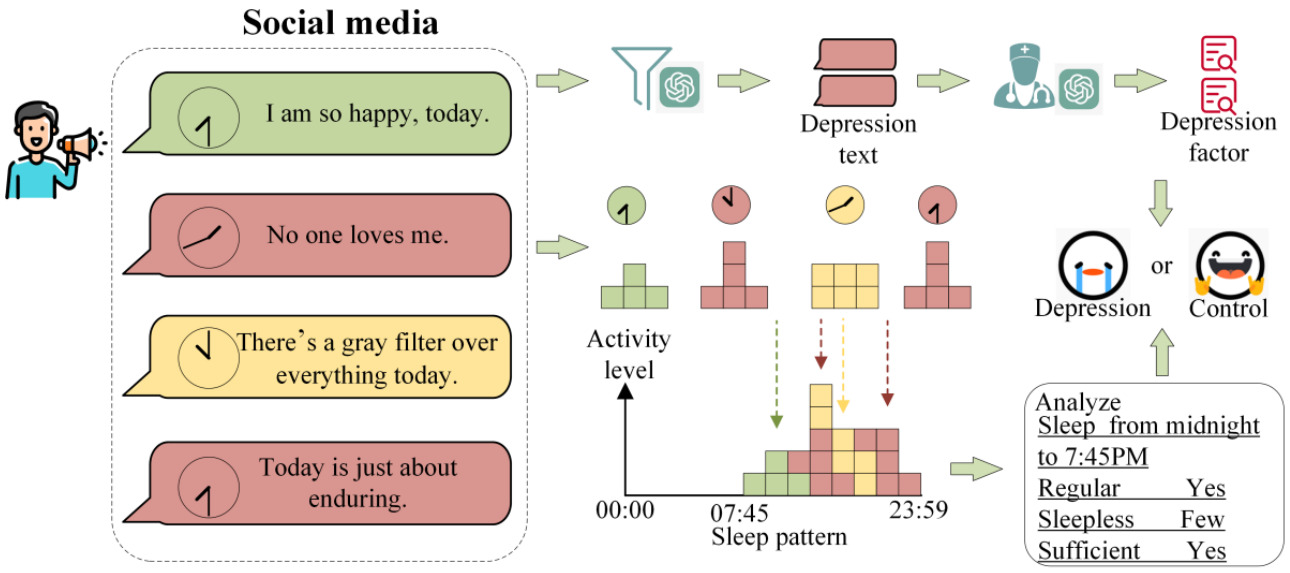
Study Objectives

To overcome these challenges, we propose the Monitoring of Individual Nighttime Dynamics (MIND) model, comprising the following three key modules:

1. Circadian activity dynamics (CAD): by modeling users' temporal activity patterns from social media posting time stamps and integrating these temporal features with textual information, this module investigates depressive tendencies through the interaction between activity frequency and sleep regularity. This module is grounded in psychological evidence demonstrating a strong association between disrupted sleep patterns and depression. [21-25].
2. LLM depression profiler (LDP): the integration of LLMs enables the automatic filtering of texts from users at high risk of depression and mitigates interference. Building on this, the model analyzes latent negative psychological factors in the text, which enhances information use efficiency and interpretability.
3. Text-sleep fusion (TSF): this module fuses multimodal features (sleep and text) for final early risk predictions.

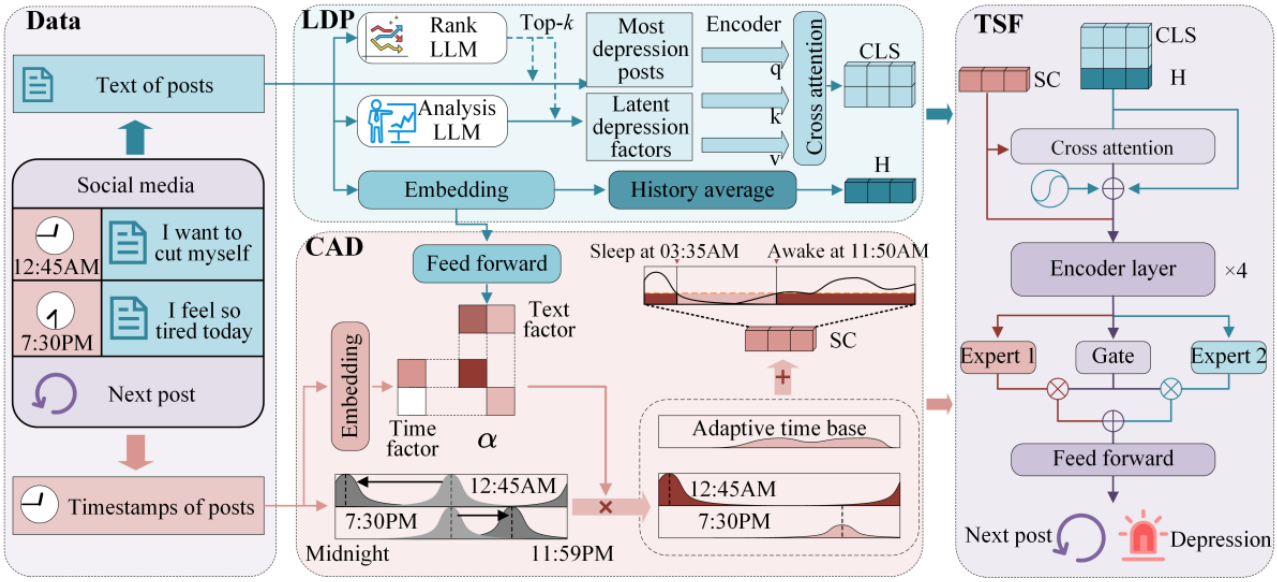
Experiments on the eRisk 2017 dataset validate the effectiveness of MIND. Our contributions include (1) introducing timestamp-based temporal activity patterns inference as an ERD feature; (2) leveraging LLM for text filtering and depression factor analysis to improve efficiency and interpretability; and (3) designing a novel multimodal fusion model (MIND) that significantly enhances early prediction performance. Figure 2 illustrates how user data are transformed into the required features, and Figure 3 provides the overall architecture of the proposed MIND model.

Figure 2. Illustration of innovation points. This figure outlines how user data are transformed into sleep-related features and depression profile features.



XSL•FO
RenderX

Figure 3. Overview of the Monitoring of Individual Nighttime Dynamics (MIND) model. H: historical representation; LDP: LLM depression profiler; LLM: large language model; SD: sleep dynamics; TSF: text-sleep fusion.



Methods

Task Definition

Let a user's collection of posts on a social media platform be denoted as $P = \{p_1, p_2, \ldots, p_L\}$, with each post associated with a time stamp, forming a corresponding set of posting time stamps $T = \{t_1, t_2, \ldots, t_L\}$. Each user is labeled according to their psychological state y_i (depressed or control).

In TRD tasks, models typically leverage all available user history to predict psychological states. By contrast, the ERD task is defined as a sequential prediction process, where posts are ordered chronologically, and each post is associated with a time step j (j = 1, 2,…, L). At each time step j (where j < L), the model processes available posts $P_i^j = \{p_1, p_2, \ldots, p_j\}$ and time stamps $T_i^j = \{t_1, t_2, \ldots, t_j\}$ to generate a prediction $\hat{y}_i^j$. This task aims to identify the user's depressive state as early as possible (ie, at the earliest possible time step $j \in (1,L)$), with only one final prediction $\hat{y}^j$ allowed. The evaluation metric emphasizes whether the model can make an accurate prediction at an earlier step rather than waiting for the complete history.

CAD Module

To capture the longitudinal behavioral patterns and temporal regularity of users, we propose a CAD module. This module transforms discrete posting time stamps into a continuous representation of activity density, serving as a robust digital proxy for the user's temporal activity patterns and circadian stability.

Activity Rhythm Initialization

We discretize the 24-hour daily cycle into M=768 time slots (same as the Bidirectional Encoder Representation from Transformers (BERT) output dimension), yielding a temporal resolution of approximately 1.875 minutes per slot. This choice of granularity is motivated by two factors: first, it ensures architectural compatibility for efficient feature fusion with BERT-derived embeddings; second, from a clinical perspective, this high-resolution discretization allows the model to capture subtle and transient fluctuations in nighttime activity patterns, which are often critical indicators of depressive behavioral shifts that coarser temporal scales might overlook. To model the natural diffusion of activity around a specific posting event, we use a Gaussian distribution assumption. For computational efficiency, rather than calculating Gaussian values for every time stamp, we implement a cyclic shift mechanism.

We first initialize a fixed base activity vector $v_{base} \in R^M$ derived from a standard Gaussian distribution ($N(\mu=0$ and $\sigma=1)$). σ is set to 1 for ease of computation and regularization, as subsequent influence factors will adjust the distribution dynamics anyway. The distribution covers the interval (–384 to 383), centered at index c_{base}=384. Here, σ is a fixed hyperparameter defining the temporal spread of activity associated with a single post. For a post occurring at minute $m_n \in (0,1440)$, we determine its target slot index as . The individual activity vector is then generated by performing a cyclic left shift on v_{base} by  and . This rotation ensures that the peak of the Gaussian distribution aligns precisely with the user's posting time on the 768-dimensional grid, regardless of the absolute time zone.

Adaptive Influence Factor Calculation

We posit that the clinical significance of a post varies depending on when it was posted and what was said. We introduce a composite weighting factor α_n to modulate the intensity of each activity vector:

(1) Hourly influence (α_n^{time}) to capture the inherent importance of specific periods, we use a learnable lookup table. The 24 hours of a day are mapped to a learnable embedding matrix $E \in R^{24\times1}$, which is initialized randomly and updated during training. For a post in an hour h_n, the factor is .

XSL•FO
RenderX

(2) Textual influence (α_n^{text}) to incorporate psychological intensity, the post text is encoded into a semantic vector $h_{text} \in R^{384}$ via the paraphrase-MiniLM-L6-v2 model. A trainable linear layer then projects this embedding to a scalar influence score $\alpha_n^{text} \in R$.

Aggregation

The final representation $SC_j \in R^{768}$ for a user at time step j is the accumulated weighted activity across all historical posts ( ). The factor (4/j) is a simple normalization term used to prevent numerical overflow during the accumulation of the SC vector and provides minimal performance gain.

By aggregating these weighted Gaussian kernels,SC forms a dense, continuous waveform. This representation effectively captures the regularity, intensity, and shifts in the user's daily rhythms. Crucially, by focusing on individual temporal consistency rather than absolute wall-clock time, this formulation remains robust to time zone variations across the global Reddit (Reddit Inc) population.

LDP Analysis

To overcome rigidity and lack of interpretability in traditional text filtering, we redesign the depression-text selection process by leveraging LLMs with few-shot and zero-shot learning paradigms [26-29]. The emotion categories used in the Rank LLM prompt were designed to conceptually align with symptom dimensions described in established depression screening instruments such as the Patient Health Questionnaire–9 (PHQ-9) and the Beck Depression Inventory. Rather than replicating questionnaire items verbatim, we abstracted common symptom constructs (eg, anhedonia, hopelessness, suicidal ideation, fatigue, appetite disturbance, and guilt) into concise semantic descriptors suitable for LLM-based text scoring. The corresponding prompt is provided in Textbox 1. Importantly, the LLM scoring is not intended as a clinical diagnosis, but as a weakly supervised semantic proxy for depression-related symptom expression in natural language. We further implemented a zero-shot large language model (LLM) prompt for psychological issue detection (Textbox 2).

Textbox 1. Prompt for a large language model (LLM).

Prompt for rank large language model (LLM):

You are a psychologist who needs to assess whether a patient's text contains depression-related emotions in order to score based on the severity of emotions. Don't explain why, just output the result.

emotion: {sad, hopeless, failure, suicide, expected to be punished, guilty, cry, lose interest, worthless, low energy, change pattern, irritable, change appetite, tired}

response: 0~5

<example>

text: Just it looks so lonely

score: 3

text: I still feel awful whenever I make a mistake

score: 5

</example>

{input text}

Prompt for analysis LLM:

You are a clinical psychologist. Read the following text and identify any potential psychological issues.

–Your reply must be very brief (one sentence or less).

–Do not explain your reasoning.

–If no issues are found, reply exactly "None."

{input text}

XSL•FO
RenderX

Textbox 2. Large language model (LLM) prompt for analysis (zero-shot).

System:

- You are a clinical psychologist.

User:

- Read the following text and identify any potential psychological issues:
- Your reply must be very brief (one sentence or less).
- Do not explain your reasoning.
- If no issues are found, reply exactly "None."

Text:

- My job is unstable. It's fine now, but what about the future?

Response:

- Potential issues with stress or anxiety related to work hours.

Large language model application programming interface (API):

- Due to regional restrictions, the experiments used a third-party API proxy. Based on communication with the provider, the default parameters were fixed at temperature = 1.0 and top_p=1.0 to ensure stability and consistency.

Specifically, the Rank LLM first assigns a depression-related score to each post $P_{i,j}^{topk}=p_{top1}, p_{top2},\ldots, p_{topk}$. Under the guidance of a small number (2) of expert-labeled examples, each post is scored from 0 to 5, and the top-*k* posts with the highest scores are selected. These selected posts are then analyzed by the Analysis LLM, which generates short free-text descriptions of potential depression-related factors. Each description represents a latent psychological issue inferred from the post (eg, hopelessness, self-harm ideation, or emotional instability). The resulting factor set is denoted as $F_{i,j}^{topk}=f_{top1}, f_{top2},\ldots, f_{topk}$.

To encode both the selected posts and the generated factors, we use a pretrained text encoder, denoted as encoder. In this work, an encoder is implemented using BERT, but the design allows replacing it with other pretrained language models to enable future improvements without changing the overall framework. While the BERT has a maximum input length of 512 tokens. To ensure stable processing, each post is truncated to the first 512 tokens before being fed into the encoder. Each text sequence is tokenized and fed into the encoder independently, which outputs contextualized token-level representations. For the factor texts, the encoder output is , where $M_F \in R^{k\times seq\times dim}$, k is the number of selected posts, L is the sequence length, and dim is the hidden dimension of the encoder. Similarly, the selected posts are encoded as an , where $M_P \in R^{k\times seq\times dim}$, k is the number of selected posts, L is the sequence length, and dim is the hidden dimension of the encoder.

These matrices are not directly reduced to fixed-length vectors. Instead, we use a cross-attention mechanism to allow the depression factors to guide the representation of the texts. Specifically, MT is used as the query, and MF is used as the key and value (), where $M \in R^{k\times seq\times dim}$, k is the number of selected posts, L is the sequence length, and dim is the hidden dimension of the encoder. W∈ trainable $R^{dim\times dim}$.

The cross-attention output is a contextualized representation that integrates both the semantic content of the posts and the depression-related factors inferred by the LLM. After cross-attention, we extract the hidden state corresponding to the CLS token as the final representation of each selected post, as in , where $CLS \in R^{k\times dim}$, k is the number of selected posts, and dim is the hidden dimension of the encoder.

At this point, we have obtained the user depression-text profile feature CLS enhanced by the LLM. Additionally, to ensure that the historical information posted by the user is not overlooked by the model, we average the embeddings of all the user's texts and use this as the coarse-grained historical text feature ( and ). Where history_∈ R^{384}, history∈ R^{384}, and W_h∈ trainable $R^{768\times384}$ represent the corresponding feature vectors and parameter matrix with specific dimensions.

Text-Sleep Fusion

First, for sleep-aware text feature enhancement, we leverage the cross-attention mechanism and use temporal activity patterns features as prior knowledge to guide the fusion of depressive text profile features and historical information features, thereby achieving the goal of targeted text feature enhancement as , , and ), where F_{text}, $F_{fusion} \in R^{k+1\times dim}$, $W_{q,k,v}$∈ trainable $R^{dim\times dim}$ represent the corresponding feature vectors and parameter matrix with specific dimensions.

Next, to preserve the order of text and distinguish between historical texts and filtered texts, we add trainable positional encoding. The resulting features are then concatenated with temporal activity patterns SC and input into the transformer encoder as  and , where $F \in R^{k+2\times dim}$ and PE∈ trainable $R^{k+2\times dim}$ represent the corresponding feature vectors and parameter matrix with specific dimensions. As global prior

XSL•FO
RenderX

knowledge, temporal activity patterns features SC are inherently different from the sequential attribute of text sequences due to their time-independence; therefore, SC is excluded from positional encoding to ensure the model can correctly distinguish between the two. After in-depth data mining by the transformer encoder, more accurate multimodal features are obtained.

To better address the information compression issue when mapping multimodal features to a single user representation, we adopt the mixture of experts (MoE) approach—a framework originally proposed by Jacobs et al [30] that has recently reemerged in LLM research. By dynamically selecting expert combinations, MoE enables more flexible feature fusion and avoids potential information loss caused by traditional pooling operations, making it well-suited for the demands of contemporary multimodal and large-scale model scenarios. Finally, the final prediction is generated through a linear layer as provided in , , , and , where e, $e_i \in R^{dim}$, $g_i \in R$, $W^e \in$ trainable $R^{dim \times dim}$, and $W^g \in$ trainable $R^{dim \times 1}$ represent the corresponding feature vectors and parameter matrix with specific dimensions. Since this is a classification task, we use the traditional cross-entropy loss function, defined as . We intentionally did not adjust the class imbalance by reweighting the loss function because the original class distribution reflects the real-world prior probability in ERD tasks. Reweighting the loss may increase false positives and harm early warning precision.

Sequential Prediction Strategy

The ERD task requires the model to make predictions as user posts arrive over time. Therefore, the model performs inference sequentially over time steps. At each time step j, the model receives all posts up to step j and generates a prediction $\hat{y} \in \{0,1\}$. The model continues predicting at each time step until a positive prediction ($\hat{y}=1$) is produced. Once the model outputs 1, the inference process stops, and the user is labeled as depressed. The time step j at which the first positive prediction occurs is recorded as the decision time, which is used to evaluate the earliness metric in ERD.

If the model never outputs 1, the final prediction at the last time step is used, and the user is labeled as control. This decision strategy follows the standard ERD evaluation protocol, where only one final decision is allowed for each user, and earlier correct predictions are rewarded.

Responsible Deployment Framework

To ensure responsible deployment, several safeguards should be maintained in practice. First, model outputs should serve only as preliminary risk signals. Any high-risk identification must be reviewed by qualified professionals or trained moderators before action is taken. Second, predictions should not trigger account suspension, content removal, or other restrictive actions without human verification. The system is intended to support assistance-oriented interventions rather than enforcement. Third, when feasible, users should be informed that automated risk monitoring systems are in place. The interpretability mechanisms provided by temporal activity patterns and LLM-generated factors may assist human reviewers in understanding the basis of risk signals. Fourth, only publicly available data are processed, and storage duration is limited. Fifth, the model should not replace professional mental health assessment; its role is limited to identifying behavioral patterns potentially associated with depression risk and recommending further evaluation where appropriate.

Ethical Considerations

The proposed MIND model is designed as a risk estimation framework rather than a diagnostic system. It does not infer clinical diagnoses and should not be used as a stand-alone decision-making tool in medical or administrative contexts. The model should not replace professional mental health assessment; its role is limited to identifying behavioral patterns potentially associated with depression risk and recommending further evaluation where appropriate. Only publicly available data are processed, and storage duration is limited. In real-world deployment, stronger anonymization and encryption mechanisms would be necessary. By explicitly distinguishing risk estimation from clinical diagnosis, the framework aims to balance early intervention benefits with ethical responsibility and user autonomy. In accordance with institutional and relevant ethical guidelines, ethical approval and informed consent were not required for this study, as only publicly available data were used and no human participants were involved. Privacy and confidentiality considerations were not applicable.

Results

Dataset

We conducted experiments on the eRisk 2017 dataset, the standard benchmark for ERD tasks [20]. The dataset contains 137 depressed users and 755 control users, divided into training (486) and testing (406) sets, collected from the Reddit platform. Depressed users were identified through self-reports in depression-related forums, with explicit confirmation of clinical diagnosis. Control users were sampled from nondepression subforums and depression subforums where users exhibited no depressive symptoms.

To ensure privacy, all user IDs were anonymized, and only text content and posting times were preserved. Table 1 summarizes the dataset statistics.

Table 1. Dataset statistics.

Characteristic	Train (depression)	Train (control)	Test (depression)	Test (control)
Users	83	403	52	349
Posts	30,851	172,834	18,706	217,665
Average posts per user	371.7	655.5	359.7	623.7

XSL•FO
RenderX

The dataset does not provide user time zone information, and all time stamps are recorded in platform time. Therefore, the proposed method does not rely on absolute clock time, but instead models the relative distribution of posting activity within each user's history. Because the representation is learned from long-term aggregated activity, it reflects stable temporal patterns rather than specific clock-based sleep schedules.

Implementation Details

The experiments were conducted via a third-party application programming interface (API) proxy service (due to regional access restrictions), which implemented a preconfigured environment for model deployment. To ensure system stability and experimental consistency within this framework, we used the default decoding parameters as confirmed by the service provider: temperature = 1.0 and top_p = 1.0. Other auxiliary parameters were maintained at their default settings to ensure the reproducibility of the scoring results under the same API environment. The encoder was implemented with BERT [31], and embeddings were generated using the frozen paraphrase-MiniLM-L6-v2 model. In the CAD module, the Gaussian parameter σ was set to 1. All attention layers adopted multihead attention with 8 heads.

In the text-sleep fusion module, the transformer encoder was configured with 4 layers. Training was performed using the Adam optimizer with a learning rate of 2e-5, a batch size of 4, and 7-14 epochs. Early stopping was applied based on validation performance, where training was terminated if the validation metric did not improve for 3 consecutive epochs. Training was conducted on NVIDIA A30 graphics processing units (GPUs) with PyTorch (version 2.3.0; Meta AI), supported by an AMD EPYC 7543 CPU and 64 GB memory. To mitigate randomness, 5 random seeds were used.

Comparison Methods

We compared MIND with multiple baselines, including official dataset baselines and subsequent studies. "Logistic regression" uses term frequency–inverse document frequency features and logistic regression, reframing ERD as multiclass classification [32]. "Feature-Rich" methods rely on feature engineering, including topic extraction using latent Dirichlet allocation and user-level statistical features [33]. "BiLSTM+Attention" models sequence information from the full user history using LSTM networks with attention mechanisms [16]. The "risk window" approach applies a sliding-window strategy to trigger warnings when depressive posts appear consecutively [16]. The "τ-SS3" model uses an enhanced n-gram-based approach for dynamic phrase sequence detection [34]. The "early list" introduces reinforcement learning-based gating to encourage early predictions [12]. "HAN-Psych" uses hierarchical attention networks inspired by clinical questionnaire structure [15]. Finally, "ESDM" balances earliness and accuracy using a loss function combining LSTM and linear attention mechanisms [13].

For fairness, all models followed their original decision-making rules. For methods without explicit rules, we adopted the default "diagnose-as-soon-as-identified" strategy.

Due to GPU memory constraints, the original implementations of ESDM and EarlyList require substantially higher computational resources. Therefore, in our experiments, these 2 methods were limited to processing the most recent 64 posts per user. In the ERD setting, later posts usually contain stronger depressive signals, so this restriction does not disadvantage the baselines and may even simplify the prediction task. We also verified that the performance obtained under this setting is consistent with the results reported in the original papers. This constraint mainly affects runtime measurement; therefore, the inference time of these methods is not reported for fair comparison. The inference time of τ-SS3 is difficult to compute and is not discussed in this paper.

Metrics

ERD introduces a specialized evaluation metric balancing "earliness" and "accuracy." This metric penalizes late predictions exponentially beyond a certain time threshold, ensuring models are rewarded for early and correct predictions, as provided in  and , where $l_{co}(j) \in [0,1]$, c_{fp}=positive/total.

In addition, we also introduced the F_1-score of the final early warning results for the ERD task (F_{1ERD}) and the F_1-score for the TRD task (F_{1TRD}) respectively, to more accurately evaluate the model's accuracy and its generalization ability on the TRD task. Overall, a superior model is characterized by smaller values in metrics $ERDE_5$ and $ERDE_{50}$, and larger values in metrics F_{1ERD} and F_{1TRD}.

Evaluation Outcomes

Tables 2 and 3 show that our proposed MIND model consistently outperforms all baseline models across all key metrics. Specifically, based on the best-performing runs reported in Table 3, MIND achieves a relative improvement of 25.4% in $ERDE_5$ and 10.7% in $ERDE_{50}$ compared to the strongest baseline. Furthermore, to address the potential bias of single-run results, we also evaluated the average performance across 5 random seeds. As shown in Table 3, the model maintains a robust advantage with mean improvements of 22% ($ERDE_5$) and 13.1% ($ERDE_{50}$), respectively. This demonstrates that the performance gain of MIND is not a result of stochastic variance but represents a systemic enhancement in early depression detection. It can be observed that two of the F_1-scores achieve only slight improvements, while the increased inference time is mainly attributed to the introduction of LLMs. We also present the Cohen *d* comparison results between our model and the 2 current state-of-the-art models in Table 4.

XSL•FO
RenderX

Table 2. Best-performing metrics across 5 random seeds.

Model	↓ERDE$_5$[a]	↓ERDE$_{50}$	↑F_{1ERD}[b]	↑F_{1TRD}[c]	↓Time cost (s)
LR[d]	13.70	8.49	40.50	60.20	4710
Feature-Rich	13.10	8.45	35.85	63.08	7558
BiLSTM+Attention	12.49	9.63	56.22	62.93	1101
Riskwindow	13.01	9.79	60.61	62.90	920
τ-SS3	13.36	8.60	49.76	54.69	—[e]
EarlyList	16.45	14.87	27.34	17.54	—
HAN-Psych	10.77	8.11	63.19	70.31	1330
ESDM	10.98	7.78	66.23	71.26	—
MIND[f]	8.03	6.94	67.17	72.72	2208

[a]ERDE: early risk detection error.
[b]ERD: early risk detection.
[c]TRD: traditional risk detection.
[d]LR: logistic regression.
[e]Not available.
[f]MIND: monitoring of individual nighttime dynamics.

Table 3. Mean (SD) of metrics across 5 random seeds.

Model	↓ERDE$_5$[a]	↓ERDE$_{50}$	↑F_{1ERD}[b]	↑F_{1TRD}[c]
LR[d]	13.70 (0.00)	8.49 (0.00)	40.50 (0.00)	60.20 (0.00)
Feature-Rich	13.35 (0.23)	8.66 (0.22)	34.68 (1.23)	62.10 (0.68)
BiLSTM+Attention	12.71 (0.40)	10.37 (0.49)	54.87 (1.30)	62.17 (0.80)
Riskwindow	13.57 (0.55)	9.89 (0.21)	58.79 (2.29)	62.36 (0.33)
τ-SS3	13.66 (0.72)	9.25 (0.53)	47.72 (1.50)	53.20 (1.18)
EarlyList	16.97 (0.86)	15.73 (0.63)	26.50 (0.91)	17.19 (0.27)
HAN-Psych	11.27 (0.56)	8.36 (0.43)	60.27 (1.80)	66.59 (2.14)
ESDM	12.37 (0.99)	8.81 (0.95)	58.81 (4.50)	66.03 (3.05)
MIND[e]	8.79 (0.67)	7.26 (0.44)	64.54 (2.80)	69.57 (3.44)

[a]ERDE: early risk detection error.
[b]ERD: early risk detection.
[c]TRD: traditional risk detection.
[d]LR: logistic regression.
[e]MIND: Monitoring of Individual Nighttime Dynamics.

Table 4. Comparison of Cohen d metrics between Monitoring of Individual Nighttime Dynamics (MIND) and SOTA baselines.

Model	ERDE$_5$[a]	ERDE$_{50}$	F_{1ERD}[b]
HAN-Psych	3.998	2.719	1.800
ESDM	4.217	2.198	1.527

[a]ERDE: early risk detection error.
[b]ERD: early risk detection.

We computed the Cohen *d* values between the state-of-the-art models and our MIND model on the ERD task based on experimental results from 5 random seed runs, so as to quantify the statistical significance of performance improvements achieved by our model.

XSL•FO
RenderX

Ablation Study

To evaluate each component's contribution, we performed ablation experiments. The "Sleep" variant removed the CAD module, relying solely on text. The "Text" removed the LLM depression profiler, using only temporal activity patterns features. The "LLM" variant replaced LLM-based filtering with the previous method, the template matching-based approach. The "History" variant removed historical text features. The "Fusion" variant removed cross-attention and MoE, using only a transformer encoder.

Table 5 presents the results of our ablation experiments. It can be observed that removing the CAD has a significant impact on the model; furthermore, the CAD alone struggles to achieve favorable performance. Replacing the LLM component in the model with existing methods exerts a substantial influence on the experimental results. Removing user historical features has a more pronounced effect on $ERDE_5$, which is more sensitive to earliness, but a relatively minor impact on $ERDE_{50}$, which has a higher tolerance for earliness. Removing the cross-attention and expert gating mechanisms only has a slight impact on the overall performance of the model.

Table 5. Results of the ablation study.

Model	↓$ERDE_5$[a]	↓$ERDE_{50}$	↑F_{1ERD}[b]
MIND[c]	0.080	0.069	0.671
Sleep	0.094	0.078	0.633
Text	0.153	0.122	0.430
LLM[d]	0.103	0.093	0.655
History	0.106	0.078	0.668
Fusion	0.084	0.071	0.669

[a]ERDE: early risk detection error.

[b]ERD: early risk detection.

[c]MIND: Monitoring of Individual Nighttime Dynamics.

[d]LLM: large language model.

Case Study

We visualize temporal activity patterns in heatmaps for depressed and control users. Control users (A) typically show near-zero activity between 6 AM and 11 AM, indicating sufficient and regular sleep. In contrast, depressed users (B) display irregular activity across all hours, often engaging in late-night posting, with loosely defined sleep windows (eg, 5 AM to noon).

Table 6 shows examples of LLM scoring and analysis. These examples highlight MIND's interpretability: high-risk posts (eg, self-harm) are prioritized, while latent depressive tendencies are also captured, providing actionable insights for clinical support.

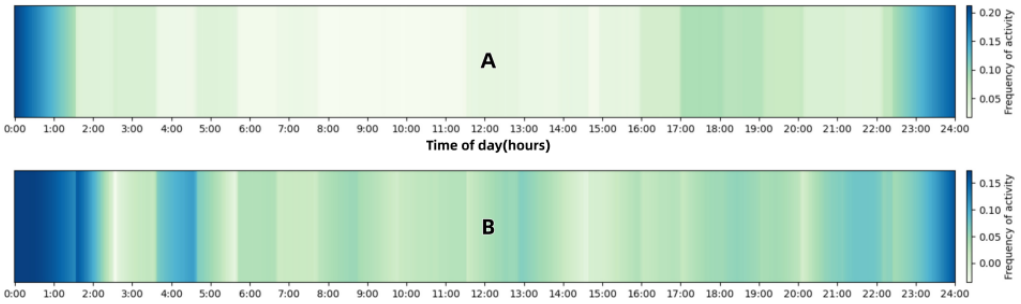
It should be noted that due to the lack of time zone information in the dataset, direct quantitative comparisons of absolute posting times (eg, mean active hour) are not statistically robust. Instead, our analysis of the heatmaps in Figures 4A and B focuses on relative activity rhythms. We specifically look for indicators of circadian rhythm disruption, such as fragmented activity during typical rest periods and the lack of a stable, continuous inactivity window. These patterns provide more reliable clinical signals for depression than absolute temporal metrics.

Table 6. Examples of large language model ranking scores and corresponding depression factor analysis.

Score	User post	Depression factor
5	Today is a bad day, maybe I should cut my wrist.	Potential self-harm ideation
2	My job is unstable. It's fine now, but what about the future.	Potential stress or anxiety related to work hours.
1	I tried to express this meaning without directly accusing OP of abusive behavior.	Potential fear of confrontation or difficulty addressing conflict directly.
0	I always thought this was why boys urinate standing up.	None

XSL•FO
RenderX

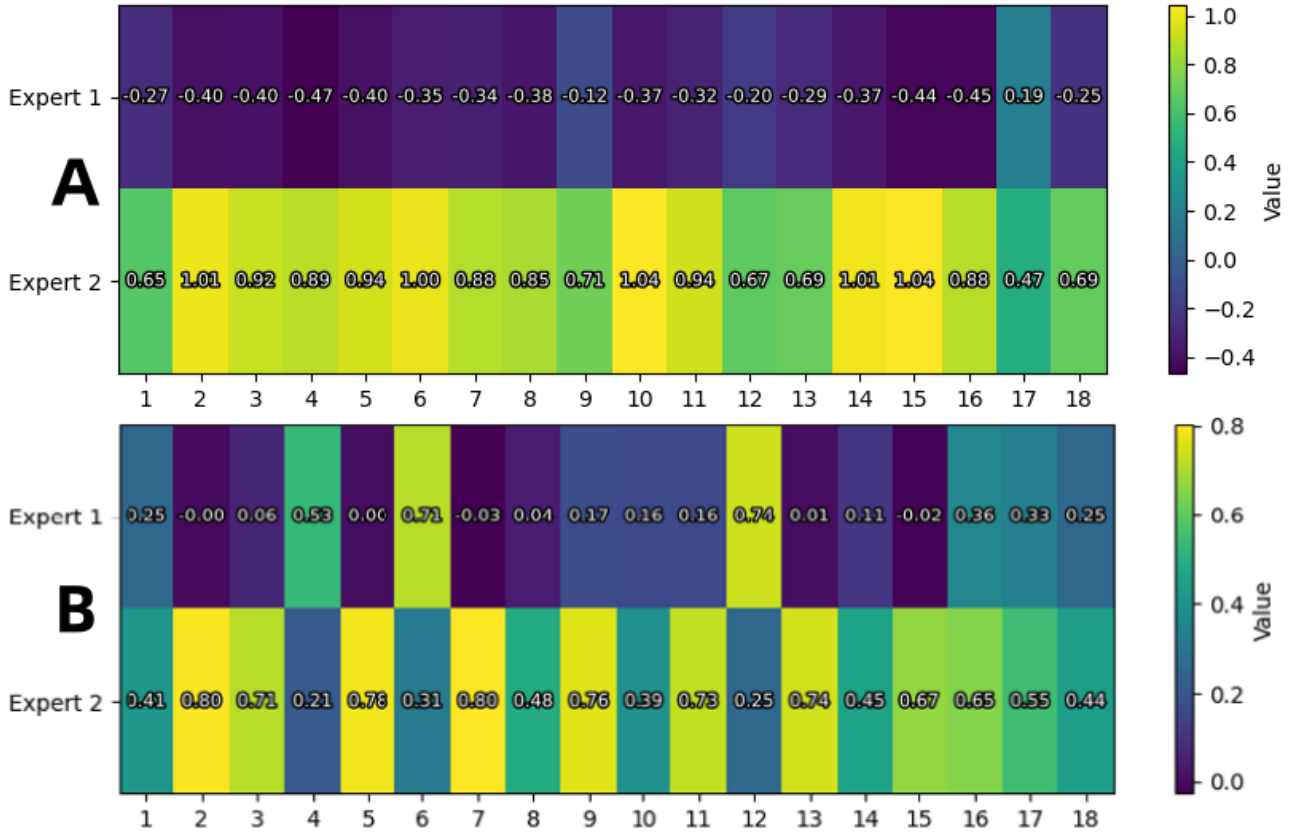
Figure 4. Visualization of temporal activity patterns in depressed vs control users.



Contribution Analysis

We visualized the contribution degree of different features to the final classification. Column 1 corresponds to the user's temporal activity patterns, Column 18 corresponds to the historical features, and Columns 2-17 correspond to the texts with the strongest depressive state among the texts posted by users. "Expert" represents the contribution degree of the expert model in Moe to the result. A higher contribution to the final discrimination result is indicated by a larger number at the corresponding position and a color that leans more toward yellow, while a lower contribution is shown by a smaller number and a color that leans more toward dark blue. As shown in Figure 5, Expert 1 is sensitive to the features of depressed users, whereas Expert 2 is more sensitive to those of control users (Figure 5A and 5B). Although temporal activity patterns and historical features have a relatively stable contribution ratio to the discrimination, the text features of users contribute the most to the final result.

Figure 5. Heatmaps of feature contribution analysis.

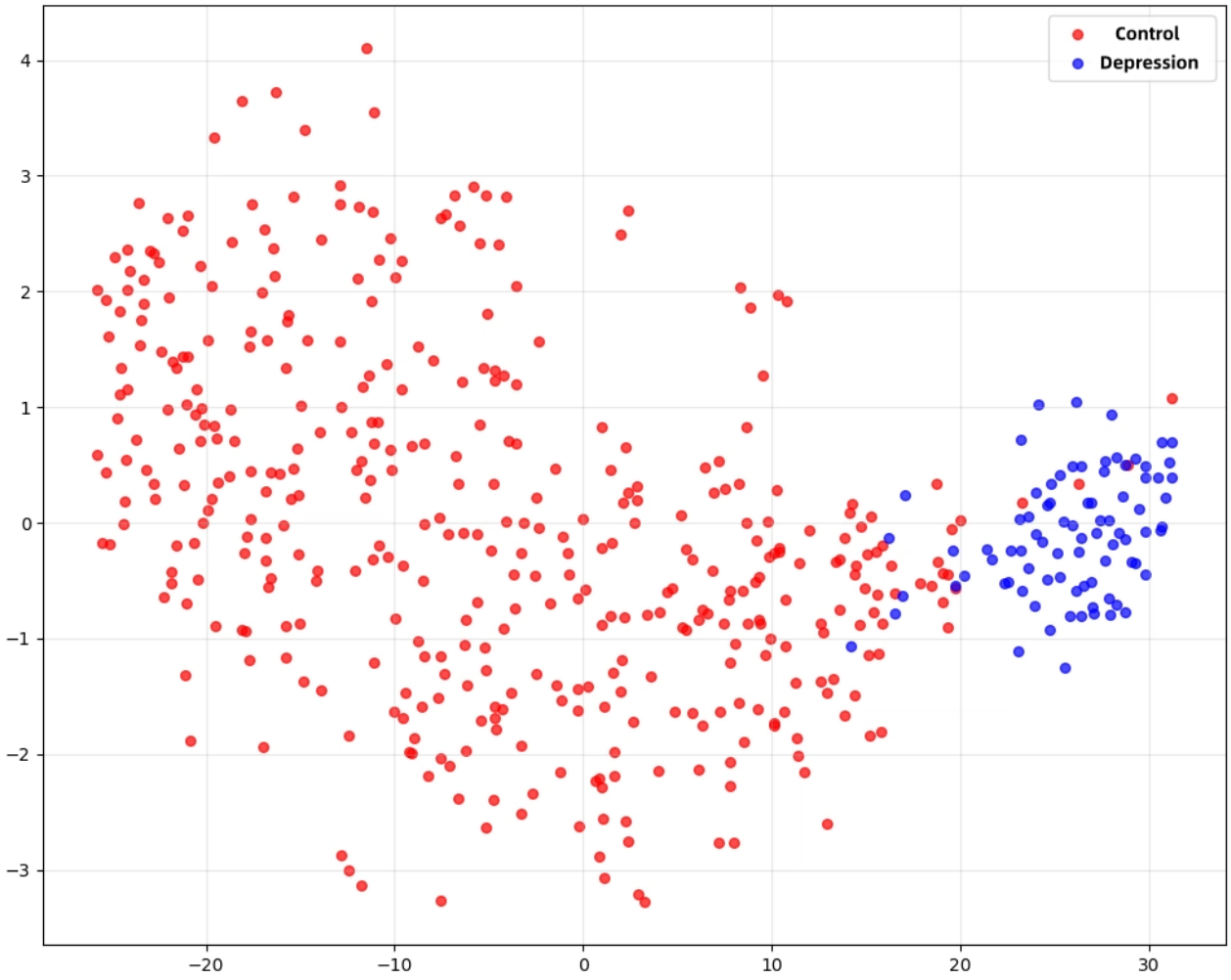


Reduction-Based Validation

This subsection presents the visualization of user features mapped in a 2D space, where the features have been processed using our proposed method. For dimensionality reduction, we adopted t-distributed stochastic neighbor embedding [35] and principal component analysis [36]. Specifically, features of depressed users are displayed in blue, while those of nondepressed users are shown in red. As illustrated in Figure 6, the features of depressed users are clustered in a distinct region. Meanwhile, the features of most nondepressed users are far apart from those of depressed users; in contrast, the difference between the features of a small number of nondepressed users and those of depressed users is not significant.

XSL•FO
RenderX

Figure 6. Dimensionality reduction visualization of the distinction between depressed and control users.

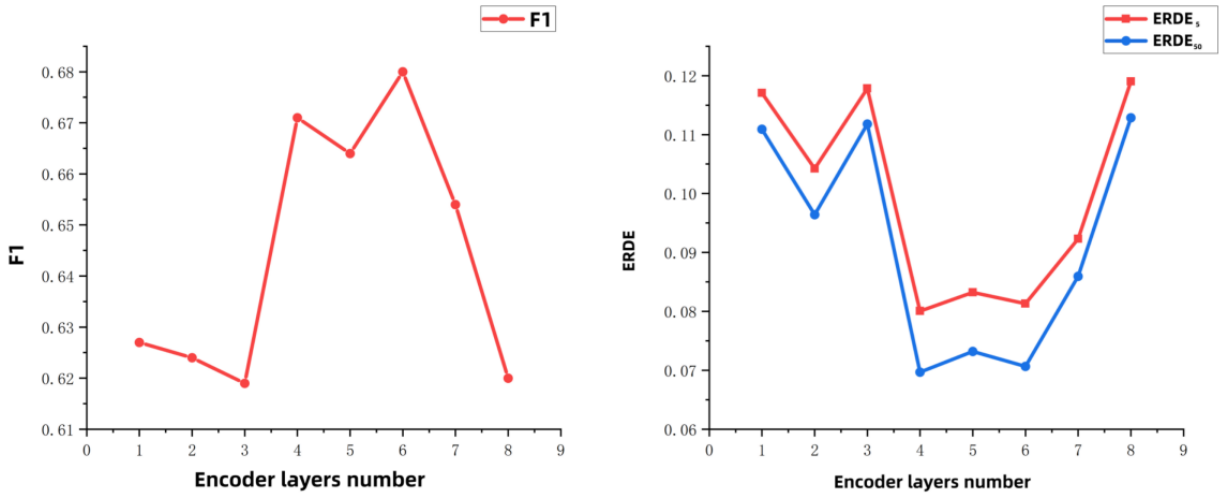


Hyperparameter Analysis

The most computationally intensive component of the entire model is the encoder layer. To strike a balance between computational efficiency and accuracy, we conducted experiments on the selection of the number of encoder layers. As shown in Figure 7, performance changes of the model in terms of F_1-score and early risk detection error as parameters vary. When the number of encoder layers is between 3 and 5, the early risk detection error values reach their minimum, while the F_1-score also peaks in a similar range. Within this range, increasing the number of layers yields limited improvement in model performance; thus, we selected 4 layers, which offer the most efficient computation speed.

XSL•FO
RenderX

Figure 7. Performance changes of the model in terms of F1-score and early risk detection error (ERDE) as model parameters vary.



Discussion

Overall Analysis

The previous section presented the performance of MIND on the eRisk 2017 dataset, along with results from ablation studies, case analyses, contribution visualization, and reduction-based validation. Overall, our findings demonstrate that the proposed approach consistently outperforms existing methods across evaluation metrics, except for higher inference-time costs. In this section, we provide a detailed analysis of each module's strengths and limitations.

CAD Analysis

In the ablation experiments, removing the CAD module resulted in a 17.5% increase in $ERDE_5$ and a 13% increase in $ERDE_{50}$, while F_{1ERD} decreased by 5.6%. These results indicate that CAD's primary contribution lies in enhancing earliness, with a relatively smaller impact on accuracy.

Although the introduction of temporal activity patterns modeling provides a novel modality for ERD, the CAD module alone yields limited performance, confirming that text features remain the dominant modality, while CAD serves as an auxiliary enhancement. This observation aligns with the contribution analysis, where temporal activity patterns provided stable contributions but were not the most decisive factor compared to the text feature.

Case studies showed that temporal activity patterns clearly distinguished depressed from nondepressed users. Depressed users exhibited irregular, insufficient temporal activity patterns, while nondepressed users displayed consistent sleep cycles. Such visualization not only aids classification but also offers clinicians interpretable evidence for interventions. However, the approach has limitations: since temporal activity patterns are inferred from aggregated posting time stamps, the model only captures coarse-grained, long-term sleep habits. While effective for users with persistent depressive tendencies, it is less sensitive to short-term depressive episodes, underscoring the need for finer-grained integration with contextual textual analysis.

Clinical Alignment With Depression-Related Sleep Disturbances

Extensive clinical research has established sleep disturbance as a core feature of major depressive disorder. Patients with depression frequently exhibit insomnia, early morning awakening, sleep fragmentation, circadian rhythm instability, shortened rapid eye movement latency, and altered sleep architecture [21-24].

Although the proposed CAD module does not measure physiological sleep architecture, the extracted temporal activity patterns may reflect behavioral correlates of these clinically observed abnormalities. For example, early morning awakening may correspond to premature termination of low-activity intervals in posting behavior; sleep fragmentation may manifest as increased nighttime posting density; and circadian rhythm instability may be reflected by higher variance in inferred rest-activity windows across days.

Therefore, the CAD representation should be interpreted as a digital behavioral proxy of circadian irregularity rather than a direct physiological sleep measurement. This alignment strengthens the clinical interpretability of the model while remaining within the scope of computational social media analysis.

LDP Analysis

Replacing the LDP module with template matching in ablation studies caused the $ERDE_5$ to increase by 28.7% and $ERDE_{50}$ by 34.7%, with a 2.3% decrease in F_{1ERD}. This demonstrates that the LLM component substantially improves early detection capability, although gains in accuracy are more modest.

One of the most significant benefits of LDP is its improvement in interpretability. Unlike static filtering, the LLM not only identifies depressive posts but also traces underlying risk factors (eg, hopelessness and suicidal ideation). This provides actionable insights into users' depressive triggers and key posts.

XSL•FO
RenderX

However, performance was found to be sensitive to the choice of base model. During experiments, different LLMs yielded varying results, with GPT-4o (OpenAI) showing the most balanced performance. This suggests that model selection and optimization remain critical directions for future research.

Moreover, reliance on external APIs and longer inference time limits scalability. Deploying locally optimized LLM could address both efficiency and privacy concerns. Historical features, while less impactful overall, contributed significantly at earlier time steps, indicating their importance for very early detection.

TSF Analysis

The TSF module was designed to align and integrate multimodal features. Ablation experiments confirmed its role: removing TSF increased the ERD metric by 9.16% and earliness by 7.2%, although F_1 remained stable (fluctuations <1%).

This demonstrates that TSF consistently enhances early warning performance while maintaining accuracy. Its lightweight fusion process effectively balances multimodal collaboration without incurring significant overhead.

Further analysis revealed that when only text features were available, TSF offered no clear advantage over a simple transformer encoder. This suggests that its benefit primarily arises from the cross-modal alignment between text and temporal activity patterns. Improving TSF's generalization ability to text-only scenarios is a potential avenue for future improvement.

Computational Cost of the LDP

As provided in Table 3, our model exhibits a relatively noticeable increase in time consumption, with an additional 878 seconds compared to the existing state-of-the-art models. The LDP introduces additional computational overhead due to repeated API calls. For each user, the Rank LLM evaluates all posts, and the Analysis LLM processes the top-*k* posts (k=16). In the eRisk2017 dataset, the average number of posts per user is approximately 419. Thus, each user requires about 419 Rank LLM calls and 16 Analysis LLM calls. The experiments were conducted using an online LLM API service. Because the API server is located overseas, all requests must be sent through a virtual private network proxy, which introduces significant network latency. In addition, the experimental environment processes requests sequentially, without multithreading or batching. Under this setting, the average time cost for processing a single post (including network communication and API response) is approximately 1.7-2.5 seconds. It should be noted that most of the time, overhead comes from network communication rather than model inference itself. In practical deployment, API calls can be parallelized and executed without virtual private network routing, so the actual runtime can be significantly lower than in the experimental environment.

Deployment and Generalization

The core method of this paper, CAD, aims to analyze users' sleep patterns through the timing of their posts, enabling earlier warning of users' depressive states. In practical deployment, platforms can obtain more detailed user information—not only time stamps of users' posting behavior but also those of their reading behavior. This allows the model to generate more accurate temporal activity patterns, thereby improving performance.

The proposed method should achieve favorable results on most social media platforms. However, it is difficult to generalize to platforms focusing primarily on visual content, such as Instagram (Meta Platforms). This is because the method takes textual information as the primary input and temporal activity patterns as an auxiliary feature to predict user depression.

Regarding generalization to other mental disorders, the method proposed in this paper cannot be directly applied. This is due to the lack of authoritative and widely accepted evidence linking other mental illnesses to sleep patterns, which requires further research.

Because the LLM component operates via explicit prompt instructions rather than model-specific fine-tuning, the framework is model-agnostic. Future researchers can substitute GPT-4o with other LLMs by preserving the prompt structure and scoring scale, facilitating adaptation as LLM technologies evolve.

Limitations

Due to GPU memory limitations, 2 baselines (ESDM and EarlyList) were evaluated using only the most recent 64 posts per user. Although this setting follows the practical constraints of the original implementations and does not negatively affect performance in the ERD scenario (where later posts usually contain stronger signals), it may still introduce minor differences compared with fully unrestricted evaluation. Future work will include experiments under larger computational resources to further verify fairness across all baselines.

The proposed CAD module models long-term temporal activity patterns derived from posting time stamps rather than actual sleep behavior. Reduced posting activity does not necessarily indicate sleep, as it may also result from work schedules, social engagement, travel, or decreased platform usage. Therefore, the extracted features should be interpreted as proxies of rest–activity regularity rather than direct measurements of sleep-wake rhythms.

Conclusions

In this study, we proposed a novel method for early risk detection of depression on social media, integrating temporal activity patterns modeling and LLM-based text filtering and analysis. Unlike prior work, which exclusively relies on text features for depressive state analysis, our approach incorporates posting time stamps to infer users' temporal activity patterns and leverages LLM to improve upon traditional template-matching methods, thereby constructing a more accurate profile of depressive texts.

Experiments on the eRisk 2017 benchmark demonstrate that MIND not only achieves superior performance across multiple evaluation metrics, despite higher inference costs, but also provides enhanced interpretability (via temporal activity patterns visualization and LLM-based tracing of latent depressive factors). The temporal activity patterns and the LLM's identification of latent depressive factors allow the model to

XSL•FO
RenderX

trace back depressive risk signals, offering strong support for clinical intervention and treatment.

Overall, MIND establishes a new paradigm for multimodal ERD by combining temporal activity patterns and advanced language modeling, breaking the limitation of prior text-only methods—with high practical value for real-world applications in mental health monitoring.

Acknowledgments

We acknowledge the support of artificial intelligence tools in completing this paper. Within the research methodology, large language models were adopted as a functional component to facilitate the study's technical implementation. For the paper text, GPT (OpenAI) was used to provide partial linguistic polishing assistance. Notably, the author retains full responsibility for the paper's academic integrity, including all research content, analytical processes, and conclusions, which have undergone independent validation.

Data Availability

Our research on depression risk may raise ethical concerns. The data used in this study come from a publicly available dataset shared by other researchers. To protect user privacy, all social media data underwent a strict anonymization process conducted by the dataset providers before usage. We followed all relevant ethical guidelines and legal requirements to ensure that no privacy violations occurred during the research process. The proposed model is not intended as a diagnostic tool but rather as a risk estimation framework to support monitoring and early intervention. All textual examples shown in this paper were paraphrased by the authors to remove sensitive information and do not represent original user posts.

Funding

This study is supported by the National Natural Science Foundation of China (62163033), Talent Innovation and Entrepreneurship Project of Lanzhou, China (2021-RC-49), and the Gansu Provincial Department of Education: Industrial Support Program Projects (2025CYZC-010). The funding bodies had no role in study design, data collection and analysis, decision to publish, or preparation of the paper.

Authors' Contributions

The author independently designed the research, developed the methodology, conducted experiments, analyzed data, and drafted the entire paper. The corresponding author provided critical guidance on research design, paper structure, and result interpretation. A fellow student offered auxiliary support, such as partial data collection and preprocessing.

Conflicts of Interest

None declared.

References

1. YuBicheng/MIND. GitHub. URL: https://github.com/YuBicheng/MIND [accessed 2026-05-01]
2. GBD 2017 DiseaseInjury IncidencePrevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1789-1858. [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
3. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study. PLoS Med 2013;10(11):e1001547. [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]
4. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Presented at: International AAAI Conference on Web and Social Media; July 8-11, 2013; Cambridge, MA URL: https://www.researchgate.net/publication/259948193_Predicting_Depression_via_Social_Media [doi: 10.1609/icwsm.v7i1.14432]
5. Parapar J, Martin-Rodilla P, Losada DF, Crestani F. Overview of eRisk 2022: early risk prediction on the internet. 2022 Presented at: Proceedings of the International Conference on Runtime Verification (RV 2022); October 2-5, 2022; Tbilisi, Georgia p. 233-256. [doi: 10.1007/978-3-031-13643-6_18]
6. Parapar J, Martin-Rodilla P, Losada DF, Crestani F. Overview of eRisk 2023: early risk prediction on the internet. 2023 Presented at: Proceedings of the IFIP International Conference on Entertainment Computing (ICEC 2023); September 27-29, 2023; Dundee, United Kingdom p. 294-315. [doi: 10.1007/978-3-031-42448-9_22]
7. Parapar J, Martin-Rodilla P, Losada DF, Crestani F. Overview of eRisk 2024: early risk prediction on the internet. 2024 Presented at: Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024); November 17-19, 2024; Porto, Portugal p. 73-92. [doi: 10.1007/978-3-031-71908-0_4]

XSL•FO
RenderX

8. Marco JH, Alonso S, Andani J. Early intervention with cognitive behavioral therapy reduces sick leave duration in people with adjustment, anxiety and depressive disorders. J Ment Health 2020;29(3):247-255. [doi: 10.1080/09638237.2018.1521937] [Medline: 30322314]
9. Elman JL. Finding structure in time. Cognitive Science 1990;14(2):179-211. [doi: 10.1016/0364-0213(90)90002-e]
10. Yoon K. Convolutional neural networks for sentence classification. 2024 Presented at: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October 25-29, 2014; Doha, Qatar p. 1746-1751. [doi: 10.3115/v1/d14-1181]
11. Zhang E, Masoud N, Bandegi M, Lull J, Malhan R. Step attention: sequential pedestrian trajectory prediction. 2022 Presented at: 2022 IEEE International Conference on Data Science and Information System (ICDSIS 2022); July 21-23, 2022; Hassan, Karnataka, India p. 8071-8083. [doi: 10.1109/jsen.2022.3158271]
12. Hartvigsen T, Sen C, Kong X, Rundensteiner E. Adaptive-halting policy network for early classification. 2019 Presented at: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; August 4-8, 2019; Anchorage, AK p. 101-110. [doi: 10.1145/3292500.3330974]
13. Bichen W, Yuzhe Z, Yanyan Z, Pengfei D, Bing Q. ESDM: Early sensing depression model in social media streams. 2024 Presented at: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024); May 22, 2024; Torino, Italy. [doi: 10.63317/2yha3uabx9j4]
14. Trotzek M, Koitka S, Friedrich CM. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. IEEE Trans Knowl Data Eng 2020;32(3):588-601. [doi: 10.1109/tkde.2018.2885515]
15. Zhang Z, Chen S, Wu M, Zhu KQ. Psychiatric scale guided risky post screening for early detection of depression. 2022 Presented at: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence; July 23-29, 2022; Vienna, Austria p. 5220-5226. [doi: 10.24963/ijcai.2022/725]
16. Sadeque F, Xu D, Bethard S. Measuring the latency of depression detection in social media. 2018 Presented at: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining; February 5-9, 2018; Marina Del Rey, CA p. 495-503. [doi: 10.1145/3159652.3159725]
17. Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. Health Inf Sci Syst 2018;6(1):8 [FREE Full text] [doi: 10.1007/s13755-018-0046-0] [Medline: 30186594]
18. Yang L, Jiang D, He L, Pei E, Oveneke MC, Sahli H. Decision tree based depression classification from audio video and language information. 2016 Presented at: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge; October 15-19, 2016; Amsterdam, The Netherlands p. 89-96. [doi: 10.1145/2988257.2988269]
19. An A, Wang J, Li S, Zhou G. Multimodal topic-enriched auxiliary learning for depression detection. 2020 Presented at: Proceedings of the 28th International Conference on Computational Linguistics; Dec 8, 2020; Barcelona, Spain p. 1078-1089. [doi: 10.18653/v1/2020.coling-main.94]
20. Chen Z, Deng J, Zhou J, Wu J, Qian T, Huang M. Depression detection in clinical interviews with llm-empowered structural element graph. 2024 Presented at: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 16-21, 2024; Mexico City, Mexico p. 8181-8194. [doi: 10.18653/v1/2024.naacl-long.452]
21. Nutt D, Wilson S, Paterson L. Sleep disorders as core symptoms of depression. Dialogues Clin Neurosci 2022;10(3):329-336. [doi: 10.31887/dcns.2008.10.3/dnutt]
22. Franzen PL, Buysse DJ. Sleep disturbances and depression: risk relationships for subsequent depression and therapeutic implications. Dialogues Clin Neurosci 2022;10(4):473-481. [doi: 10.31887/dcns.2008.10.4/plfranzen]
23. Riemann D, Berger M, Voderholzer U. Sleep and depression--results from psychobiological studies: an overview. Biol Psychol 2001;57(1-3):67-103. [doi: 10.1016/s0301-0511(01)00090-4] [Medline: 11454435]
24. Yasugaki S, Okamura H, Kaneko A, Hayashi Y. Bidirectional relationship between sleep and depression. Neurosci Res 2025 Feb;211:57-64 [FREE Full text] [doi: 10.1016/j.neures.2023.04.006] [Medline: 37116584]
25. Akre S, Cohen ZD, Welborn A, Zbozinek TD, Balliu B, Craske MG, et al. Comparing self reported and physiological sleep quality from consumer devices to depression and neurocognitive performance. NPJ Digit Med 2025;8(1):92 [FREE Full text] [doi: 10.1038/s41746-025-01493-6] [Medline: 39924572]
26. Fei-Fei L, Fergus R, Perona P. A bayesian approach to unsupervised one-shot learning of object categories. 2003 Presented at: Proceedings Ninth IEEE International Conference on Computer Vision (ICCV); October 13-16, 2003; Nice, France p. 1134-1141. [doi: 10.1109/iccv.2003.1238476]
27. Palatucci M, Hinton GE, Pomerleau D, Mitchell TM. Zero-shot learning with semantic output codes. 2009 Presented at: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS 2009); December 7-12, 2009; Vancouver, BC p. 1410-1418. [doi: 10.5555/2984093.2984252]
28. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. Cambridge: MIT Press; 2020 Presented at: Proceedings of the 34th International Conference on Neural Information Processing Systems; December 6-12, 2020; Vancouver, BC p. 1877-1901. [doi: 10.5555/3495724.3495883]

XSL•FO
RenderX

29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. 2017 Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17); December 4-9, 2017; Long Beach, CA p. 6000-6010. [doi: 10.65215/ctdc8e75]
30. Jacobs RA, Jordan MI, Nowlan SJ, Hinton GE. Adaptive mixtures of local experts. Neural Comput 1991;3(1):79-87. [doi: 10.1162/neco.1991.3.1.79] [Medline: 31141872]
31. Jacob D, Ming-Wei C, Kenton L, Kristina T. BERT: Pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/N19-1423]
32. Losada DE, Crestani F. A test collection for research on depression and language use. 2016 Presented at: Proceedings of the 7th International Conference of the CLEF Association (CLEF 2016); September 5-8, 2016; Évora, Portugal p. 28-39. [doi: 10.1007/978-3-319-44564-9_3]
33. Koltcov S, Koltsova O, Nikolenko S. Latent dirichlet allocation: stability and applications to studies of user-generated content. 2014 Presented at: Proceedings of the 2014 ACM conference on Web science; June 23-26, 2014; Bloomington, IN. [doi: 10.1145/2615569.2615680]
34. Burdisso SG, Errecalde M, Montes-y-Gómez M. τ-SS3: a text classifier with dynamic n-grams for early risk detection over text streams. Pattern Recognit Lett 2020;138:130-137. [doi: 10.1016/j.patrec.2020.07.001]
35. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9(86):2579-2605 [FREE Full text]
36. Felipe LG, Gustavo RF, Henrique FDA, Filipi NS, Cesar HC, Diego RA, et al. Principal component analysis: a natural approach to data exploration. ACM Comput 2022:34. [doi: 10.1145/3447755]

Abbreviations

BDI: Beck Depression Inventory
BERT: Bidirectional Encoder Representation from Transformers
CAD: circadian activity dynamics
ERD: early risk detection
GPU: graphics processing unit
LDP: LLM depression profiler
LLM: large language model
MIND: Monitoring of Individual Nighttime Dynamics
MoE: mixture of experts
PHQ-9: Patient Health Questionnaire–9
TRD: traditional risk detection,
TSF: text-sleep fusion

Edited by W Ahmed; submitted 05.Nov.2025; peer-reviewed by B Patel, S Sazzed; comments to author 29.Jan.2026; revised version received 30.Mar.2026; accepted 29.Apr.2026; published 29.May.2026.

Please cite as:
Yu B, Zhang Z, Ma L, Cai J, Zhang Y
Early Depression Detection in Social Media: Monitoring of Individual Nighttime Dynamics and Large Language Model Analysis
JMIR Infodemiology 2026;6:e87138
URL: https://infodemiology.jmir.org/2026/1/e87138
doi:10.2196/87138
PMID:

©Bicheng Yu, Zhichang Zhang, Lulu Ma, Jiongfu Cai, Yuanyuan Zhang. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 29.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Applying the Vertically Integrated Project Model to Youth-Inclusive Social Media Coding of Zyn-Related YouTube Shorts: Methodological Case Study

Julia (Pengyue) Dou[1], BA, MPH; Beth L Hoffman[1], MPH, PhD; Piper Narendorf[1], BS; Tatiana Grinberg Limoncic[1], BA; Grace Carver[2]; Natalia Connor[2]; Christine Larkin[1], BS; Jaime E Sidani[1], MPH, PhD

[1]Department of Behavioral and Community Health Sciences, School of Public Health, University of Pittsburgh, Pittsburgh, PA, United States
[2]MYTH (Misinformation-identifying Youth in Tobacco and Health) Youth Collaborative, Pittsburgh, PA, United States

Corresponding Author:
Julia (Pengyue) Dou, BA, MPH
Department of Behavioral and Community Health Sciences
School of Public Health
University of Pittsburgh
Public Health Bldg, 6th Fl
130 De Soto St
Pittsburgh, PA, 15261
United States
Phone: 1 9179072438
Email: PED44@pitt.edu

Abstract

Background: The rapid evolution of the nicotine and tobacco product (NTP) marketplace continues to outpace existing research. Although adolescents are active participants on social media who both consume and generate content, this age group has rarely been included in analyses of NTP content, raising a methodological challenge for accurate interpretation and coding of youth-oriented NTP content. Most research examining oral nicotine product content online has focused on TikTok, despite YouTube's widespread adolescent use. One promising framework is the vertically integrated project (VIP) model. Situating content analysis within the VIP model suggests the methodological potential of age-diverse teams to capture nuances in social media data coding, reveal areas of interpretive divergence, and ultimately advance research methodologies in public health.

Objective: This study aimed to understand how perspectives across different age groups, academic levels, and lived experiences shape the interpretation of Zyn-related YouTube shorts by engaging high school (HS) students, undergraduate public health students, and faculty researchers in parallel coding.

Methods: We manually collected 300 publicly available YouTube shorts containing #Zyn on September 18, 2024, as a methodological case example. Three coder pairs, including 2 HS students (GC and NC), 2 undergraduate students (PN and TGL), and 2 public health faculty members (BLH and JES), independently coded the same 50 videos per round across 6 iterative rounds (N=300). For each round, interrater reliability was assessed using Cohen κ and percent agreement as complementary indicators. After all coding rounds were completed, the full coding team participated in a debriefing session. The transcript was analyzed using a combined deductive-inductive thematic approach to contextualize reliability patterns and identify challenges and practical insights related to age-diverse collaborative coding within the VIP-informed model.

Results: Interrater reliability varied across 6 iterative rounds, with the undergraduate team showing the highest κ values in the final coding round (κ range: 0.40-0.63). Several constructs showed consistently high percent agreement but low and variable κ values, particularly within the HS team. For individual codes, mean κ values for male stereotype were higher for the undergraduate and faculty teams than for the HS team (0.56, 0.47, and 0.10, respectively). Debriefing findings indicated that disagreement reflected factors such as evolving code definitions, interpretive drift across rounds, and differences in cultural and experiential lenses. Coders noted that the codebook struggled to capture ambiguity in short-form content and that teams developed distinct heuristics for resolving uncertainty. Suggested improvements included rotating coding pairs and expanding codebook examples.

Conclusions: This methodological case study illustrates how a VIP-informed, age-diverse coding structure can be used to examine interpretive challenges in coding youth-oriented nicotine-related short-form video content. For public health studies

XSL•FO
RenderX

examining how different groups interpret online health messages, the VIP model may offer a practical structure for incorporating those perspectives into the research process.

(JMIR Infodemiology 2026;6:e95363) doi:10.2196/95363

KEYWORDS

collaborative coding; youth-inclusive research; social media content analysis; interrater reliability; vertically integrated project; VIP; nicotine and tobacco products

Introduction

Social Media and Youth Nicotine and Tobacco Product Use

Prior research examining nicotine and tobacco product (NTP)–related social media exposure and e-cigarette use suggests that such exposure may shape youth perceptions and behaviors. A 2025 scoping review of 30 studies reported that most identified significant positive associations between NTP content exposure and e-cigarette use across diverse demographic groups [1].

The rapid evolution of the NTP marketplace continues to outpace existing research. While combustible cigarette use has declined, youth now encounter a broad array of alternatives, including oral nicotine products (ONPs), such as Zyn [2]. Unit sales of ONPs increased nearly 300-fold between 2016 and 2020 [3]. These tobacco-free, discreet pouches are available in youth-appealing flavors and are often marketed as “cleaner” alternatives [4]. Against this backdrop of increasing digital exposure and regulatory shifts following restrictions on flavored e-cigarettes [5], ONP manufacturers expanded their presence on digital platforms [6-8]. Such shifts raise concerns that nicotine use may be reintroduced to nicotine-naive adolescents through less overt, more culturally embedded pathways.

Despite a 2025 Pew Research Center survey indicating that more than 90% of teens aged 13-17 years use YouTube compared to 68% who say they use TikTok, to date, most research examining ONP content online has focused on TikTok [6,9,10]. The research has found that ONP manufacturers expanded their presence on digital platforms, often relying on youth-oriented cues such as lifestyle framing, humor, social appeal, and the minimization of health risks [6-10]. This is a notable gap in the literature, as about 96% of participants in a 2023 survey who reported current NTP use used YouTube, and 40% had seen YouTube content promoting NTP use [11]. Additionally, a previous qualitative study also found that YouTube videos were a main source of information about modifying e-cigarette liquids and devices [12]. Therefore, it would be valuable for researchers to extend the current literature examining ONP content on TikTok to YouTube.

Moreover, despite adolescents being active participants on social media who both consume and generate large volumes of content [13], to the best of our knowledge, this age group has yet to be included in analyses of NTP content on social media. Previous research has indicated that adolescents and adult researchers interpret health content on television differently [14], and a 2022 study that engaged young adults in the analysis of TikTok videos related to nicotine addiction reported differences in content analysis between young adults and the academic adult researchers [9]. These findings point to a potential methodological challenge regarding accurately interpreting and coding youth-oriented NTP social media content and suggest that when coding decisions rely solely on adult researchers’ perspectives or automated approaches, there is a substantial risk of misinterpreting the nuanced meaning of the content. Specifically, although the authors of the above-mentioned studies analyzing ONP content on TikTok noted that humorous and youth-appealing portrayals of Zyn were prevalent and could contribute to harm-minimization and normalization of Zyn use, it is unknown if adolescents and young adults, the primary audience for this content, would also classify this content as humorous and appealing to youth [6,9,10].

Considering these previous studies and the growing attention to NTP-related content on social media, it is prudent to begin to address this research gap and examine if there are differences between adolescent, young adult, and adult interpretations of this content [15].

Vertically Integrated Project Model

One promising framework for examining this is the vertically integrated project (VIP) model. Originally developed at the Georgia Institute of Technology [16], the VIP model is a well-established approach that emphasizes long-term, faculty-led projects where students collaborate in multidisciplinary teams over multiple semesters or years [17,18]. Traditionally, VIP teams are led by 1 or more faculty members to facilitate greater mentorship and cross-collaboration. The VIP model attempts to create a supportable and long-lasting research community of faculty and students and can be used across various research modalities from basic science to community health [18]. This multigenerational approach facilitates knowledge transfer by enabling students to learn from and build upon the experience of more advanced scholars.

Regarding NTP content specifically, by incorporating youth as co-researchers within a structured, longitudinal framework, the VIP approach may enhance contextual interpretation of digital content while reducing interpretive bias. In other words, a VIP approach can allow for the examination of differences between adolescent, young adult, and adult interpretations of NTP social media content, as well as strengthen the interpretation of ambiguous NTP-related content, potentially identify engagement considerations among different audiences, and offer insight into how such material is targeted toward specific age segments.

Therefore, given this potential for a VIP approach to address the gap in adolescent participation in analyzing NTP content on social media and the gap in the literature regarding an examination of ONP content on YouTube, this study used a

XSL•FO
RenderX

VIP model to conduct a qualitative analysis of ONP-related YouTube videos. By engaging high school (HS) students, undergraduate public health students, and faculty researchers in parallel coding, this methodological case study sought to understand how perspectives across different age groups, academic levels, and lived experiences shape the interpretation of Zyn-related YouTube shorts. Situating content analysis within the VIP model suggests the methodological potential of age-diverse teams to capture nuances in social media data coding, reveal areas of interpretive divergence, and ultimately advance research methodologies in public health.

Methods

Data Collection

Following discussions with the Misinformation-identifying Youth in Tobacco and Health (MYTH) Youth Collaborative, a youth participatory action research (YPAR) program that engages adolescents aged 13-17 years as partners in the examination of NTP-related social media content [19], we decided to narrow our analysis to YouTube shorts as MYTH members identified these as far more frequently viewed by their peers than longer form YouTube content.

To prepare HS coders who had no prior experience analyzing social media content and to retrain experienced coders, we first manually collected a time-stamped sample of YouTube shorts containing the hashtag #Zyn on September 5, 2024. This separate training set was used only for coder training and calibration and was not included in the analytic sample. Data collection was conducted on a single day, using the same device and geographic location, to capture a consistent platform snapshot. The search was conducted through the YouTube shorts search interface rather than the general YouTube search page. The researcher opened a new browser page without logging into a YouTube or Google account to reduce the influence of account-based personalization on the returned results, entered the search term "#Zyn," and collected the first 100 publicly available shorts returned by the platform at the time of collection. No additional filters or sorting options were applied; videos were collected in the algorithmically ranked order returned by the YouTube shorts search interface. Because the videos were publicly available through the YouTube shorts interface, no private content was included. The research team then documented the video URLs for coding purposes, and videos were not downloaded or archived.

After completing the training and calibration process, we then collected a set of 300 YouTube shorts on September 18, 2024, for the main analysis, following the same search procedure described above.

Because this manuscript examines the VIP-informed coding process rather than estimating the prevalence of Zyn-related content on YouTube, the dataset serves as a methodological case example rather than a representative sample of all #Zyn content on YouTube shorts.

Coding Team

To facilitate the VIP approach, this study was conducted by a coding team consisting of HS, undergraduate, and faculty coders. The team included 2 public health faculty members with over a decade of experience in social media and substance use research, and 2 undergraduate students with approximately 2-3 years of prior experience coding social media data related to NTPs and other health-related topics. Additionally, the team included 2 HS students (GC and NC) affiliated with the MYTH Youth Collaborative. One HS coder had experience in NTP-related coding before this study, while the other had no previous formal coding experience.

Furthermore, the team was characterized by high levels of educational attainment and institutional homogeneity: faculty and undergraduate coders were affiliated with the same academic institution and were trained in public health–related disciplines, and both HS coders attended well-funded local institutions. While this shared background supported efficient collaboration and a common research orientation, it also highlights the importance of reflexivity in considering how researchers' social positions may shape the interpretation of youth-oriented social media content. The coding team was predominantly composed of individuals who identified as white women.

Coding Procedures

For qualitative coding, we adapted an existing codebook previously used in research on e-cigarette-related social media [20,21]. The codebook was revised to reflect ONP/Zyn-specific terminology and short-form video cues, such as nicotine pouch packaging, visible pouch use, flavor references, comparisons with other nicotine products, Zyn-related humor, and youth or masculinity framing. The broader adapted codebook included multiple categories for coding ONP/Zyn-related short-form video content. Because the purpose of this methodological case study was to examine coding challenges rather than to report the substantive distribution of all coded content categories, the present analysis focused on a subset of constructs and individual codes that had been especially difficult for the larger research team to apply consistently in prior work and that were examined in both the interrater reliability (IRR) assessment and team debriefing.

These focal constructs included sentiment (pro-ONP, anti-ONP, and unclear), account type (the type of the account presenting the content), content type (the primary purpose and presentation style of the video), and appeal to groups of interest (stylistic, cultural, or narrative cues intended to resonate with particular audiences, including youth-oriented or gender-coded elements). These constructs were closely examined in both the quantitative assessment and the team debriefing session. Specific individual codes included youth appeal (features likely to appeal to youth, such as animation, humor, or popular cultural references), male stereotype (content invoking masculine characteristics or stereotypes, including male-coded interests or imagery), and personal experience (references to the poster's own experience with Zyn/ONPs, including direct or implied personal use, but excluding proximal experience). Other coding categories included in the broader codebook were not specifically analyzed in the present study. An abbreviated supplemental codebook table is provided in Multimedia Appendix 1 for the focal constructs and key individual codes analyzed in this manuscript,

XSL•FO
RenderX

including operational definitions, coding structures, and descriptive examples. Because the coded units were YouTube shorts, examples are provided as descriptive summaries of visual, audio, and narrative cues rather than original video clips, screenshots, usernames, or URLs.

The lead researcher for this study, PYD, oversaw all coding trainings, assignments, and adjudications. All 3 teams first participated in a structured training process. Each team independently coded a set of 100 #Zyn YouTube shorts using the adapted codebook. This practice round was not included in the analytic dataset but served as a calibration exercise to familiarize coders with the constructs, identify initial challenges, and establish baseline consistency. The main coding process was iterative and occurred across 6 rounds (N=300). In each round, 2 coders within each age group (HS, undergraduate, and faculty) independently coded the same set of 50 videos ("double-coded"). Following independent coding, all coders participated in a team debriefing session to reflect on coding discrepancies, interpretive challenges, and lessons learned throughout the coding process.

IRR Assessment

To assess the consistency with which the 2 coders within each team applied the focal constructs and key individual codes to the 300-video analytic sample, IRR was assessed in Stata 19 (StataCorp LLC) using Cohen κ and percent agreement. For multicategory constructs (eg, sentiment: pro-ONP, anti-ONP, and unclear), mean κ for each construct and coding team was calculated by averaging all available category-specific κ values across categories and coding rounds. Given the relatively small size (N=300) of our dataset and the known limitations of κ in low-prevalence coding categories, percent agreement was used as a supplemental measure to provide a more nuanced view of consensus [22]. Commonly cited benchmarks (κ=0.70 and ≥80% agreement) were used as reference points to identify constructs that warranted closer review during adjudication, rather than as strict thresholds of acceptability [23]. IRR was calculated within coder groups (eg, HS-HS, undergraduate-undergraduate, and faculty-faculty). Reliability patterns were then compared descriptively across teams (HS, undergraduate, and faculty) to examine codebook consistency and identify areas that required further reflection during the team debriefing. Because each team included only 2 coders, these comparisons were not interpreted as evidence of group-level differences attributable to age, academic level, or lived experience.

Team Debriefing and Reflective Analysis

After all 6 coding rounds were completed, we held a structured team debriefing session with all coders to discuss coding challenges, disagreements, and lessons learned from the VIP-informed coding process. We treated this session as a team reflection rather than a standalone qualitative study. The goal was to help us better understand the descriptive IRR patterns and identify practical changes that could improve future youth-inclusive social media coding projects.

The debriefing lasted approximately 90 minutes, was conducted via Zoom, and was facilitated by the lead author. All coders participated as members of the research team. A semistructured guide was used to organize the discussion, including reflections on teamwork across coding rounds, coding categories that were easier or more difficult to apply, constructs with lower agreement, and strategies for managing disagreements during adjudication.

The recording was then transcribed verbatim, after which the transcript was reviewed in Dedoose 10.34 (SocioCultural Research Consultants, LLC) using a deductive-inductive thematic approach [24]. Deductive coding was guided by a priori analytic anchors, including codes that showed consistently lower or more variable IRR in the quantitative analysis and recurring points of disagreement identified during adjudications across coding rounds. These anchors provided a structured way to examine coders' reflections on interpretive challenges and connect them to patterns observed in the reliability results.

In parallel, inductive coding was used to capture themes that extended beyond these predefined areas, including reflections on age-based perspectives, learning across rounds, codebook limitations, interpretive drift, and collaborative dynamics across research teams. The lead author conducted the initial review and organization of the transcript. Because the debriefing was designed as a reflective team discussion rather than a standalone qualitative study, the transcript was not independently coded by multiple analysts, and formal intercoder agreement was not calculated. Themes were refined to capture both shared and divergent perspectives across the 3 age groups. All coder co-authors reviewed and confirmed that the final thematic interpretation accurately reflected the team debriefing. This combined deductive-inductive approach allowed the debriefing findings to contextualize the quantitative reliability patterns while identifying practical lessons for future age-diverse collaborative coding projects.

Ethical Considerations

This study was conducted as part of a larger research project examining NTP messaging on social media. The overall research program was reviewed and approved by the Institutional Review Board (IRB) (Protocol #22080079). MYTH Youth Collaborative members and their parents or guardians provided assent and consent prior to joining the collaborative as part of our IRB protocol. Before the debriefing session, all coders were informed of its purpose and provided verbal agreement to participate and to be audio-recorded. As co-authors of this manuscript, the 2 HS coders (GC and NC) also reviewed and confirmed the thematic interpretation of the debriefing data, providing participant validation. Additionally, HS coders viewed the videos through their own YouTube accounts, so their access was limited by YouTube's existing age-based restrictions. Videos requiring adult access or age verification would therefore not have been available to them during coding. We recognize that publicly available nicotine-related Shorts may still contain profanity, sexual innuendo, or normalization of substance use, and future youth-inclusive projects should build in clearer guidance for handling content that HS coders find uncomfortable.

For this study, we did not ask coders to disclose personal NTP use or personal exposure to NTP-related content, in order to avoid unnecessary disclosure of potentially sensitive

XSL•FO
RenderX

information. Instead, we treated age, prior coding experience, and self-described cultural or experiential perspectives as reflexive factors that could shape interpretation and examined these factors through adjudication notes and the team debriefing.

In consultation with the IRB and consistent with the ethical criteria for internet research proposed by Eysenbach and Till [25], the present analysis involved publicly available social media content and did not involve human participants. Therefore, informed consent was not required. To protect anonymity, no usernames, profile descriptions, or identifiable metadata are reported.

Results

Quantitative IRR Results

Overall IRR Trends Across Coding Teams

Across 6 iterative coding rounds (N=300), Cohen κ values varied across rounds and coding teams (Table 1). Across the 6 rounds, κ values ranged from 0.30 to 0.54 for the HS team, from 0.40 to 0.63 for the undergraduate team, and from 0.47 to 0.61 for the faculty team. These ranges are presented descriptively and are not interpreted as evidence of group-level differences in reliability.

Table 1. Interrater reliability across coding rounds by coding team (N=300).

Team	κ range
HS[a]	0.30-0.54
UG[b]	0.40-0.63
Faculty	0.47-0.61

[a]HS: high school.

[b]UG: undergraduate.

Within-Team IRR for Focal Constructs

Within-team reliability for focal constructs was assessed across the 6 coding rounds for each team using both Cohen κ and percentage agreement as complementary indicators of consistency (Table 2).

Table 2. Summary of within-team interrater reliability across coding rounds by focal construct and coding team.

Construct and team	Mean κ (SD)[a]	κ range[b]	Agreement range (%)
Sentiment			
HS[c]	0.43 (0.26)	0.37-0.58	80.67-94.33
UG[d]	0.64 (0.27)	0.43-1	85-95.29
Faculty	0.53 (0.29)	0.20-0.74	78.33-95.67
Account type			
HS	0.28 (0.34)	0-0.74	79.33-98
UG	0.39 (0.40)	0-0.75	84.33-98
Faculty	0.38 (0.38)	0-0.76	83-98
Content type			
HS	0.31 (0.27)	0.21-0.64	75.26-87.97
UG	0.50 (0.38)	0.07-0.86	88.33-93.33
Faculty	0.50 (0.37)	0.14-0.76	82-94.67
Appeal to groups of interest			
HS	0.11 (0.29)	0-0.10	79.23-98
UG	0.29 (0.46)	0-0.75	81-98.50
Faculty	0.38 (0.44)	0-0.67	82.67-99

[a]For multicategory constructs (eg, sentiment: pro-ONP, anti-ONP, and unclear), mean κ and SD for each construct and coding team were calculated using all available category-specific κ values across categories and coding rounds.

[b]The κ range represents the minimum and maximum of the round-level construct κ values (ie, the average κ across the codes comprising the construct within each coding round) across the six coding rounds. Undefined κ values were excluded from the calculation of mean κ, SD, and κ range.

[c]HS: high school.

[d]UG: undergraduate.

XSL•FO
RenderX

This analysis focused on the reliability of focal constructs to capture how coder agreement evolved over time. For sentiment, mean κ values varied across teams, with the undergraduate team showing the highest mean κ (0.64, SD 0.27), followed by the faculty team (mean 0.53, SD 0.29) and the HS team (mean 0.43, SD 0.26). κ values also varied across rounds, ranging from 0.43 to 1 for the undergraduate team, 0.20 to 0.74 for the faculty team, and 0.37 to 0.58 for the HS team. Percentage agreement remained relatively high across all teams, ranging from 85% to 95.29% for the undergraduate team, 78.33% to 95.67% for the faculty team, and 80.67% to 94.33% for the HS team.

For account type, mean κ values were modest across teams (HS: 0.28, SD 0.34; undergraduate: 0.39, SD 0.40; faculty: 0.38, SD 0.38), although κ ranges varied across rounds. The HS team showed κ values ranging from 0 to 0.74, the undergraduate team from 0 to 0.75, and the faculty team from 0 to 0.76. However, percentage agreement remained consistently high across all teams, ranging from 79.33% to 98% for the HS team, 84.33% to 98% for the undergraduate team, and 83% to 98% for the faculty team.

For content type, mean κ values were higher for the undergraduate and faculty teams (mean 0.50, SD 0.38 for undergraduate; mean 0.50, SD 0.37 for faculty) than for the HS team (mean 0.31, SD 0.27). κ values also varied across rounds, ranging from 0.21 to 0.64 for the HS team, 0.07 to 0.86 for the undergraduate team, and 0.14 to 0.76 for the faculty team. Percentage agreement remained relatively high across all teams, ranging from 75.26% to 87.97% for the HS team, 88.33% to 93.33% for the undergraduate team, and 82% to 94.67% for the faculty team.

For appeal to groups of interest, mean κ values were highest for the faculty team (mean 0.38, SD 0.44), followed by the undergraduate team (mean 0.29, SD 0.46) and the HS team (mean 0.11, SD 0.29). κ values also varied across rounds, ranging from 0 to 0.10 for the HS team, 0 to 0.75 for the undergraduate team, and 0 to 0.67 for the faculty team. Despite these lower κ values, percentage agreement remained high across all teams, ranging from 79.23% to 98% for the HS team, 81% to 98.50% for the undergraduate team, and 82.67% to 99% for the faculty team. In some instances, complete agreement among coders produced undefined κ values. These values were excluded from the calculation of mean κ and SD, as well as from the reported κ ranges.

IRR for Key Individual Codes

Analyses of key individual codes were conducted to examine overall IRR for each team across coding rounds (Table 3). For these codes, mean κ and mean percentage agreement across rounds were reported to summarize overall coding consistency.

Table 3. Mean within-team interrater reliability for key individual codes by coding team.

Code and team	Mean κ (SD)[a]	Mean % agreement[b]
Youth appeal		
HS[c]	–0.01 (0.01)	97
UG[d]	0.28 (0.13)	81
Faculty	0.44 (0.13)	82.67
Male stereotype		
HS	0.10 (0.06)	79.20
UG	0.56 (0.14)	81.33
Faculty	0.47 (0.22)	88
Personal experience		
HS	0.51 (0.35)	90.33
UG	0.30 (0.39)	90.67
Faculty	0.62 (0.18)	85

[a]For each individual code and coding team, mean κ was calculated by averaging round-level values across the six coding rounds; undefined κ values were excluded from both calculations.

[b]For each individual code and coding team, mean percent agreement was calculated by averaging round-level values across the six coding rounds.

[c]HS: high school.

[d]UG: undergraduate.

For youth appeal, the HS team had a near-zero mean κ (–0.01, SD 0.01), despite very high mean percentage agreement of 97%, suggesting that low code prevalence likely affected κ estimates. For male stereotype, mean κ values were higher for the undergraduate and faculty teams than for the HS team (mean 0.56, SD 0.14; mean 0.47, SD 0.22; and mean 0.1, SD 0.06 respectively), and percent agreement also remained higher for the undergraduate and faculty teams than for the HS team (81.33%, 88%, and 79.20%, respectively). For personal experience, mean κ values ranged from 0.30 (SD 0.39) to 0.62 (SD 0.18) across teams, and percent agreement remained high across all 3 teams, at or above 85%.

Qualitative Results From Team Debriefing Session

The thematic analysis of the debriefing session revealed several key themes related to the collaborative coding process,

XSL•FO
RenderX

highlighting the benefits, challenges, and nuanced interpretive differences that emerged across coding teams (Table 4).

XSL•FO
RenderX

Table 4. Themes from team debriefing session on the collaborative coding process.

Theme	Description	Illustrative quote
Interpretive drift and peer influence	Coders' interpretations of the codebook evolved over time. Rather than relying solely on the codebook, coders often tried to anticipate how their partner would code, leading to "flip-flopping" definitions across rounds and creating a coding logic based on interpersonal dynamics.	• "I kept putting what I thought the other faculty would think, you know, because I was always coding with the other faculty...you start to think just like them. You don't think of the code book." (Faculty) • "I was really remembering the conversations that we were having maybe even more so than the code book at times, because I knew that I was trying to think like the other coder…" (UG[a])
Codebook limitations and tacit knowledge	The static codebook struggled to accommodate the novelty and ambiguity of short-form video content. Participants found it difficult to translate intuitive feelings about content into concrete, replicable coding rules.	• "So much of it (the content) just to us, felt male stereotype for many different reasons, but we couldn't really define it to put it in the code book..." (UG) • "I would feel like, okay, we've discussed everything we needed to discuss; we have very clear definitions, and then the next round would bring like, whole different type of videos that I hadn't really considered." (Faculty)
Perceived role of cultural and experiential lenses	Coders described their interpretations as being influenced by cultural background, prior research experience, and lived experience. This was especially evident in the appeal to groups of interest construct.	• "I think in that sense, it's very hard to define what is male stereotype. And I think with the other undergraduate coder and I specifically, of course, we had worked together before in the lab, but I feel like in this codebook in particular, our cultural differences really impacted how we saw it." (UG) • "I always have trouble trying to judge targeting of a group that I don't belong to...it also feels sometimes icky to me to try to code things that doesn't just make it seem like a terrible stereotype..." (Faculty)
Subjectivity of content-specific codes	Constructs such as sentiment, content type, and appeal to groups of interest were consistently challenging for all teams due to their inherent subjectivity.	• "I remember at one point we sort of redefined comedy as it is supposed to be entertaining. And I think that helped. But...at first, I didn't code it as comedy, because I was like, it's not funny. But then it was like, okay, but they're trying to be funny." (Faculty) • "I feel like I struggled with the other substance and Zyn, because a lot of the videos we saw were talking about nicotine free pouches...my head would want to be like, oh, it's a Zyn, but then it wasn't." (HS[b]) • "...Initially, I would say she looks 15 to me, that's what I think. Um, after going to her account, sometimes they put their ages, they're like, oh, 23..." (UG)
Divergent heuristics between coder groups	The 3 teams developed different approaches to resolving ambiguity. Experienced faculty and UG coders often referred to definitions from previous projects, while the HS team used "fresh eyes" and concrete strategies like detailed notes to reach consensus more quickly.	• "We were able to lean on some of that (prior experience)...the high school students...just got it really quickly...I think we were just kind of having analysis paralysis, right? Whereas they were able to come with fresh eyes too." (Faculty) • "...to always make notes on the certain sections so that we would establish, like, this one is only if they show Zyn in the video...adding notes and just kind of agreeing on certain things...definitely helped." (HS)

XSL•FO
RenderX

Theme	Description	Illustrative quote
Divergent emotional responses to content	The teams exhibited divergent emotional responses to the repetitive and often "annoying" video content. While the faculty team explicitly reported experiencing cognitive fatigue and frustration, the UG team developed humor as a coping mechanism.	• "...just some of these videos were so annoying and so hard to watch, and I feel like maybe I could have done a better job if I was not just being like, oh my gosh, I can't watch another guy lifting weights with weird music in the background..." (Faculty) • "...we just kind of took it (the video) with humor, and we were just seeing everything as, like, hallucination of videos, because some of them were so insane." (UG)
Methodological suggestions for future projects	Participants proposed strategies to improve future projects, including re-pairing coders and enhancing the codebook with more specific examples to aid clarity.	• "I do think it'd be interesting in the future if we did...the switching up of pairs...switching it up by age." (Faculty) • "...something that would be helpful is adding something in the example category...if there was an example, we could just easily refer back to that." (HS)

[a]UG: undergraduate.

[b]HS: high school.

Interpretive Drift and Peer Influence

The coding process was not a static application of the codebook. Participants described a phenomenon of "interpretive drift" where their interpretations evolved over time, influenced by anticipating their partner's logic. This created a coding logic based on interpersonal dynamics rather than relying consistently on the written codebook definitions. As one faculty coder noted, "I kept putting what I thought the other faculty would think… you start to think just like them. You don't think of the codebook".

Codebook Limitations and Evolving Definitions

Coders consistently expressed that the formal codebook struggled to capture the novelty and ambiguity of short-form video content. Participants found it difficult to translate intuitive feelings about content into concrete, replicable coding rules. A faculty coder explained, "I would feel like, okay, we've discussed everything... and then the next round would bring like, whole different types of videos that I hadn't really considered".

Perceived Role of Cultural and Experiential Lenses

Coders described their interpretations as being influenced by cultural background, prior research experience, and lived experience, particularly when applying subjective codes such as male stereotype. One undergraduate coder reflected on this, mentioning, "I feel like in this codebook in particular, our cultural differences really impacted how we saw it".

Subjectivity of Content-Specific Codes

Several constructs (eg, content type and appeal to groups of interest) were consistently challenging for all teams due to their inherent subjectivity. For instance, a faculty coder shared the struggle with coding comedic video: "at first, I didn't code it as comedy, because I was like, it's not funny. But then it was like, okay, but they're trying to be funny".

Divergent Heuristics Between Coder Groups

The 3 teams developed different approaches to resolving ambiguity. The more experienced coders often relied on prior research strategies. In contrast, the HS team, without the burden of previous project methods, created their own concrete methods to reach consensus. A HS coder explained their approach, saying they learned to "always make notes on the certain sections so that we would establish, like, this one is only if they show Zyn in the video... adding notes and just kind of agreeing on certain things... definitely helped".

Divergent Emotional Responses to Content

The teams exhibited different emotional responses to the repetitive and often "annoying" video content. The faculty team reported experiencing cognitive fatigue and frustration, with one member mentioning they "can't watch another guy lifting weights with weird music in the background". Conversely, the undergraduate team developed humor as a coping mechanism, with one coder saying, "we just kind of took it (the video) with humor... because some of them were so insane."

Methodological Suggestions for Future Projects

Participants proposed several strategies to improve future collaborative coding projects. These suggestions included re-pairing coders to reduce interpretive drift and enhancing the codebook with more specific examples to aid clarity. One faculty coder suggested, "I do think it'd be interesting in the future if we did... the switching up of pairs... switching it up by age."

Discussion

This methodological case study examined how perspectives across different age groups, academic levels, and lived experience shaped the qualitative coding of youth-oriented short-form nicotine-related content within a VIP-informed, age-diverse coding structure. Rather than evaluating the prevalence or distribution of Zyn-related content, the study focused on the coding process itself, including descriptive IRR

XSL•FO
RenderX

patterns and team reflections on interpretive challenges. Together, these findings illustrate how coder experience, age-related perspective, prior research exposure, and cultural lenses may shape the application of codes to ambiguous social media content.

The descriptive IRR patterns suggest that differences across teams may have reflected not only coding consistency, but also coders' prior experience and familiarity with the content. The more experienced undergraduate and faculty teams showed higher κ values on complex, subjective constructs (eg, sentiment and content type). Their familiarity with established heuristics may have contributed to more consistent application of some codes. In contrast, the HS team demonstrated a notable learning curve, with IRR showing steady improvement across rounds. Their "fresh eyes" and development of concrete coding strategies appeared to help them apply the codes more consistently across rounds, particularly on youth-oriented content from a platform with which they were familiar.

The team debriefing findings helped contextualize the descriptive IRR patterns by suggesting that numerical disagreements reflected not only reliability challenges, but also the different forms of experience and perspective that coders brought to the collaborative structure of the VIP model [17]. For youth appeal, the near-zero κ among HS coders was likely influenced by the low prevalence of the code, as described in the quantitative results. The debriefing findings should therefore be understood not as an alternative explanation for the κ value, but as context for how coders understood and applied the youth appeal construct. Faculty coders expressed discomfort applying the code to a group to which they did not belong, whereas the HS team's perspective as peers of the target audience likely shaped their application of the code differently. This suggests that the quantitative patterns were not simply measures of "correctness," but rather reflected the distinct interpretive frameworks each team brought to the coding process. Consistent with the broader pattern observed for complex constructs, the higher κ values for male stereotype among the undergraduate and faculty teams may similarly reflect more stable within-pair coding heuristics and greater prior coding experience, rather than clearer or more accurate interpretation of the construct. The debriefing theme of Perceived Role of Cultural and Experiential Lenses helped explain why this code remained difficult, as coders from different cultural backgrounds interpreted masculinity cues differently. Fluctuations in IRR across rounds were explained by the themes of interpretive drift and codebook limitations. The team debriefing revealed that teams often developed their own internal logic and that the static codebook struggled to capture the nuances of fast-paced video content, leading to inconsistent codebook application. Finally, although personal experience showed moderate reliability differences across teams, this code was not a prominent topic during the debriefing session. Coders rarely reflected on challenges related to identifying personal experience narratives. This absence may partly reflect the relatively clear operational definition provided in the codebook, which specified explicit self-referential statements, making the code easier to apply consistently and less likely to become a major topic of discussion. Together, by integrating the descriptive IRR patterns with the team debriefing findings, these code-level examples illustrate a potential methodological benefit of age-diverse coding teams: they may help researchers recognize when constructs such as humor, youth appeal, masculinity, or harm-minimizing framing depend on audience-specific and culturally situated interpretations rather than on unambiguous content features. In future substantive content analyses, these interpretive differences could inform the refinement of codebook categories used to examine audience-oriented engagement cues; however, we also wanted to note that the present case study was not designed to infer creators' intent, platform targeting, exposure effects, or associations with youth nicotine use.

This study has several limitations. First, the sample size of 300 YouTube shorts, while sufficient for a methodological exploration, limits the extent to which these coding-process insights can be generalized beyond this specific project. Our analysis was also confined to a single platform (YouTube) and a single hashtag (#Zyn); future research could benefit from including other short-form video platforms such as TikTok to capture a broader range of Zyn-related social media content.

Second, although the VIP model intentionally incorporated coders across multiple age groups, the coding team was relatively homogeneous. Coders generally had high levels of educational attainment or academic exposure, shared public health–related interests, and were predominantly White women. This shared background may have shaped how the team interpreted youth-oriented social media content, particularly for culturally specific or identity-related constructs. Future studies would benefit from more demographically and educationally diverse coding teams to further examine how coder positionality shapes collaborative coding processes and reliability patterns.

Third, although HS coders viewed the videos through their own YouTube accounts, we did not have a separate protocol for handling videos that HS coders might find uncomfortable. Future youth-inclusive coding projects should include clearer guidance for this situation.

Fourth, the team debriefing was facilitated and analyzed by the lead author, who also coordinated the coding process. Because the debriefing included HS, undergraduate, and faculty coders in the same discussion, participants may have been influenced by social desirability, team hierarchy, or group consensus. For this reason, we interpreted the debriefing findings as reflective team insights that help contextualize the coding process, rather than as generalizable evidence about how youth interpret Zyn-related content.

Finally, IRR was calculated using Cohen κ within fixed 2-coder teams and was used descriptively in this case study. Given the small coding team, the small number of videos per round, and the low prevalence of several codes, differences across teams should be interpreted only as case-based insights into how this specific team approached difficult-to-code Zyn-related YouTube shorts. Future reliability-focused studies, particularly those involving low-prevalence codes, may benefit from including Krippendorff α as an additional metric [26].

To our knowledge, this study is among the first to apply a VIP model to qualitative social media coding in public health. Our

XSL•FO
RenderX

findings suggest that this approach is feasible and may be useful for identifying parts of the coding process that require additional clarification. In this study, when interpreted alongside the debriefing findings, lower IRR did not simply indicate poor reliability. We found that it also helped reveal where codebook definitions, coder assumptions, and team decision-making needed further discussion. However, persistent disagreement after training and adjudication should still be treated as a potential threat to classification validity and as a signal that code definitions, examples, or coder training require further refinement before substantive findings are interpreted. Overall, this perspective is especially important in studies of youth-oriented digital media, where meaning often depends on rapidly shifting humor, aesthetics, and platform-specific norms.

In addition to its methodological value, this study also highlights the educational and equity-related potential of adapting the VIP model for public health research. By extending a model originally designed for undergraduate training to include HS students as full analytic partners, the project created opportunities for early research skill development, mentorship, and exposure to scientific career pathways. At the same time, the team benefited from bidirectional learning in which the 2 HS coders (GC and NC) brought their own experiences as youth research partners, while the more experienced coders helped guide the coding process and codebook application. Future research could examine whether alternative team structures, such as cross-age coding pairs, help clarify how coders with different backgrounds resolve ambiguity during social media coding. Such designs may provide a more direct assessment of whether and how age-diverse collaboration shapes codebook refinement, adjudication, and consensus-building.

In conclusion, this methodological case study illustrates how a VIP-informed, age-diverse coding team can be used to examine interpretive challenges in coding youth-oriented nicotine-related short-form video content. The findings should be interpreted as case-based insights into one project's coding process rather than as evidence that the VIP model improves reliability or generates more accurate interpretations. For public health studies examining how different groups interpret online health messages, the VIP model may offer a practical structure for incorporating those perspectives into the research process.

Data Availability

The dataset analyzed in this study consists of publicly available YouTube shorts containing the hashtag #Zyn, manually collected on September 18, 2024. Due to platform dynamics and potential changes in content availability, direct redistribution of the original videos is not provided. The coding results that support the findings of this study are presented within the manuscript tables.

Funding

This study was supported by the National Institutes of Health under award 5R01MD018543-04 (principal investigator: JES).

Authors' Contributions

Conceptualization: JPD (lead), BLH (equal), JES (equal)
Data curation: JPD
Formal analysis: JPD
Funding acquisition: JES (lead), BLH (supporting)
Investigation: JPD (lead), BLH (equal), GC (equal), JES (equal), NC (equal), PN (equal), TGL (equal)
Methodology: JPD (lead), BLH (supporting), JES (supporting)
Project administration: JPD (lead), CL (supporting)
Supervision: JPD, JES (supporting)
Visualization: JPD (lead), JES (supporting)
Writing – original draft: JPD
Writing – review and editing: JPD (lead), JES (equal), BLH (supporting), CL (supporting), GC (supporting), NC (supporting), PN (supporting), TGL (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Abbreviated codebook for focal constructs and individual codes.
[PDF File (Adobe PDF File), 179 KB - infodemiology_v6i1e95363_app1.pdf]

References

1. Aldukhail S. Relationship between e-cigarette media content and product use: A scoping review. Tob Induc Dis 2025;23:1-20. [doi: 10.18332/tid/200547] [Medline: 40018080]
2. Zyn. URL: https://www.zyn.com [accessed 2026-02-12]

XSL•FO
RenderX

3. Marynak KL, Wang X, Borowiecki M, Kim Y, Tynan MA, Emery S, et al. Nicotine pouch unit sales in the US, 2016-2020. JAMA 2021;326(6):566-568 [FREE Full text] [doi: 10.1001/jama.2021.10366] [Medline: 34374729]
4. Leventhal AM, Cho J, Vogel EA, Tackett AP, Harlow AF, Wong M, et al. Differences in intention to use flavored oral nicotine products among young adult e-cigarette users and non-users. Prev Med Rep 2022;30:102027 [FREE Full text] [doi: 10.1016/j.pmedr.2022.102027] [Medline: 36304078]
5. Regulation and enforcement of non-tobacco nicotine (NTN) products. US Food and Drug Administration. URL: https://www.fda.gov/tobacco-products/products-ingredients-components/regulation-and-enforcement-non-tobacco-nicotine-ntn-products [accessed 2026-02-12]
6. Zenone M, Harries B, Hartwell G. The promotion of oral nicotine pouches for non-smoking cessation purposes on tiktok. Nicotine Tob Res 2026;28(2):282-286 [FREE Full text] [doi: 10.1093/ntr/ntaf024] [Medline: 39879071]
7. Tackett AP, Barrington-Trimis JL, Leventhal AM. 'Flavour ban approved': new marketing strategies from tobacco-free nicotine pouch maker Zyn. Tob Control 2023;32(e1):e134-e135 [FREE Full text] [doi: 10.1136/tobaccocontrol-2021-057222] [Medline: 35459750]
8. Mand A, Fonteyne K, Struik L. Examining how oral nicotine pouches are trending on TikTok: a qualitative descriptive study. JMIR Form Res 2025 Nov 14;9:e73032 [FREE Full text] [doi: 10.2196/73032] [Medline: 41236804]
9. Donaldson SI, La Capria K, DeJesus A, Ganz O, Delnevo CD, Allem J. Describing ZYN-related content on TikTok: content analysis. Nicotine Tob Res 2026 Jan 22;28(2):277-281. [doi: 10.1093/ntr/ntaf016] [Medline: 39823221]
10. Basch C, Clarke Hillyer G, Yousaf H, Basch C. Content trends and tone of videos posted on the popular hashtag #Zyns on TikTok. J Consumer Health Internet 2025 Jul 14;29(3):375-386. [doi: 10.1080/15398285.2025.2521592]
11. Vogel EA, Ranker LR, Harrell PT, Hart JL, Kong G, McIntosh S, et al. Characteristics of adolescents' and young adults' exposure to and engagement with nicotine and tobacco product content on social media. J Health Commun 2024;29(6):383-393 [FREE Full text] [doi: 10.1080/10810730.2024.2355291] [Medline: 38775659]
12. Churchill V, Fairman R, Brown D, Massey Z, Ashley D, Popova L. "I Get the Flavors and It Makes Me Love Vaping More": how and why youth users modify electronic nicotine delivery systems. Nicotine Tob Res 2023;25(11):1791-1797 [FREE Full text] [doi: 10.1093/ntr/ntad104] [Medline: 37358211]
13. Kolotouchkina O, Rangel C, Gómez PN. Digital media and younger audiences. Media Commun 2023;11(4):124-128. [doi: 10.17645/mac.v11i4.7647]
14. Manganello JA, Henderson VR, Jordan A, Trentacoste N, Martin S, Hennessy M, et al. Adolescent judgment of sexual content on television: implications for future content analysis research. J Sex Res 2010;47(4):364-373. [doi: 10.1080/00224490903015868] [Medline: 19536684]
15. Morales M, Fahrion A, Watkins SL. #NicotineAddictionCheck: puff bar culture, addiction apathy, and promotion of e-cigarettes on tiktok. Int J Environ Res Public Health 2022;19(3):1820 [FREE Full text] [doi: 10.3390/ijerph19031820] [Medline: 35162846]
16. Vertically integrated projects. University of Bath. URL: https://www.bath.ac.uk/guides/vertically-integrated-projects/ [accessed 2026-02-13]
17. Coyle E, Sonnenberg-Klein J. A pathway to systemic reform: VIP programs and the VIP consortium. SPUR 2025;8(3):7-13. [doi: 10.18833/spur/8/3/9]
18. Goins E, Mahabaduge H. Introduction – spring 2025. SPUR 2025;8(3):2-3. [doi: 10.18833/spur/8/3/8]
19. MYTH youth collaborative. Pitt+Me. URL: https://pittplusme.org/studyarms/publicdetails?guid=8ba1530c-b650-4e3a-a6f1-c2f110c4ad41 [accessed 2026-02-13]
20. Sidani JE, Colditz JB, Barrett EL, Chu KH, James AE, Primack BA. JUUL on Twitter: analyzing tweets about use of a new nicotine delivery system. J Sch Health 2020 Feb;90(2):135-142 [FREE Full text] [doi: 10.1111/josh.12858] [Medline: 31828791]
21. Hoffman B, Tripathi A, Shensa A, Dou JP, Narendorf P, Hundi N, et al. Response to the Netflix docuseries "Big Vape: The Rise and Fall of JUUL": mixed methods analysis of YouTube comments using qualitative coding and topic modeling. JMIR Form Res 2025 Sep 19;9:e76737 [FREE Full text] [doi: 10.2196/76737] [Medline: 40971587]
22. Dettori JR, Norvell DC. Kappa and beyond: is there agreement? Global Spine J 2020;10(4):499-501 [FREE Full text] [doi: 10.1177/2192568220911648] [Medline: 32435572]
23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Res Psychol 2008;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001;323(7321):1103-1105 [FREE Full text] [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
26. Marzi G, Balzano M, Marchiori D. K-alpha calculator-krippendorff's alpha calculator: a user-friendly tool for computing krippendorff's alpha inter-rater reliability coefficient. MethodsX 2024;12:102545 [FREE Full text] [doi: 10.1016/j.mex.2023.102545] [Medline: 39669968]

XSL•FO
RenderX

Abbreviations

HS: high school
IRB: institutional review board
IRR: interrater reliability
NTP: nicotine and tobacco product
MYTH: Misinformation-identifying Youth in Tobacco and Health
ONP: oral nicotine product
VIP: vertically integrated project
YPAR: youth participatory action research

Edited by J Yang; submitted 17.Mar.2026; peer-reviewed by C Li-Chuan, W Qin, S Kulkarni; comments to author 22.Jun.2026; revised version received 26.Jul.2026; accepted 09.Aug.2026; published 02.Sep.2026.

Please cite as:
(Pengyue) Dou J, Hoffman BL, Narendorf P, Limoncic TG, Carver G, Connor N, Larkin C, Sidani JE
Applying the Vertically Integrated Project Model to Youth-Inclusive Social Media Coding of Zyn-Related YouTube Shorts: Methodological Case Study
JMIR Infodemiology 2026;6:e95363
URL: https://infodemiology.jmir.org/2026/1/e95363
doi:10.2196/95363
PMID:

©Julia (Pengyue) Dou, Beth L Hoffman, Piper Narendorf, Tatiana Grinberg Limoncic, Grace Carver, Natalia Connor, Christine Larkin, Jaime E Sidani. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 02.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Characterizing Family Abuse in Suicidal Ideation Posts on Reddit: Large Language Model–Assisted Content Analysis

Salim Sazzed[1], PhD

Georgia Southern University, Statesboro, GA, United States

Corresponding Author:
Salim Sazzed, PhD
Georgia Southern University
1332 Southern Dr
Statesboro, GA, 30460
United States
Phone: 1 912 478 2727
Email: ssazzed@georgiasouthern.edu

Abstract

Background: Suicidal ideation is a significant global public health concern, with family abuse recognized as a key risk factor. Although clinical studies have examined the link between family abuse and mental health, its representation in social media discussions—particularly in relation to suicidal ideation—remains largely unexplored.

Objective: This study aimed to examine how family abuse manifests in suicidal ideation narratives on social media, focusing on the demographic characteristics of individuals reporting abuse, the types of abuse described, and relationships to perpetrators.

Methods: We analyzed 27,434 posts from the Reddit *SuicideWatch* forum to identify posts in which suicidal ideation and family abuse co-occurred. Using a combination of manual review, large language model–assisted keyword expansion, natural language processing–based information extraction, and human validation, we examined self-disclosed age and gender among Reddit users who reported family abuse in suicidal ideation posts, along with abuse types and relationships to perpetrators.

Results: Among posts with self-reported age information, individuals aged 18 to 24 years accounted for the largest proportion (144/313, 46.0%). Among posts with self-reported gender information, men represented 57.3% (86/150). Physical abuse was the most frequently reported form of family abuse (683/975, 70.1%; 95% CI 67.18%-72.93%), followed by emotional or psychological abuse (350/975, 35.9%; 95% CI 32.89%-38.91%) and sexual abuse (296/975, 30.4%; 95% CI 27.47%-33.25%). The parent category was the most frequently identified perpetrator category (435/975, 44.6%; 95% CI: 41.52%-47.75%).

Conclusions: This study enhances our understanding of how family abuse appears in social media posts involving suicidal ideation. The findings demonstrate the potential of computational social media analysis to support descriptive, hypothesis-generating research and inform the design of online support resources for family abuse–related suicidal distress.

(JMIR Infodemiology 2026;6:e90340) doi:10.2196/90340

KEYWORDS

suicidal ideation; family abuse; domestic violence; social media; Reddit; mental health; large language models

Introduction

Background

Suicidal ideation is a major public health concern and a leading cause of death, particularly among young individuals [1]. According to the World Health Organization, approximately 700,000 people die by suicide each year [2]. Suicidal ideation arises from a complex interplay of psychological, social, and environmental factors, with profound consequences for individuals, families, and communities [3,4]. Among these factors, family abuse is recognized as a significant contributor to suicidal ideation [5,6]. Family abuse—encompassing physical, emotional, and psychological harm—induces chronic stress and emotional instability, severely affecting mental health [7]. Survivors often report helplessness, low self-worth, and social isolation, all of which are strongly associated with suicidal ideation [8,9]. The relationship between family abuse and suicidal ideation is complex and multidimensional, influenced by intersecting factors such as the severity of abuse, the presence of protective social supports, and individual psychological resilience [10,11]. Understanding these intricate dynamics is essential for advancing our knowledge of suicide risk and developing nuanced interventions tailored to those affected by

XSL•FO
RenderX

familial abuse. Existing clinical and survey-based studies demonstrate the profound impact of family abuse on mental health and its strong association with suicidal behaviors [12-14].

Individuals with a history of family abuse are at higher risk for psychological distress and suicidal ideation [8,15,16]. Joiner's [3] interpersonal theory of suicide highlights perceived burdensomeness and thwarted belongingness as critical risk factors, both exacerbated by abusive experiences. Intimate partner violence, a key form of family abuse, has been linked to severe psychological consequences [5]. Furthermore, the adverse childhood experiences (ACEs) study, along with systematic reviews and meta-analyses, underscore the long-term effects of childhood maltreatment on suicide risk [8,17-19]. Research by Duprey et al [6] and Charak et al [7] further demonstrates the correlation between early abuse and increased risk of suicidal ideation and attempts across the lifespan. Studies have also examined vulnerabilities in specific populations: Kwok et al [9] identified adolescents as particularly susceptible to the effects of childhood maltreatment, while Kaplan et al [11] explored the experiences of adolescent suicide attempters with a history of physical abuse. Together, these studies emphasize the multifaceted nature of family abuse and the need for targeted, nuanced approaches to understanding its psychological and behavioral consequences. Unlike clinical assessments, social media mental health posts provide largely unfiltered, freeform accounts of distress and experiences of family abuse, offering insights that may not emerge in formal assessments. The growing body of research examining mental health through digital platforms has highlighted the unique potential of online spaces to reveal critical insights into individuals' lived experiences [20,21]. Social media platforms, which are increasingly used as outlets for expressing emotional distress, offer an unprecedented opportunity to study the real-time manifestations of suicidal ideation and its associated triggers [22-25]. Through the analysis of social media health data, researchers can gain critical insights into the various factors that shape suicidal thoughts, ultimately contributing to the development of evidence-based strategies for prevention and intervention [26]. However, despite recent progress, gaps remain in understanding the multifaceted nature and risk factors of suicidal ideation, including the dynamics of family abuse within online communities and discourses.

To address this gap, this study explores the intersection of family abuse and suicidal ideation in online discourse by analyzing textual content from the Reddit (Reddit, Inc) forum *r/SuicideWatch*. We aim to uncover the multidimensional nature of family abuse among individuals reporting suicidal thoughts by addressing the following three research questions (RQs):

RQ1 addresses the demographic characteristics of individuals reporting family abuse. This research question examines the demographic distributions of individuals who report experiencing family abuse alongside suicidal ideation by extracting attributes such as gender and age. Understanding these characteristics is essential for identifying vulnerable groups and informing the development of tailored interventions.

RQ2 addresses the relationships to perpetrators. What are the relationships to perpetrators in cases of family abuse, including roles such as parent (eg, father or mother), sibling, or intimate partner (eg, boyfriend or spouse)? Examining these relationships is critical for identifying the dynamics that contribute to suicidal ideation and informing targeted intervention strategies.

RQ3 addresses the types of family abuse. What is the prevalence of different forms of family abuse (eg, physical and emotional), and what specific characteristics define each type?

Understanding these patterns clarifies the range of harm individuals experience and informs the development of more comprehensive and targeted interventions.

To explore these research questions, we analyze social media posts related to suicidal ideation written by individuals reporting experiences of family abuse. Using a dataset of approximately 27,434 posts collected from the Reddit forum *r/SuicideWatch*, a keyword-based search was applied to identify about 975 posts that explicitly describe experiences of family abuse. We then examine key demographic characteristics of individuals reporting family abuse; explore relationships to perpetrators; and identify prevalent forms of abuse, including physical, emotional, and sexual abuse, as well as neglect. Our findings reveal demographic distributions and the prevalence and forms of family abuse across different groups. These insights enhance our understanding of family abuse in relation to suicidal ideation and can inform targeted suicide prevention and mental health interventions.

Contributions

Table 1 summarizes the main contributions and novelty of the study across 6 key dimensions: data source, demographics, perpetrator relationships, abuse types, methodology, and implications. For each dimension, we describe the specific contribution of this work and the corresponding gap in the existing literature that it addresses.

XSL•FO
RenderX

Table 1. Summary of study contributions and gaps addressed.

Aspect	Contribution	Novelty or gap addressed
Data source	Analyzed 27,434 Reddit posts; identified 975 posts explicitly mentioning family abuse in suicidal ideation	No prior studies have specifically examined family abuse in suicidal posts on social media.
Demographics	Extracted self-reported age and gender; highlighted high representation of young adults and LGBTQ+[a] individuals	Prior work focused on detecting suicidal content rather than examining demographic patterns in abuse-related posts.
Perpetrator relationships	Identified abuse across parents, siblings, partners, stepparents, and extended family	First systematic analysis of relational contexts in social media posts reporting suicidal ideation.
Abuse types	Characterized physical, emotional, and sexual abuse and neglect with associated keywords and frequency counts	No prior social media studies mapped abuse types in posts reporting suicidal ideation.
Methodology	Developed an LLM[b]-supported keyword- and context-based extraction module to detect abuse mentions and perpetrators	Novel approach for systematic extraction from social media text, combining LLM capabilities with rule-based methods.
Implications	Insights derived from this study may help guide the development of interventions for groups at higher risk of familial abuse and suicidal ideation	Demonstrates how insights from social media posts can inform intervention strategies for familial abuse–related suicidal ideation, an approach not previously explored.

[a]LGBTQ+: lesbian, gay, bisexual, transgender, queer or questioning, and other sexual and gender identities.

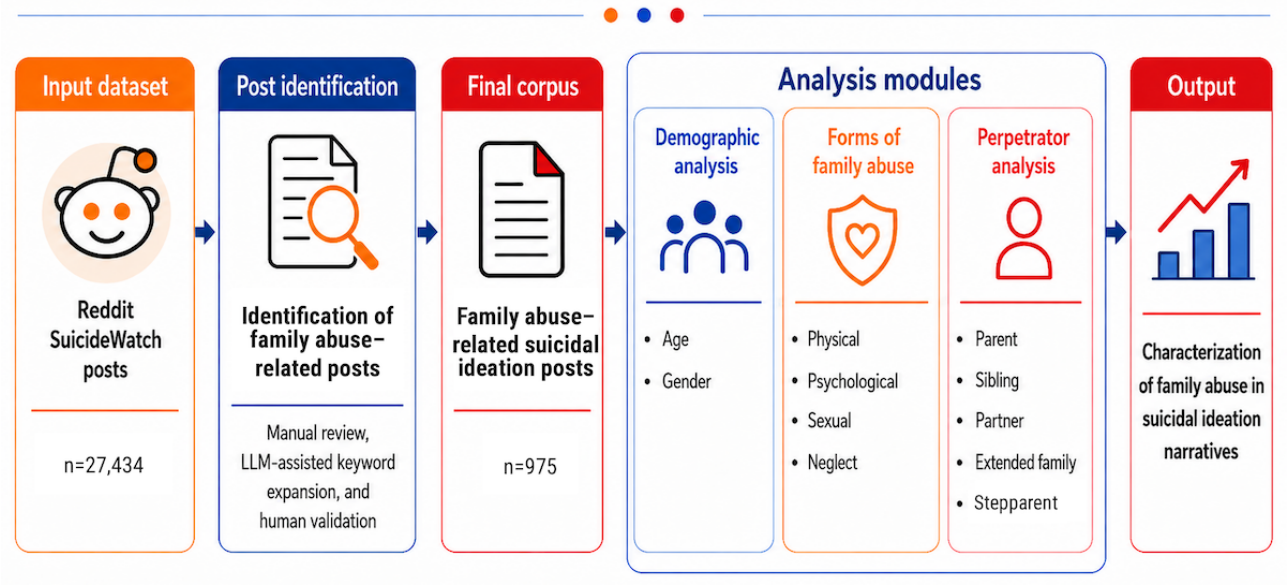
[b]LLM: large language model.

Methods

Study Overview

Figure 1 provides an overview of the study workflow. Starting from the Reddit *SuicideWatch* corpus, we first identified family abuse–related suicidal ideation posts using a keyword-based selection process informed by manual review, large language model (LLM)–assisted keyword expansion, and human validation. The identified corpus was then analyzed across 3 dimensions: demographic characteristics, forms of family abuse, and perpetrator categories.

Figure 1. Overall workflow for identifying and characterizing family abuse–related suicidal ideation posts. LLM: large language model.



Dataset Creation and Description

Collection of Suicidal Ideation Posts

The suicidal ideation posts analyzed in this study were sourced from Reddit, a widely used social media platform where users engage in discussions on diverse topics. The corpus used in this study was developed in our previous work [27] using a publicly available Reddit dataset hosted on Kaggle. The original dataset contained approximately 230,000 Reddit posts, including suicidal posts collected from the *r/SuicideWatch* subreddit and nonsuicidal posts collected from multiple other subreddits. As the present study focuses on suicidal discourse, only posts from *r/SuicideWatch* were included, resulting in a corpus of 27,434 posts.

The *r/SuicideWatch* subreddit is a discussion forum where users anonymously share experiences related to suicidal ideation and severe psychological distress. Posts in this forum typically consist of personal experiences, thoughts, and help-seeking

XSL•FO
RenderX

behaviors, providing largely unfiltered accounts of distress and family abuse as reported directly by affected individuals. As a result, posts from *r/SuicideWatch* provide a rich source of naturally occurring text that facilitates the analysis of family abuse in the context of suicidal ideation.

The posts were published between December 16, 2008, and January 2, 2021, covering a broad temporal span. The dataset consists solely of original user posts, excluding all responses and comments. During preprocessing, only English-language posts containing meaningful textual content were retained.

Selection of Suicidal Posts Involving Family Abuse

Manually reviewing these 27,434 suicidal posts to identify incidents related to family abuse would be highly resource intensive, particularly given that each post contains, on average, more than 700 words. To address this challenge, an automated keyword-based selection approach was used, where initial keywords are derived through manual review and later extended by an LLM to ensure broad coverage of expressions related to family abuse. Although keyword-based filtering has limitations in contextual understanding, the substantial reduction in manual effort, especially in the initial screening phase, makes it a feasible and widely adopted method for large-scale data selection and annotation [28,29].

Keyword Selection Procedure

Overview

The keyword selection process follows a systematic procedure designed to ensure both precision and comprehensive coverage. The procedure consists of 2 phases. In phase 1, a manual review is performed on a subset of posts to identify naturally occurring terms and expressions associated with family abuse. In phase 2, an LLM is used to generate additional relevant keywords, expanding coverage to include variations and contextually related expressions not captured in the initial manual review.

Phase 1: Manual Review

A representative subset of 2000 posts, randomly selected from the corpus, was manually reviewed to ensure an unbiased sample. Each post in the corpus was assigned a unique identifier, and 2000 identifiers were selected using a random number generator, the *random.sample()* function of Python (Python Software Foundation), so that every post had an equal chance of inclusion. This subset was chosen to enable detailed examination of naturally occurring expressions of family abuse across diverse contexts in the corpus.

The manual review and keyword extraction in phase 1 were conducted by 2 trained annotators with expertise in social media discourse and mental health: annotator 1, who has more than 8 years of experience analyzing social media and mental health–related content, and annotator 2, who has several years of experience analyzing social media data.

During phase 1, reviewers extracted phrases that explicitly or implicitly referred to abusive behaviors, harmful family dynamics, or distressing household environments and compiled them into an initial candidate keyword set. This phase was designed as an exploratory keyword-generation step rather than a fixed-label double-coding task.

The goal was to identify candidate family abuse–related expressions, not to assign abuse or nonabuse labels or abuse-type labels to a fixed set of posts. Therefore, Cohen κ was not calculated for phase 1 because there was no fixed set of categorical labels or shared annotation units on which agreement could be computed. To ensure consistency, reviewers followed predefined inclusion criteria and retained only terms whose relevance to family abuse was clear and unambiguous, as described below. Any discrepancies were resolved through discussion until consensus was reached.

Keywords for types of abuse included terms representing different forms of abuse, including emotional, physical, and financial abuse and neglect. Keywords for relational contexts included terms indicating specific familial roles and dynamics, such as *abusive parent*, *toxic household*, or *manipulative sibling*, highlighting interactions between perpetrators and individuals experiencing abuse. Keywords for experience of abuse included phrases reflecting the emotional and psychological toll of abuse, such as *helpless at home*, *scared to go home*, and *living in fear.* Other keywords included terms not covered by the above categories but still relevant to family abuse.

Phase 2: LLM-Based Keyword Expansion

The initial keyword set from the manual review was subsequently expanded in phase 2 using an LLM to capture additional terms and contextual variations. We used ChatGPT (GPT-4o; OpenAI), an LLM, to generate additional keywords relevant to family abuse; the following prompt was used to produce a set of keywords: “Provide a list of comma-separated keywords associated with family abuse including terms related to domestic violence, abusive family structures, and various forms of abuse such as emotional, physical, sexual, and verbal abuse. Focus on terms like abusive parents, sibling abuse, gaslighting, psychological manipulation, family trauma, and the impacts of living in fear or experiencing an unsafe home environment.”

The LLM-generated keywords (not already included in the phase 1 manual set) were validated for contextual relevance by the senior annotator (annotator 1). For each keyword, the first 10 occurrences in the dataset (or all occurrences if fewer than 10) were examined. This sample size was chosen to balance efficiency and representativeness, ensuring that retained keywords accurately reflected typical contexts. A keyword was retained only if at least 70% of its sampled occurrences explicitly referenced family abuse, ensuring high contextual relevance and reliability. The validated keywords were then combined with the initial set from phase 1 to form the final comprehensive, nonredundant keyword list. Note that the LLM was not used to directly classify posts as family abuse–related or nonfamily abuse–related. Instead, it was used only to expand the manually derived keyword list by providing additional candidate terms.

The phase 1 manual review produced 53 initial candidate keywords and phrases. In phase 2, the LLM generated 114 candidate keywords and phrases, of which 26 nonredundant terms were retained after duplicate or overlapping terms were removed and contextual relevance was validated using the 70% threshold. After combining the phase 1 terms with the retained

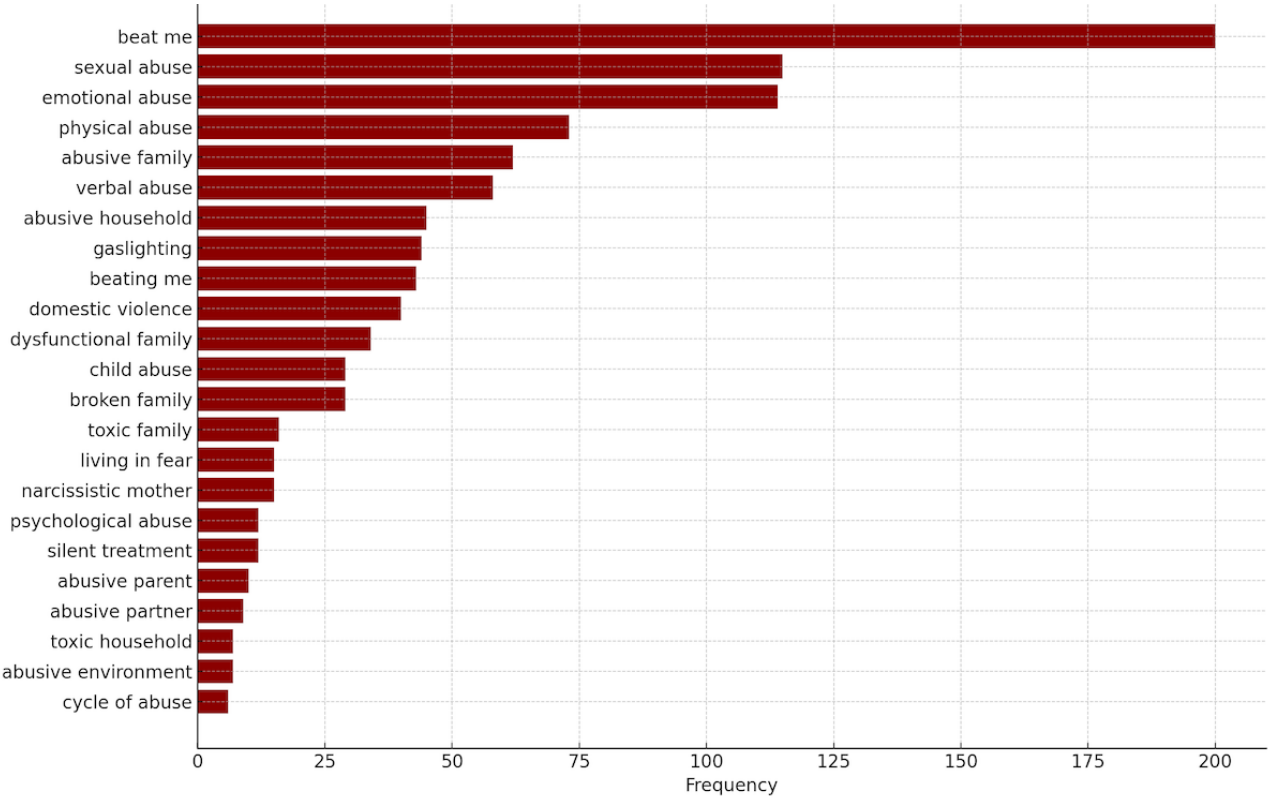
XSL•FO
RenderX

phase 2 terms, the final postselection keyword list contained 79 nonredundant keywords and phrases (presented in Multimedia Appendix 1).

Description of the Final Dataset

Applying the final comprehensive keyword list derived from the keyword selection procedure (refer to the Phase 1 and Phase 2 subsections) to the full corpus, 975 posts were identified explicitly referencing family abuse. The resulting corpus has a mean word count of 730.27 (SD 609.75) per post. It is important to note that while this keyword-based approach captures a broad range of family abuse–related expressions in suicidal posts, it may not identify posts that use indirect, unconventional, or rare language to describe abusive experiences. The top family abuse–related keywords are shown in Figure 2.

Figure 2. Distribution of family abuse–related keywords in the corpus. Frequency indicates the number of posts in which each keyword appears.



Demographic Characteristics of Individuals Reporting Family Abuse

Overview

We examined the demographic characteristics, specifically gender and age, of individuals reporting suicidal ideation in the context of family abuse. To extract age and gender information, a self-reported demographic information extraction tool, SocialDemoExtract, developed by our research group [27], was used. The tool uses a combination of keywords and rule-based patterns to identify demographic information.

Age Distributions

The age distribution of individuals who experienced family abuse was analyzed to understand their representation across age groups. Age groups were categorized as adolescence (aged 13-17 years), young adulthood (aged 18-24 years), early to middle adulthood (aged 25-39 years), and middle-aged adulthood (aged 40-59 years), consistent with the approach used by Wang et al [30]. The other category includes individuals aged <13 years and those aged ≥60 years.

Gender Distributions

Family abuse affects individuals across all genders, though its prevalence and forms vary [31]. To examine gender-related patterns in this dataset, self-reported gender and identity mentions were identified and categorized into three groups: (1) men, (2) women, and (3) people who identify as lesbian, gay, bisexual, transgender, queer or questioning, and other sexual and gender identities (LGBTQ+).

Forms of Family Abuse

Overview

Familial abuse encompasses a range of harmful behaviors, including neglect, emotional manipulation, sexual abuse, and physical violence, all of which have been strongly associated with suicidal ideation in a population-based study by Calder et al [32]. A deeper understanding of the frequency, context, and nuances of abuse, explored through social media discourse analysis, can offer valuable insights into these distributions, thereby complementing clinical and population-based studies and informing the development of targeted, evidence-based interventions. To support this analysis, keywords for each form of abuse were curated using ChatGPT (GPT-4; OpenAI), generating comprehensive lists of formal terms, social media

XSL•FO
RenderX

expressions, slang, and abbreviations. All candidate keywords for the different abuse types were manually validated for contextual relevance before inclusion. For each keyword, the first 10 occurrences in the dataset were examined, and a keyword was retained only if at least 70% of these occurrences clearly represented the intended form of abuse, excluding metaphorical, ambiguous, or unrelated uses. After validation, the final category-specific lists contained 47 physical abuse terms, 53 emotional or psychological abuse terms, 45 sexual abuse terms, and 40 neglect terms; the full lists are presented in Multimedia Appendix 1. The validated category-specific keyword lists were then used to assign abuse-type categories to posts based on the presence of keywords from the corresponding abuse category. Therefore, the resulting abuse-type assignments should be interpreted as scalable keyword-supported categorizations based on manually validated keyword lists, rather than as independently assigned manual post-level labels. This approach was used to enable systematic analysis of a larger corpus while retaining human validation at the keyword selection stage.

Physical Abuse

Physical abuse involves the deliberate use of physical force to inflict harm or injury upon a family member. This category includes actions such as hitting, slapping, punching, beating, and choking, which are frequently used as means of control or intimidation. Some example keywords include *physical abuse*, *hitting*, *slapping*, *punching*, *beating*, *choking*, *bruises*, and *physical assault* (the full list is presented in Multimedia Appendix 1).

Emotional or Psychological Abuse

Emotional or psychological abuse refers to behaviors that undermine an individual's emotional well-being and sense of self-worth. These behaviors include verbal assaults, gaslighting, threats, humiliation, and manipulation, all of which are intended to instill fear, shame, or insecurity in the individual experiencing abuse. Some associated keywords for emotional abuse considered in this study include *emotional abuse*, *psychological abuse*, *verbal assault*, *gaslighting*, *threats*, *humiliation*, *manipulation*, and *belittling* (the full list is presented in Multimedia Appendix 1).

Sexual Abuse

Sexual abuse within the family refers to nonconsensual sexual behaviors or exploitation, including incest, molestation, inappropriate touching, sexual assault, and forced sexual activity. These acts occur without the individual's consent and often involve coercion, manipulation, or the exploitation of trust [33]. Some keywords used to identify such sexual abuse include *sexual abuse*, *molestation*, *incest*, *inappropriate touching*, *sexual assault*, *sexual exploitation*, and *forced sex* (the full list is presented in Multimedia Appendix 1).

Neglect

Neglect occurs when a caregiver fails to provide essential care or resources—such as food, medical attention, emotional support, or shelter—to a dependent family member. This form of abuse can result in long-term physical, emotional, and psychological consequences. Some keywords used to identify neglect include *neglect*, *abandonment*, *lack of food*, *parent neglect*, and *child neglect* (the full list is presented in Multimedia Appendix 1).

Perpetrators of Family Abuse

Overview

We aim to identify which family members—parents, siblings, or intimate partners—are most frequently associated with abuse among individuals experiencing suicidal ideation, with a focus on understanding the relational dynamics involved.

Parents

Parents are commonly identified as perpetrators of abuse within the family context. Research shows that parental abuse often stems from stressors such as financial hardship, mental health challenges, or substance use and can lead to long-term consequences for children, including increased susceptibility to suicidal ideation [34]. Parental neglect, which often co-occurs with abuse, is also associated with developmental delays and elevated mental health risks. The following keywords were used to identify parental figures: *mom*, *dad*, *mother*, *father*, *parents*, *mum*, *mama*, *papa*, *ma*, *pa*, *dada*, and *mummy*.

Intimate Partners

Intimate partner violence (IPV) is a well-established form of abuse that disproportionately affects women, although men can also experience IPV [35,36]. IPV is strongly associated with suicidal ideation, particularly when emotional and physical abuse are persistent and severe [37]. The following terms were used to identify intimate partners: *partner*, *boyfriend*, *girlfriend*, *husband*, *wife*, *spouse*, *significant other*, *lover*, *companion*, *fiancée*, *fiancé*, *gf*, and *bf*.

Siblings

Sibling abuse, though less frequently studied, can significantly contribute to family trauma. It often involves emotional manipulation, bullying, or physical violence and is strongly linked to depressive symptoms and suicidal ideation [38]. The following terms were used to identify siblings: *brother*, *sister*, *sibling*, *bro*, *sis*, *half-brother*, *half-sister*, *stepbrother*, and *stepsister*.

Stepparents

Stepparents often play complex roles in family dynamics, and their involvement in abuse is linked to strained relationships with stepchildren [39,40]. Such abuse may include neglect, emotional abuse, or physical violence, particularly when children feel alienated or rejected. The following terms were used to identify stepparents: *stepfather*, *stepmother*, *stepdad*, and *stepmom*.

Extended Family Members

Extended family members, including grandparents, uncles, aunts, and cousins, can also be involved in family abuse. The following terms were used to identify such perpetrators: *grandfather*, *grandmother*, *grandpa*, *grandma*, *uncle*, *aunt*, *cousin*, *nephew*, *niece*, *in-law*, *brother-in-law*, *sister-in-law*, *mother-in-law*, and *father-in-law*.

XSL•FO
RenderX

Perpetrator Identification and Validation

To identify perpetrators of family abuse, a keyword- and rule-based module was developed to extract potential perpetrator mentions from posts. The module first detects familial terms (eg, *father*, *mother*, *parent*, *sibling*, and *stepparent*) and then searches for family abuse keywords—using the same list applied to identify family abuse–related posts—within a window of ±1 sentence, with sentence boundaries determined by spaCy's (Explosion) sentence splitter. The identified familial term–abuse keyword pairs were flagged as candidate perpetrator mentions and subsequently verified by a senior annotator (annotator 1) to ensure that the familial term explicitly referred to the perpetrator.

Verified candidate perpetrator mentions were then aggregated at the post level so that each perpetrator category was counted at most once per post.

An additional sample-based interannotator agreement analysis was conducted to assess the reliability of the perpetrator verification procedure. A second annotator independently reviewed 617 candidate perpetrator mentions identified by the keyword- and rule-based module from a random sample of 195 posts, representing 20% of the family abuse–related corpus. Each candidate was coded as valid or not valid based on whether the familial term explicitly referred to the perpetrator in context. The annotators agreed on 572 of 617 candidates, including 205 valid and 367 nonvalid cases, yielding 92.71% agreement and a Cohen κ of 0.84. This reliability check indicated high agreement and was used to assess the consistency of the verification procedure; the second annotator's labels were not used to reassign perpetrator categories or recalculate the prevalence estimates reported in the Results section.

Ethical Considerations

No human participants were recruited or directly involved in this study. The study analyzes only publicly available, anonymized Reddit posts related to suicidal ideation, and the original corpus does not contain personally identifiable information. No direct quotes containing identifiable details are presented.

Results

Demographic Characteristics of Individuals Reporting Family Abuse

Table 2 presents the self-reported gender and age distribution of suicidal individuals who reported experiences of family abuse.

Table 2. Self-reported demographic characteristics of family abuse–related suicidal ideation posts.

Characteristic	Posts, n (%)[a]
Gender and/or identity (n=150)	
Men	86 (57.3)
Women	47 (31.3)
LGBTQ+	17 (11.3)
Age (n=313)	
Adolescents (13-17 years)	78 (24.9)
Young adults (18-24 years)	144 (46.0)
Early to mid-adulthood (25-39 years)	78 (24.9)
Middle-aged adults (40-59 years)	6 (1.9)
Other (<13 or ≥60 years)	7 (2.2)

[a]Percentages are calculated within posts containing self-reported information for the corresponding demographic characteristic.

SocialDemoExtract identified self-reported gender information in 150 posts. Among these, 57.3% (n=86) were authored by individuals identifying as men, 31.3% (n=47) by those identifying as women, and 11.3% (n=17) by individuals identifying as LGBTQ+.

SocialDemoExtract shows that, of the 975 posts, 313 contain self-reported age information. Among the 313 posts with self-reported age information, young adults (aged 18-24 years) accounted for the largest proportion, with 144 (46.0%) posts, followed by adolescents (aged 13-17 years) with 78 (24.9%) posts. In contrast, middle-aged adults (aged 40-59 years) and individuals in the other age category (aged <13 or ≥60 years) are underrepresented.

Forms of Family Abuse

Table 3 shows the distribution of familial abuse types identified in suicidal posts, emphasizing their prevalence. Each abuse category is represented by post counts and percentages, while associated keywords are reported using occurrence counts, offering insights into the nature, frequency, and severity of these self-reported experiences.

Physical abuse (683 posts) emerged as the most frequently reported category, with high prevalence of terms such as beat me (146 occurrences) and physical abuse (80 occurrences). Specific expressions such as beating (54 occurrences), hit me (53 occurrences), and kicked (40 occurrences) reflect common violent behaviors.

Emotional or psychological abuse (350 posts) also appeared prominently, with terms such as emotional abuse (119

XSL•FO
RenderX

occurrences) and verbal abuse (62 occurrences) being frequently mentioned. Other expressions, including gaslighting (36 occurrences), yelling (36 occurrences), and insults (15 occurrences), further underscore the psychological harm.

Sexual abuse (296 posts) is primarily represented by the terms sexual abuse (125 occurrences) and rape (93 occurrences), along with molested (25 occurrences) and sexual assault (20 occurrences).

Neglect (126 posts) is reflected in terms such as neglect (63 occurrences), abandoned (35 occurrences), and abandonment (17 occurrences).

Table 4 provides examples of text excerpts from posts illustrating various types of family abuse.

Table 3. Distribution of family abuse types identified in suicidal ideation posts (N=975).[a]

Abuse type	Posts, n (%)
Physical abuse	683 (70.1)
Emotional or psychological abuse	350 (35.9)
Sexual abuse	296 (30.4)
Neglect	126 (12.9)

[a]Abuse type categories were not mutually exclusive; a post could be assigned to more than one category.

Table 4. Example excerpts illustrating the forms of family abuse identified in suicidal ideation posts.

Type of abuse	Example
Physical	"...When I was a kid my mom used to abuse me relentlessly. She would beat me with belts and extension cords, and punch me in the face and tell me I'm stupid, and burned me with cigarettes a few times..."
Emotional or psychological	"...Anyways, I have an emotionally abusive dad, who constantly calls my brothers and I jackasses, shit-head, dumbasses, and he calls me a bitch all the time…"
Sexual	"...I was sexually abused when I was 4-6 years old by my older male cousin..."
Neglect	"...alcoholic enabling father gave me a childhood that was so filled with neglect..."

Perpetrators of Family Abuse

Table 5 presents the post-level distribution of perpetrator categories and associated keywords. A perpetrator category was counted once per post if at least 1 familial term from that category was identified in an abuse-related context and verified as referring to the perpetrator. Therefore, the counts reflect the number of posts mentioning each perpetrator category, not the total number of keyword occurrences.

Parent (435 posts; 435/975, 44.6%; 95% CI 41.52%-47.75%) was a predominant theme, with suicidal individuals most frequently mentioning *mom* (112 occurrences), *mother* (86 occurrences), and *dad* (84 occurrences), followed by *parents* (67 occurrences), *father* (56 occurrences), *parent* (22 occurrences), and *mum* (9 occurrences), reflecting diverse references to parental figures in abusive contexts.

Mentions of sibling abuse (104 posts; 104/975, 10.7%; 95% CI 8.88%-12.76%) most frequently involved *brother* (50 occurrences) and *sister* (32 occurrences), with occasional references to *siblings*, *stepbrother*, or other specific cases.

Partner (37 posts; 37/975, 3.8%; 95% CI 2.77%-5.19%) abuse in intimate partnerships was reflected by mentions of *girlfriend* (13 occurrences), *husband* (11 occurrences), and *wife* (8 occurrences), with less frequent informal references such as *gf* (4 occurrences) and *bf* (1 occurrence).

Abuse by extended family members (34 posts; 34/975, 3.5%; 95% CI 2.51%-4.83%) was most frequently associated with *uncle* (12 occurrences) and *grandparents* (7 occurrences), with occasional mentions of *aunt* (5 occurrences), *cousin* (4 occurrences), *grandpa* (3 occurrences), and *grandma* (3 occurrences).

Stepparent abuse (12 posts; 12/975, 1.23%; 95% CI 0.71%-2.14%) was less frequently mentioned, most often involving *stepdad* (8 occurrences), followed by *stepmother* (3 occurrences) and *stepfather* (1 occurrence).

Table 6 presents example excerpts illustrating the perpetrator categories identified in family abuse–related suicidal ideation posts.

XSL•FO
RenderX

Table 5. Perpetrator categories identified in family abuse–related suicidal ideation posts (N=975).[a]

Perpetrator category	Posts, n (%)	Most frequent associated terms, n
Parent	435 (44.6)	• mom: n=112 • mother: n=86 • dad: n=84 • parents: n=67 • father: n=56 • parent: n=22 • mum: n=9
Sibling	104 (10.7)	• brother: n=50 • sister: n=32 • sibling: n=14 • siblings: n=5 • stepbrother: n=1
Partner	37 (3.8)	• girlfriend: n=13 • husband: n=11 • wife: n=8 • gf: n=4 • bf: n=1
Extended family	34 (3.5)	• uncle: n=12 • grandparents: n=7 • aunt: n=5 • cousin: n=4 • grandpa: n=3 • grandma: n=3
Stepparent	12 (1.2)	• stepdad: n=8 • stepmother: n=3 • stepfather: n=1

[a]Perpetrator categories were counted at the post level, with each category counted at most once per post. Categories were not mutually exclusive; therefore, a post could mention more than one perpetrator category.

Table 6. Example excerpts illustrating perpetrator categories in family abuse–related suicidal ideation posts.

Perpetrator type	Example
Parent	"...My father was an alcoholic as well as *abusive*, after he died when I was 12..."
Sibling	"...My *older brother* also used to *beat me up* when I was little..."
Partner	"...My *boyfriend* started to *control me by gaslighting*, stonewalling, denying my needs, withholding sex and affection..."
Extended family	"...My uncle violently *raped* me from the ages of 7 to 13 while my aunt watched and/or held me down..."
Stepparent	"...My *stepdad* started physically and mentally *abusing* me starting around ages 5 or 6. the physical abuse slowed way down when i hit puberty and he moved to *touching me* instead of hitting me..."

Discussion

Principal Results

Among posts with self-reported age information, young adults and adolescents were the most frequently represented age groups. This suggests that online suicidal ideation posts may offer useful descriptive insight into abuse-related distress among younger users, who may be more inclined to disclose personal experiences in online spaces.

The lower representation of older adults may reflect reduced social media engagement among older populations [41] or a lower likelihood of disclosing sensitive family abuse experiences in online settings. Because demographic information was available only when users self-disclosed it, these findings should be interpreted as patterns in disclosed online narratives rather than as population-level prevalence estimates.

Among posts with available gender and identity information, users identifying as men represented the largest group, followed by users identifying as women and those with LGBTQ+ self-disclosures. This pattern may reflect disclosure behaviors specific to anonymous online communities. The presence of LGBTQ+ self-disclosures suggests that online forums may capture abuse-related suicidal narratives from groups that are particularly vulnerable to family rejection, stigma, or interpersonal stress, although this finding should be interpreted cautiously given that demographic disclosure was incomplete.

Physical abuse was the most frequently identified abuse type, followed by emotional or psychological abuse, sexual abuse,

XSL•FO
RenderX

and neglect. This pattern is consistent with prior research linking physical, sexual, and emotional or psychological abuse to suicidal ideation and suicidal behaviors [9,42]. These findings suggest that family abuse–related suicidal ideation posts frequently reference more than 1 form of abuse, indicating the heterogeneous nature of family abuse described in online mental health narratives.

Parental relationships were the most frequently identified perpetrator category, followed by siblings, intimate partners, extended family members, and stepparents. The predominance of parental perpetrators is consistent with prior research highlighting the central role of parent-child dynamics in family abuse and its association with suicidal ideation [6,11]. The presence of multiple perpetrator categories further indicates that abuse-related suicidal narratives are not restricted to a single family role or relationship type but instead reflect the varied and complex household dynamics in which individuals describe abuse-related distress.

Translational Implications

This study has practical implications for computational mental health research and online support systems. The findings show that suicidal ideation posts involving family abuse often contain identifiable abuse-related terms and references to family relationships.

These features may help researchers organize large collections of online mental health posts into interpretable themes, such as physical abuse, emotional or psychological abuse, sexual abuse, neglect, and perpetrator relationship categories.

The findings may also inform the design of online support resources. For example, posts describing unsafe home environments, parental abuse, or ongoing family violence may require support resources that extend beyond general suicide prevention messaging to include information about family violence, abuse support services, and crisis resources. In this way, aggregate analysis of online narratives can help identify the types of support needs that users disclose in public mental health forums.

These implications should be interpreted cautiously. Because this study provides descriptive, aggregate-level analysis of public online disclosures, the findings should not be interpreted as clinical assessment, diagnosis, or individual-level suicide risk prediction.

Limitations

This study has several limitations that should be considered when interpreting the findings. The analysis relies on self-reported English-language posts from the Reddit *SuicideWatch* forum, where users may omit details, describe experiences selectively, or frame events differently than they would in clinical or survey settings. Social media disclosures may also reflect platform-specific norms, limiting generalizability to other languages, cultural contexts, online communities, or offline populations. In addition, demographic information such as age and gender was available only when users explicitly disclosed it, limiting subgroup-level comparisons and the interpretation of demographic patterns. Finally, this study is descriptive in nature and does not establish a causal relationship between family abuse and suicidal ideation.

Conclusions

This study analyzed 975 suicidal ideation posts involving family abuse to characterize demographic patterns, forms of abuse, and relationships to perpetrators in online mental health narratives. The findings show that family abuse–related suicidal discourse contains diverse abuse-type and relational-context signals, with physical abuse and parental relationships appearing most frequently.

Overall, this work demonstrates the potential of computational social media analysis to support descriptive, hypothesis-generating research on family abuse and suicidal ideation. Future research may examine how family abuse co-occurs with other psychosocial stressors and how these patterns evolve over time in online suicidal ideation narratives.

Funding

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

Data Availability

The raw Reddit posts used in this study were obtained from a publicly available Reddit dataset. The keyword lists used for post selection and abuse-type categorization are provided in Multimedia Appendix 1. Aggregate counts, nonidentifying summary outputs, and analysis code will be made available through the project GitHub repository upon publication. Additional derived materials are available from the corresponding author upon reasonable request.

Authors' Contributions

SS was responsible for writing—original draft, project administration, methodology, investigation, formal analysis, data curation, and conceptualization.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Full keyword lists and supplementary materials, including (1) the complete 79-term postselection keyword list used to identify family abuse–related suicidal ideation posts; (2) the large language model prompts used to generate candidate keywords for each abuse type; (3) the category-specific keyword lists for physical abuse (47 terms), emotional or psychological abuse (53 terms), sexual abuse (45 terms), and neglect (40 terms); and (4) examples of age and gender information extracted using SocialDemoExtract.
[DOCX File , 30 KB - infodemiology_v6i1e90340_app1.docx]

References

1. Suicide. National Institute of Mental Health. URL: https://www.nimh.nih.gov/health/statistics/suicide [accessed 2025-12-25]
2. Suicide. World Health Organization. 2025 Mar 25. URL: https://www.who.int/news-room/fact-sheets/detail/suicide [accessed 2025-12-25]
3. Joiner T. Why People Die by Suicide. Cambridge, MA: Harvard University Press; 2005.
4. Iemmi V, Bantjes J, Coast E, Channer K, Leone T, McDaid D, et al. Suicide and poverty in low-income and middle-income countries: a systematic review. Lancet Psychiatry 2016 Aug;3(8):774-783. [doi: 10.1016/s2215-0366(16)30066-9]
5. Hines DA, Malley-Morrison K, Dutton LB. Family Violence in the United States: Defining, Understanding, and Combating Abuse. Thousand Oaks, CA: SAGE Publications; 2012.
6. Duprey EB, Handley ED, Manly JT, Cicchetti D, Toth SL. Child maltreatment, recent stressful life events, and suicide ideation: a test of the stress sensitivity hypothesis. Child Abuse Negl 2021 Mar;113:104926 [FREE Full text] [doi: 10.1016/j.chiabu.2020.104926] [Medline: 33444849]
7. Charak R, Tromp NB, Koot HM. Associations of specific and multiple types of childhood abuse and neglect with personality pathology among adolescents referred for mental health services. Psychiatry Res 2018 Dec;270:906-914. [doi: 10.1016/j.psychres.2018.11.016] [Medline: 30551343]
8. Dube SR, Anda RF, Felitti VJ, Chapman DP, Williamson DF, Giles WH. Childhood abuse, household dysfunction, and the risk of attempted suicide throughout the life span: findings from the Adverse Childhood Experiences Study. JAMA 2001 Dec 26;286(24):3089-3096. [doi: 10.1001/jama.286.24.3089] [Medline: 11754674]
9. Kwok SY, Chai W, He X. Child abuse and suicidal ideation among adolescents in China. Child Abuse Negl 2013 Nov;37(11):986-996. [doi: 10.1016/j.chiabu.2013.06.006] [Medline: 23899534]
10. Johnson MP. A Typology of Domestic Violence: Intimate Terrorism, Violent Resistance, and Situational Couple Violence. Lebanon, NH: Northeastern University Press; 2008.
11. Kaplan SJ, Pelcovitz D, Salzinger S, Mandel F, Weiner M. Adolescent physical abuse and suicide attempts. J Am Acad Child Adolesc Psychiatry 1997 Jun;36(6):799-808 [FREE Full text] [doi: 10.1097/00004583-199706000-00017] [Medline: 9183135]
12. Pu M, Guo L, Cheng P, Gao Q, Zhu H. Family dysfunction and risk of suicidal behavior in adolescents: a systematic review and meta-analysis. J Affect Disord 2025 Feb 01;370:427-433. [doi: 10.1016/j.jad.2024.11.025] [Medline: 39542110]
13. Angelakis I, Gillespie EL, Panagioti M. Childhood maltreatment and adult suicidality: a comprehensive systematic review with meta-analysis. Psychol Med 2019 Jan 4;49(07):1057-1078. [doi: 10.1017/s0033291718003823]
14. Hari S, Ruch DA, Bridge JA, Brink FW. The evaluation of emotional maltreatment's effect on family dynamics and suicidal behaviors. Child Abuse Negl 2023 Oct;144:106351. [doi: 10.1016/j.chiabu.2023.106351] [Medline: 37515917]
15. Thompson MP, Kingree JB, Lamis D. Associations of adverse childhood experiences and suicidal behaviors in adulthood in a U.S. nationally representative sample. Child Care Health Dev 2019 Jan;45(1):121-128. [doi: 10.1111/cch.12617] [Medline: 30175459]
16. Angelakis I, Austin JL, Gooding P. Association of childhood maltreatment with suicide behaviors among young people: a systematic review and meta-analysis. JAMA Netw Open 2020 Aug 03;3(8):e2012563 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.12563] [Medline: 32756929]
17. Jiang X, Wei Q, Yin W, Pan S, Dai C, Zhou L, et al. Bullying victimization and suicidal ideation among Chinese adolescents: a moderated mediation model of depressive symptoms and perceived family economic strain. BMC Public Health 2025 Jan 30;25(1):393 [FREE Full text] [doi: 10.1186/s12889-025-21579-w] [Medline: 39885401]
18. Jewett PI, Taliaferro LA, Borowsky IW, Mathiason MA, Areba EM. Structural adverse childhood experiences associated with suicidal ideation, suicide attempts, and repetitive nonsuicidal self-injury among racially and ethnically minoritized youth. Suicide Life Threat Behav 2025 Feb;55(1):e13084. [doi: 10.1111/sltb.13084] [Medline: 38651757]
19. Ye Y, Chen B, Zhen R, Li Y, Liu Z, Zhou X. Childhood maltreatment patterns and suicidal ideation: mediating roles of depression, hope, and expressive suppression. Eur Child Adolesc Psychiatry 2024 Nov;33(11):3951-3964. [doi: 10.1007/s00787-024-02442-6] [Medline: 38642118]
20. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
21. Wells I, Thelwell E, Giacco D. Guidance on how to involve people with lived experience in research on digital mental health interventions. Eur Psychiatry 2024 Aug 27;67(S1):S59-S60. [doi: 10.1192/j.eurpsy.2024.171]

XSL•FO
RenderX

22. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. Stroudsburg, PA: Association for Computational Linguistics; 2014:51-60.
23. Morese R, Gruebner O, Sykora M, Elayan S, Fadda M, Albanese E. Detecting suicide ideation in the era of social media: the population neuroscience perspective. Front Psychiatry 2022 Apr 14;13:652167 [FREE Full text] [doi: 10.3389/fpsyt.2022.652167] [Medline: 35492693]
24. Lekkas D, Klein RJ, Jacobson NC. Predicting acute suicidal ideation on Instagram using ensemble machine learning models. Internet Interv 2021 Jul 06;25:100424 [FREE Full text] [doi: 10.1016/j.invent.2021.100424] [Medline: 34401383]
25. Sazzed S. A comparative study of affective and linguistic traits in online depression and suicidal discussion forums. In: HT '23: 34th ACM Conference on Hypertext and Social Media. New York, NY: Association for Computing Machinery; 2023:1-6.
26. Balcombe L, De Leo D. Digital mental health challenges and the horizon ahead for solutions. JMIR Ment Health 2021 Mar 29;8(3):e26811 [FREE Full text] [doi: 10.2196/26811] [Medline: 33779570]
27. Sazzed S, Pial ME, Dehan FN. SocialDemoExtract: a tool for extracting self-reported age and gender from social media text. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). New York, NY: IEEE; 2025.
28. Noh H, Jo Y, Lee S. Keyword selection and processing strategy for applying text mining to patent analysis. Expert Syst Appl 2015 Jun 01;42(9):4348-4360. [doi: 10.1016/j.eswa.2015.01.050]
29. Romano W, Sharif O, Basak M, Gatto J, Preum SM. Theme-driven keyphrase extraction to analyze social media discourse. Proc Int AAAI Conf Web Soc Media 2024 May 28;18(1):1315-1327. [doi: 10.1609/icwsm.v18i1.31391]
30. Wang W, Pei T, Chen J, Song C, Wang X, Shu H, et al. Population distributions of age groups and their influencing factors based on mobile phone location data: a case study of Beijing, China. Sustainability 2019 Dec 09;11(24):7033. [doi: 10.3390/su11247033]
31. Warren A, Blundell B, Chung D, Waters R. Exploring categories of family violence across the lifespan: a scoping review. Trauma Violence Abuse 2024 Apr;25(2):965-981. [doi: 10.1177/15248380231169486] [Medline: 37154576]
32. Calder J, McVean A, Yang W. History of abuse and current suicidal ideation: results from a population based survey. J Fam Violence 2009 Oct 24;25:205-214. [doi: 10.1007/s10896-009-9284-x]
33. Finkelhor D. The prevention of childhood sexual abuse. Future Child 2009;19(2):169-194. [doi: 10.1353/foc.0.0035] [Medline: 19719027]
34. Straus MA, Gelles RJ. Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8,145 Families. Piscataway, NJ: Transaction Publishers; 1991:326-328.
35. Karakurt G, Silver KE. Emotional abuse in intimate relationships: the role of gender and age. Violence Vict 2013;28(5):804-821 [FREE Full text] [doi: 10.1891/0886-6708.vv-d-12-00041] [Medline: 24364124]
36. Hines DA, Malley-Morrison K. Psychological effects of partner abuse against men: a neglected research area. Psychol Men Masculin 2001 Jul;2(2):75-85. [doi: 10.1037/1524-9220.2.2.75]
37. Tjaden P, Thoennes N. Full report of the prevalence, incidence, and consequences of violence against women: findings from the National Violence against Women Survey. National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, and the Centers for Disease Control and Prevention. 2000. URL: https://stacks.cdc.gov/view/cdc/21948 [accessed 2026-08-25]
38. Dantchev S, Hickman M, Heron J, Zammit S, Wolke D. The independent and cumulative effects of sibling and peer bullying in childhood on depression, anxiety, suicidal ideation, and self-harm in adulthood. Front Psychiatry 2019 Sep 24;10:651 [FREE Full text] [doi: 10.3389/fpsyt.2019.00651] [Medline: 31616323]
39. Daly M, Wilson M. The Truth about Cinderella: A Darwinian View of Parental Love. New Haven, CT: Yale University Press; 1999.
40. Alexandre GC, Nadanovsky P, Moraes CL, Reichenheim M. The presence of a stepfather and child physical abuse, as reported by a sample of Brazilian mothers in Rio de Janeiro. Child Abuse Negl 2010 Dec;34(12):959-966. [doi: 10.1016/j.chiabu.2010.06.005] [Medline: 21030083]
41. 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 [FREE Full text] [doi: 10.1186/s12877-023-04437-5] [Medline: 38012565]
42. Bahk YC, Jang SK, Choi KH, Lee SH. The relationship between childhood trauma and suicidal ideation: role of maltreatment and potential mediators. Psychiatry Investig 2017 Jan;14(1):37-43 [FREE Full text] [doi: 10.4306/pi.2017.14.1.37] [Medline: 28096873]

Abbreviations

ACE: adverse childhood experience
IPV: intimate partner violence
LGBTQ+: lesbian, gay, bisexual, transgender, queer or questioning, and other sexual and gender identities
LLM: large language model

XSL•FO
RenderX

RQ: research question

Edited by T Mackey; submitted 25.Dec.2025; peer-reviewed by E Zeru, Z Xie; comments to author 21.Apr.2026; revised version received 20.Jun.2026; accepted 08.Jul.2026; published 14.Sep.2026.

Please cite as:
Sazzed S
Characterizing Family Abuse in Suicidal Ideation Posts on Reddit: Large Language Model–Assisted Content Analysis
JMIR Infodemiology 2026;6:e90340
URL: https://infodemiology.jmir.org/2026/1/e90340
doi:10.2196/90340
PMID:

©Salim Sazzed. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 14.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Monitoring Canada's Health Care Polycrisis Through News Headline Analysis: Infodemiology Study

John Weirstrass Muteba Mwamba[1], BSc, MA, MCom, PhD; Vijay Mago[1], BSc, MSc, PhD

School of Health Policy and Management, Faculty of Health, York University, Toronto, ON, Canada

Corresponding Author:
John Weirstrass Muteba Mwamba, BSc, MA, MCom, PhD
School of Health Policy and Management
Faculty of Health
York University
Stong College Building, Room 208B
4700 Keele Street
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 736 2100
Email: jwmm@yorku.ca

Abstract

Background: The health care polycrisis in Canada comprises several concurrent crises, namely workforce shortages; food insecurity; housing stress; and health inequities affecting Indigenous peoples, low-income families, and racialized communities. A scalable infoveillance system is needed to monitor these related but distinct crises concurrently using natural language processing (NLP) and statistical forecasting.

Objective: This study applied a multimethod NLP pipeline to characterize discourse dynamics, causal interdependencies among crisis themes, and the responsiveness of policy-making to those dynamics within Canada's health care system from 2020 to 2025.

Methods: We used an NLP pipeline to examine the language and sentiment of 6757 online news headlines regarding Canada's health care system over approximately 5 years (2020-2025), using 2 methods for sentiment analysis: a fine-tuned BERT (Bidirectional Encoder Representations from Transformers) model (test accuracy of 0.97) and VADER (Valence Aware Dictionary and Sentiment Reasoner) for valence and subjectivity. We used latent Dirichlet allocation to identify topics within the headlines, then applied time series analysis, using both ARIMA (autoregressive integrated moving average) and ETS (error, trend, and seasonality) methods, to predict trends in each topic. Finally, we used Granger causality testing to determine whether past values of one theme could predict future values of another.

Results: Seven major discourse topics emerged, of which 5 were retained for the surveillance analysis. The health care system crisis theme was the most prevalent (1200/6757, 17.8% of all headlines), followed by economic determinants of health (1017/6757, 15.1%), housing as a health determinant (958/6757, 14.2%), food insecurity health impact (929/6757, 13.7%), and social inequality health effects (699/6757, 10.3%). Strong statistical evidence showed that the volume of discourse around food insecurity health impact was temporally predictive of subsequent discourse volume around housing as a health determinant ($F_{1,54}$=17.361; false discovery rate–adjusted P=.002). Sentiment analysis showed no statistically significant change after any policy intervention (Cohen d <0.2), and the 12-month forecasts were flat or slightly lower for 4 of the 5 retained themes and rose only for health care system crisis (all 95% CIs included zero).

Conclusions: These findings illustrate the complexity of issues currently facing Canada and suggest that policymakers should consider implementing NLP-based infoveillance systems to monitor and analyze complex relationships between these crises. Further, our findings suggest a persistent policy perception gap with significant implications for designing health informatics surveillance systems.

(JMIR Infodemiology 2026;6:e99655) doi:10.2196/99655

KEYWORDS

infoveillance; Granger causality; health care policy; natural language processing; polycrisis; sentiment analysis

XSL•FO
RenderX

Introduction

Global health care systems face an unprecedented set of interrelated, simultaneous, and interactive crises. These crises include severe workforce shortages; economic instability; persistent and worsening social and economic disparities; and demographic pressures. These health care crises can no longer be viewed as discrete epidemiological events developing in isolation [1,2]. Rather, they represent part of a larger polycrisis: a type of event where multiple, systemic crises interact with one another and produce a total impact much greater in magnitude than if the individual crises were analyzed independently in isolation [3,4].

An important characteristic of a polycrisis is the development of upward-downward cascade failures, that is, when a major crisis develops in a key area (eg, housing or food insecurity), the downward trajectory in the population's health status and the burden placed upon the health care delivery system will accelerate dramatically [5].

Canada's health care system is currently experiencing a polycrisis. Canada faces 3 main, concurrent crises simultaneously: unprecedented levels of housing stress; record levels of food insecurity; and severe and growing workforce shortages within the health care system [6-8]. Each of these crises also results in disproportionate negative effects on Indigenous populations, low-income families, and racialized communities—contributing to historical health inequities [9]. Moreover, despite the obvious linkages between these crises, the majority of responses to date have consisted of independent policy actions across separate crisis themes without consideration for the causal relationships between the socioeconomic determinants of the population's health and the health care services provided.

To manage a health care polycrisis effectively, we need to go beyond traditional epidemiologic surveillance methods used by virtually all nations. Traditional surveillance approaches are generally disease-specific and rely on historical data.

The unstructured text generated from online digital discourse, including that derived from news media, contains a vast amount of information that reflects the interaction of the various crisis themes in real time [10]. While natural language processing (NLP) techniques, specifically topic modeling and sentiment analysis using transformer models (such as BERT [Bidirectional Encoder Representations from Transformers]), have demonstrated success in identifying public health indicators [11,12], there is a substantial lack of evidence demonstrating their application to monitor the health care polycrisis in Canada. Virtually all previous infoveillance (in this study, we use the term infoveillance to refer specifically to the use of publicly available digital data, including online news and search behavior [10], to monitor population-level health trends, and we use this term consistently rather than the related but distinct terms digital surveillance and infosurveillance) studies in Canada have focused almost exclusively on single events (eg, the COVID-19 pandemic) [13] and not on the comprehensive network of food insecurity, housing, financial difficulties, and systemic inadequacies in the delivery of health care simultaneously.

Therefore, a sizeable gap in the literature exists: a robust, multimethod infoveillance tool capable of monitoring the entire interconnected health system crises in Canada is nonexistent. In particular, the causal structure linking public discourse associated with these different crisis themes has not been quantitatively modeled, and neither has the public sentiment toward health care been statistically correlated to the timing of policy actions taken to address the health care polycrisis.

This study seeks to fill this void by creating a scalable multimethod NLP pipeline to identify the causal relationship between public discourse and the various components of the health care polycrisis in Canada between 2020 and 2025. We propose to achieve this objective through the application of a variety of complementary analytical tools, including transformer-based sentiment analysis; latent Dirichlet allocation (LDA)–based topic modeling; Granger causality-based econometric testing for causality; and ARIMA (autoregressive integrated moving average)/ETS (error, trend, and seasonality)–based statistical forecasting. Through our proposed pipeline, we plan to model the public discourse in Canadian online news headlines. Ultimately, our objective is to predict the future trends of the critical crisis themes and evaluate the responsiveness of the policy-making process to the evolving online public discourse surrounding the delivery of health care.

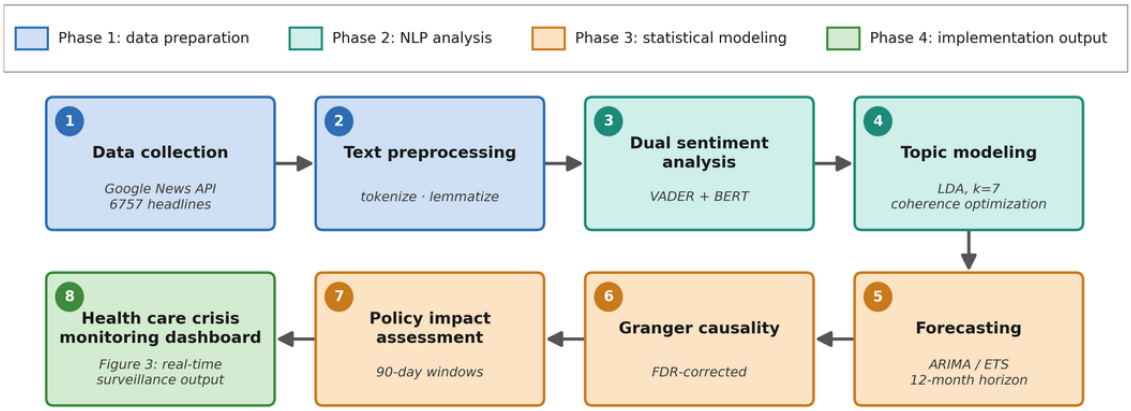
Methods

Overview

The proposed health care infoveillance tool comprised 4 sequential analytical phases shown in Figure 1.

XSL•FO
RenderX

Figure 1. The multimethod NLP digital infoveillance pipeline, with each phase of analysis represented by a different color. Blue represents the data preparation phase; teal represents both sentiment and topic modeling phases within the NLP analysis phase; orange represents forecasting, causal network analysis, and policy impact assessment within the statistical modeling phase; and green represents implementation outputs. Numbered boxes indicate the processing order. The ultimate goal of this study is to develop an analytical health care crisis monitoring prototype (see Figure 3). ARIMA: autoregressive integrated moving average; BERT: bidirectional encoder representations from transformers; ETS: error, trend, and seasonality; FDR: false discovery rate; LDA: latent Dirichlet allocation; NLP: natural language processing; VADER: Valence Aware Dictionary and Sentiment Reasoner.



Data Collection and Preprocessing

We created a digital infoveillance tool for health care through the use of publicly available online news headlines. Eysenbach [10] demonstrated that the number of news headlines is large in scale and can be used as an indicator of the most prominent issues or concerns of a population. For our analysis, we accessed online news headlines using the Google News API (GoogleNews Python library), with 12 different search terms over 5 different crisis themes: housing as a determinant of health, health care system collapse, food insecurity, financial hardship, and social inequality. The time frame for which data was collected began on October 17, 2020, and ended on September 28, 2025. In total, there were 6757 unique English-language headlines posted online by Canadian news media outlets. Each of the records included information regarding the headline itself, the date it was published online, the name of the publishing outlet, and the terms that generated the headline.

The 12 search terms that produced the corpus were the following: “Canada polycrisis,” “Canada’s economic crisis,” “Canada’s affordability crisis,” “Canada’s healthcare crisis,” “Collapse of the Canadian Healthcare System,” “Crisis in Ontario Healthcare,” “Privatizing Healthcare in Canada,” “Income Inequality in Canada,” “Food Insecurity in Canada,” “Racialized Communities’ Access to Healthcare in Canada,” “Affordability Crisis in Housing in Canada,” and “Homeless Crisis in Canada,” all limited to English-language articles and to the Canada region. The data were sourced from 259 unique news outlets; although there was an evident concentration of reporting, no one outlet represented more than half of all media content. Headlines from CBC (Canadian Broadcasting Corporation) made up 13.3% (897/6757) of the total headlines, while headlines from Global News made up 4.7% (320/6757), those from the Fraser Institute made up 3.7% (248/6757), those from the Financial Post made up 2.6% (174/6757), and those from The Conversation made up 2.5% (170/6757). Together, these top 5 outlets accounted for approximately 26.8% (1809/6757) of all headline content, while the top 10 outlets collectively accounted for 37.8% (2554/6757) of headline content. Although the retrieval of headlines via Google’s News aggregation and ranking algorithms may be less prone to sampling error than using either a truly random or a comprehensive set of Canadian news stories, this approach nonetheless introduces its own potential sources of sampling bias [14]. These sources include the composition of outlets included in the corpus and the coverage of the search terms used in the retrieval process.

A typical NLP pipeline was applied for preprocessing. In addition to converting all characters in each string of text to lowercase, all HTML tag elements were eliminated, as well as URLs and special characters. Each string of text was then broken down into individual words through the use of Natural Language Toolkit (NLTK). Using WordNet, the root form (or base) of each word (ie, “running” becomes “run”) was identified. As such, common English words (stopwords) are typically eliminated in an effort to remove noise from data. However, in this case, it is possible some of those English stopwords would be important to include because they were related to themes such as health or health care, Canada or Canadian, systems, and crises. All strings with fewer than 3 characters were removed from the dataset.

This study used only publicly available news headline titles collected from the Google News API. There were no humans involved in this study, nor were there biological samples or personally identifiable information. Therefore, Ethics Board Review was not needed at York University.

Sentiment Surveillance

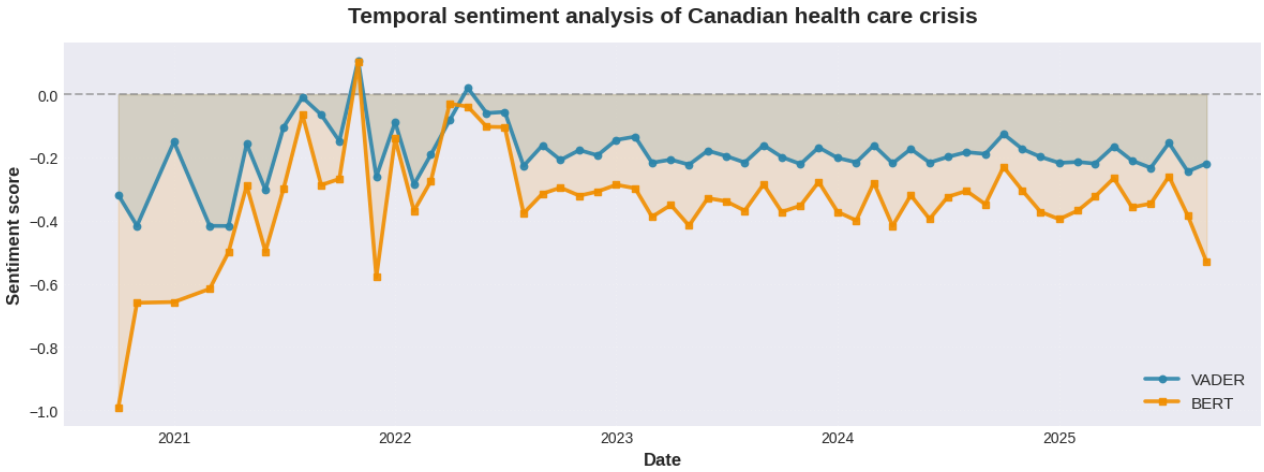
VADER (Valence Aware Dictionary and Sentiment Reasoner) [15] was run on all 6757 raw online news headline data points to create compound scores ranging from –1 to +1. The headlines were then categorized based on the VADER scores. If the VADER score was greater than or equal to 0.05, it was considered a positive headline. If the VADER score was less than or equal to –0.05, it was considered a negative headline. All other headlines were categorized as neutral. The monthly

XSL•FO
RenderX

aggregation of these scores formed the VADER surveillance series.

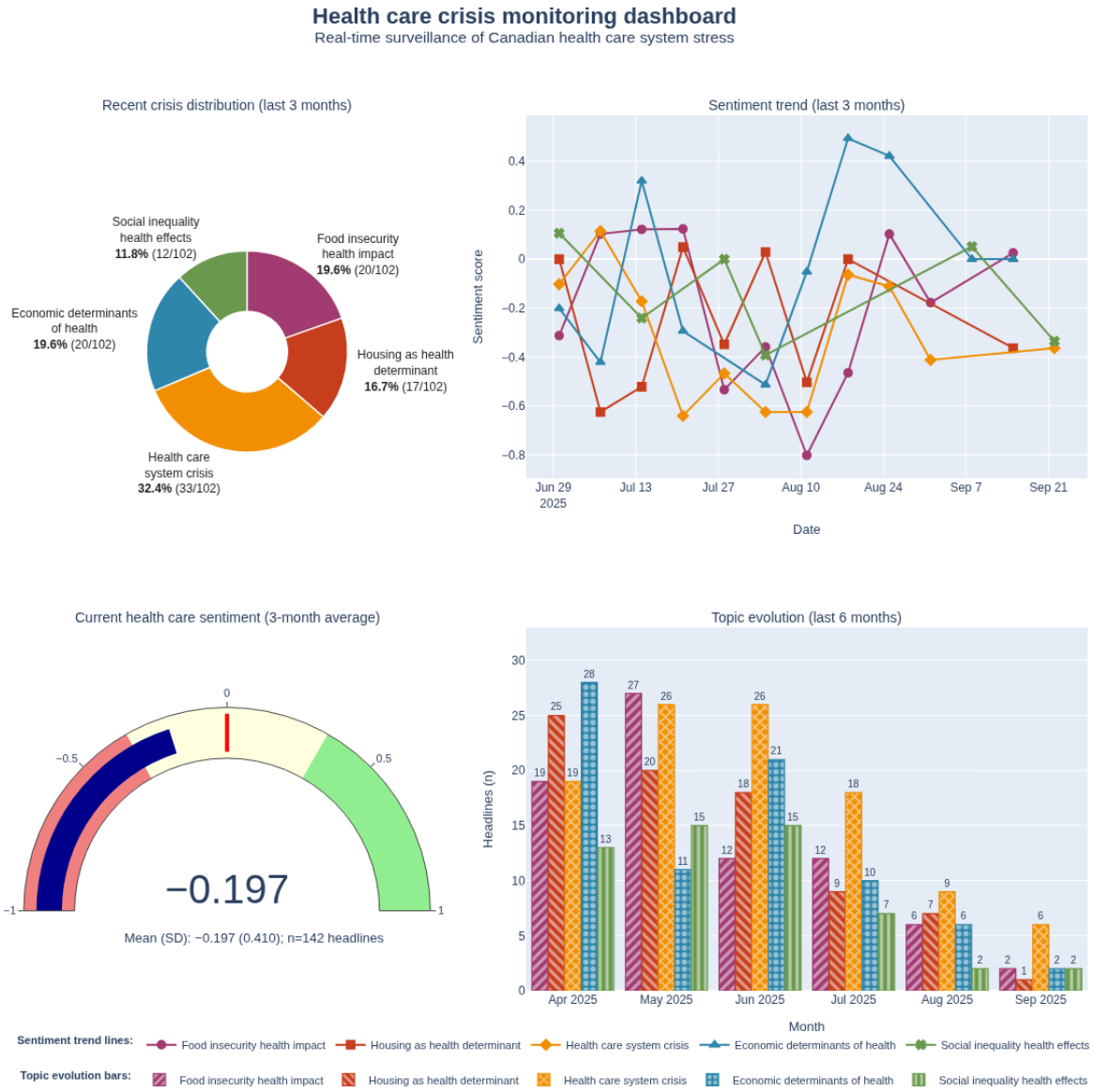
The BERT-base-uncased [16] was fine-tuned for 3-class sentiment classification. To enhance computational efficiency, a stratified random sample of 3000 headlines was sampled from the total 6757-headline dataset. The stratified random sample was then divided into an 80% (2400/3000) holdout training set and a 20% (600/3000) held-out testing set. Stratification by class label was carried out at both partitions to preserve the original distribution of class labels. The original class labels for each headline were assigned using the same VADER classification models that were originally trained. Following fine-tuning, the model was run again but in inference mode over all 6757 headlines in the entire dataset. This run through the entire dataset generated the BERT sentiment scores shown in Figure 2 and the monitoring dashboard (Figure 3), and the evaluated model agreement with VADER. The hyperparameters chosen included 2 epochs, a batch size of 16, a learning rate of 2×10^{-5}, the AdamW optimizer, and a max sequence length of 128. These hyperparameters are within acceptable ranges recommended for the BERT-base-uncased model on downstream classification tasks [16]. It is worth noting that a formal hyperparameter search was not conducted, and this is noted as a limitation. A continuous sentiment score was calculated as the difference in positive and negative sentiment probabilities from the softmax output.

Figure 2. VADER and BERT monthly sentiment scores for Canadian health care polycrisis (October 2020 to September 2025). Both time series display consistent negative orientations and periodic peaks in negative intensity. Shaded areas extend from each series to zero. BERT: bidirectional encoder representations from transformers; VADER: Valence Aware Dictionary and Sentiment Reasoner.



XSL•FO
RenderX

Figure 3. The prototype health care crisis monitoring dashboard that displays all monitoring results. At the top left, there is a donut chart illustrating the proportion of each crisis theme in the latest 3 months. At the top right, there are line charts of sentiment trends for each theme in the latest 3 months, with consistent colors used in the donut chart. The gauge indicator at the bottom left shows the rolling sentiment score averaged over the last 3 months (–0.197, SD 0.410; n=142 headlines), which is firmly negative. The bar chart at the bottom right shows how frequently each topic has been discussed each month over the last 6 months, with the aim of identifying topics with increasing discussion. This dashboard demonstrates that our pipeline can be used as a near-real-time health informatics surveillance tool.



Topic Modeling

LDA [17] was applied to the preprocessed corpus. The vocabulary was filtered so that only those terms that appeared in more than 5 and fewer than half of the documents were included. A range of topic numbers (k=3, ..., 7) was tested to determine which number of topics provided the highest value of c_v coherence. The model was run with a total of 20 iterations, with the hyperparameters set to automatically estimate the Dirichlet parameters, and the random seed was set to 42. For each headline, the most likely topic was identified and assigned to it. The monthly topic frequencies were created [12].

First, we quantitatively evaluated the topic coherence using the c_v coherence measure [18]; the common values for coherence evaluation that have been reported in the literature when evaluating short informal texts (such as headlines, X [X Corp] posts, or forum postings) typically fall into the 0.4 to 0.6 range. The reason behind these lower ranges can be attributed to the fact that shorter texts provide fewer word co-occurrences relative to longer texts (for example, full articles or abstracts), which generally produce higher c_v measures. The c_v measure calculated for this dataset (c_v=0.453) fits within the established acceptable range. In addition to representing some degree of true thematic overlap (as indicated by the presence of "care" among the keyword lists for more than one topic), this relatively low c_v score likely represents a portion of theme overlap.

The LDA was chosen over embedding-based alternatives such as BERTopic [19] and Top2Vec [20] due to its ability to

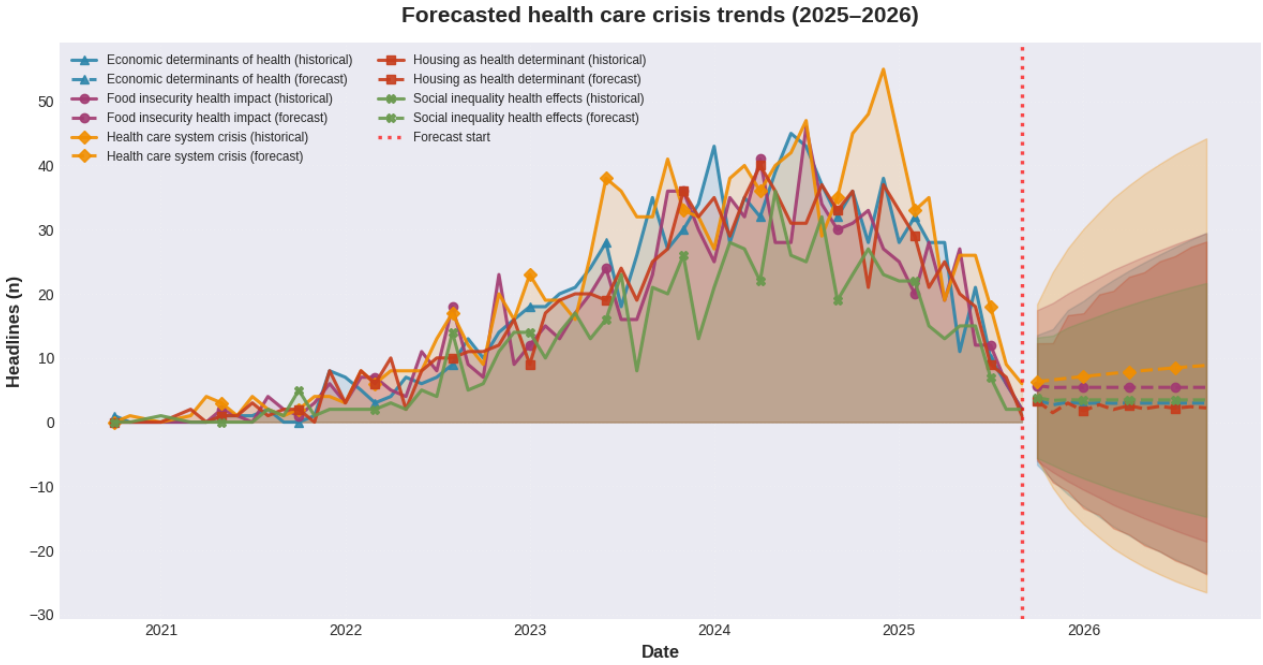
XSL•FO
RenderX

generate fully interpretable word-topic and document-topic probability distributions, as well as its deterministic nature provided that a fixed random seed is used. This combination of features is important in the context of a surveillance system that needs to be auditable by health policy stakeholders.

Time Series Forecasting

Each of the topic frequency time series was tested for stationarity using the augmented Dickey-Fuller test. Time series that failed the augmented Dickey-Fuller test were first-differenced to make them stationary. For each of the 5 primary topics, two models were compared: an ARIMA(1,1,1) and an ETS (additive trend and seasonal) model with period 12. The performance of each of these models was evaluated for the final 6 months of data available using both mean absolute error and root-mean-square error. The 12-month forecasts with 95% CIs reported in Figure 4 were generated with the ARIMA(1,1,1) model for all 5 topics; the ETS model was used only for the accuracy comparison.

Figure 4. Historical monthly frequency of discourse and 12-month forecasts (ARIMA(1,1,1) model) with 95% CIs. Forecasts are flat or slightly lower for 4 topics and rise for health care system crisis; all 95% CIs include zero. ARIMA: autoregressive integrated moving average.



Granger Causality Analysis

For each of the 20 pairs of primary topics, pairwise Granger causality tests [21,22] were carried out on the topic frequency time series from lag 1 through lag 6. Since our null hypothesis is that the topic frequency time series do not provide predictive information to each other, we are dealing with a multiple testing problem. Therefore, we used the corrected *P* values to account for this issue using Benjamini-Hochberg false discovery rate (FDR) at α=.05 [23]. In addition, all of our Granger causality results were operationalized to indicate whether or not prior discussion in 1 area provided statistically significant predictive information relative to another area (beyond what has already been observed within the prior discussion of the other area).

Policy Intervention Analysis

To evaluate the impact of 8 policy interventions on crisis themes such as economic support, food security, health care privatization, workforce expansion, and housing-health integration, we examined the topic frequency time series for each intervention. We did this by calculating the symmetric difference of the topic frequency time series before and after each policy intervention for each type of intervention. To evaluate whether the changes in the topic frequency time series were due to chance, 2-tailed independent-samples *t* tests, Cohen d effect sizes, and 1000 resample bootstrap 95% CIs were calculated. To assess how long a policy intervention could sustain a positive sentiment, we defined sentiment sustainability as the amount of time that a topic frequency time series could stay above 0.05 units of the baseline (max 180 days). As with the Granger causality analysis, we also had multiple hypotheses to test (8 policy interventions) and therefore applied FDR correction to each of the tests. A full description of all 8 interventions, including type, date, rationale, and implementation score, is provided in Multimedia Appendix 1.

Predictive Feature Analysis

To predict negative sentiment based on the features calendar month, year, and headline text length, a random forest classifier (100 trees, max depth 10, random seed 42) was trained. The importance of each feature was measured as the average reduction in Gini impurity when splitting the training data based on that feature. All analyses were carried out in Python (Python Software Foundation) using *pandas*, *NumPy*, *NLTK*, *Gensim*, *statsmodels*, *scikit-learn*, *PyTorch*, and *Hugging Face Transformers*. The random seed was applied worldwide, including to the stratified train or validation splits that were performed before BERT fine-tuning, and to both stages of LDA topic modeling. The ARIMA and ETS models, the augmented Dickey-Fuller test, the Granger causality tests, and the 2-tailed

XSL•FO
RenderX

independent samples *t* test all use deterministic methods and therefore have no randomness or stochastic components; they required no seed. Additionally, we ran a robustness check on our Random Forest feature importance scores using Shapley additive explanations (SHAP) values [24] via a TreeExplainer since it has been shown that Gini Impurity is biased toward continuous and high cardinality features [25]. These results are presented in the Predictive Feature Analysis subsection of the Results section.

Ethical Considerations

Our study consists solely of publicly available news headline titles. As such, we neither included nor referenced human participants, biological samples, nor personal identifying information. Consequently, ethics committee approval and/or participant consent were not necessary.

Results

Sentiment Surveillance

The fine-tuned BERT classifier performed with an accuracy of 0.97 (weighted F_1-score=0.97) on the held-out test set (n=600) drawn from the 3000-headline stratified subsample described in the Methods section. On this test set, for the negative class, precision, recall, and F_1-scores were all equal to 0.99; for the neutral class, precision was 1.00, recall was 0.93, and F_1-score was 0.96; for the positive class, precision was 0.93, recall was 0.98, and F_1-score was 0.95. Both VADER and BERT models indicate a generally negative monthly sentiment trajectory from 2020 to 2025, with dips in sentiment occurring during periods of crisis and a similar month-to-month pattern in the 2 models (r=0.81 across the 58 monthly means), although BERT scores were lower than VADER scores in 57 of the 58 months (mean difference –0.15), as shown in Figure 2.

Topic Modeling: Thematic Structure of the Polycrisis

Using LDA, 7 topic clusters were chosen based upon a model fit (c_v=0.453). Of these 7 topics, 5 could be substantively interpreted to form the core themes used for surveillance purposes; however, an additional two—affordable housing policy deficit (996/6757, 14.7%) and Canadian housing cost burden (958/6757, 14.2%)—represented overlapping themes associated with Canada's housing crisis. These 2 topics have been removed from the causal analysis.

Table 1 reports that health care system crisis was the most common topic at 17.8% (1200/6757) of all headlines, followed by economic determinants of health at 15.1% (1017/6757); housing as a health determinant at 14.2% (958/6757), food insecurity and health impact at 13.7% (929/6757), and social inequality and health effects at 10.3% (699/6757). Health care system crisis had been covered more than any other single theme in the sample corpus we collected; yet, when we look at the 4 upstream social-determinant themes (economic determinants of health, housing as health determinant, food insecurity health impact, and social inequality health effects) together they account for the majority of all headlines (3603/6757, 53.3%), indicating that while news outlets in Canada are covering some of the immediate manifestations of this crisis in their coverage of health care system crisis; they are also framing much of the coverage through the lens of its drivers, which are identified by the Granger causality analysis reported in the Granger Causality Network subsection of this section.

Table 1. Emergent themes from health care polycrisis discourse, identified through LDA[a] (c_v=0.453, k=7). This depicts all 7 topics. The first 5 rows represent the main surveillance topics, with the last 2 rows reflecting the intersecting threads of Canada's housing crisis, which we omitted from causal analyses.

Topic theme	Representative keywords	Values, n (%)
Food insecurity health impact	food, insecurity, care, economy, recession	929 (13.7)
Housing as health determinant	care, Ontario, recession, Ford, public	958 (14.2)
Health care system crisis	Ontario, care, hospital, privatization, doctor	1200 (17.8)
Economic determinants of health	care, housing, income, inequality, could	1017 (15.1)
Social inequality health effects	polycrisis, housing, global, say, care	699 (10.3)
Affordable housing policy deficit	housing, care, new, government, affordability	996 (14.7)
Canadian housing cost burden	housing, economic, care, say, Canadian	958 (14.2)
Total (all topics)	N/A[b]	6757 (100.0)

[a]LDA: latent Dirichlet allocation.

[b]N/A: not applicable.

Time Series Forecasting

The results of the augmented Dickey-Fuller tests indicate that all 5 crisis themes we retained were found to be nonstationary (the values of the p-statistic, which ranged from 0.055 to 0.541). However, upon taking the first difference of each nonstationary time series, the resulting series were transformed into stationary time series. All time series received the same uniform treatment based on their p-statistics' proximity to the 0.05 threshold value.

Therefore, it is assumed that this rule-based methodology reduced the potential for subjective judgments regarding stationarity based on individual cases. Regarding performance metrics for model fit, as shown in Table 2, the ETS model outperformed the ARIMA model for 3 crises: economic determinants of health, housing as a health determinant, and

XSL•FO
RenderX

social inequality health effects. Conversely, the ARIMA model performed better than the ETS model for 2 crises: health care system crisis and food insecurity health impact. Figure 4 exhibits the historical monthly frequency of each theme and the 12-month forecasts with 95% CIs.

Table 2. Time series model performance by discourse theme (MAE[a] and RMSE[b] on the final 6 months of observed data).

Theme and model	MAE	RMSE
Economic determinants of health		
ARIMA[c]	16.53	18.72
ETS[d,e]	11.77	13.14
Health care system crisis		
ARIMA[e]	20.26	21.71
ETS	27.21	28.34
Food insecurity health impact		
ARIMA[e]	13.32	15.35
ETS	16.79	19.20
Housing as a health determinant		
ARIMA	12.17	14.59
ETS[e]	6.83	8.51
Social inequality health effects		
ARIMA	10.64	12.00
ETS[e]	6.66	7.62

[a]MAE: mean absolute error.

[b]RMSE: root-mean-square error.

[c]ARIMA: autoregressive integrated moving average.

[d]ETS: error, trend, and seasonality.

[e]Best-performing model for that theme.

Granger Causality Network

The FDR-corrected Granger causality results reported in Table 3 below show that food insecurity and its health effects discourse were Granger-caused by the economic determinant of health discourse ($F_{1,54}$=7.60, FDR P value=.024).

This relationship supports an upstream ordering where coverage of economic hardship is first followed by coverage of food insecurity and then followed by coverage of housing as a health determinant and social inequality health effects. It can therefore be interpreted as an association in discourse over time. However, it would be incorrect to infer that economic hardship causes food insecurity and/or housing stress in the population. The results also show that health care system crisis discourse Granger-caused both the housing as health determinant discourse ($F_{1,54}$=5.95, FDR P value=.047) and the social inequality health effects discourse ($F_{1,54}$=6.16, FDR P value=.043). Of these 5 relationships, 2 (food insecurity health impact to housing as health determinant; economic determinants of health to food insecurity health impact) retained their FDR-significance across all 6 lags, while the other 3 either diminished or became insignificant at some of the later lags, indicating that the strength of the causal signal for each of those relationships is greatest in the immediate period. Figure 5 shows the complete network of significant relationships across lags 1 to 6.

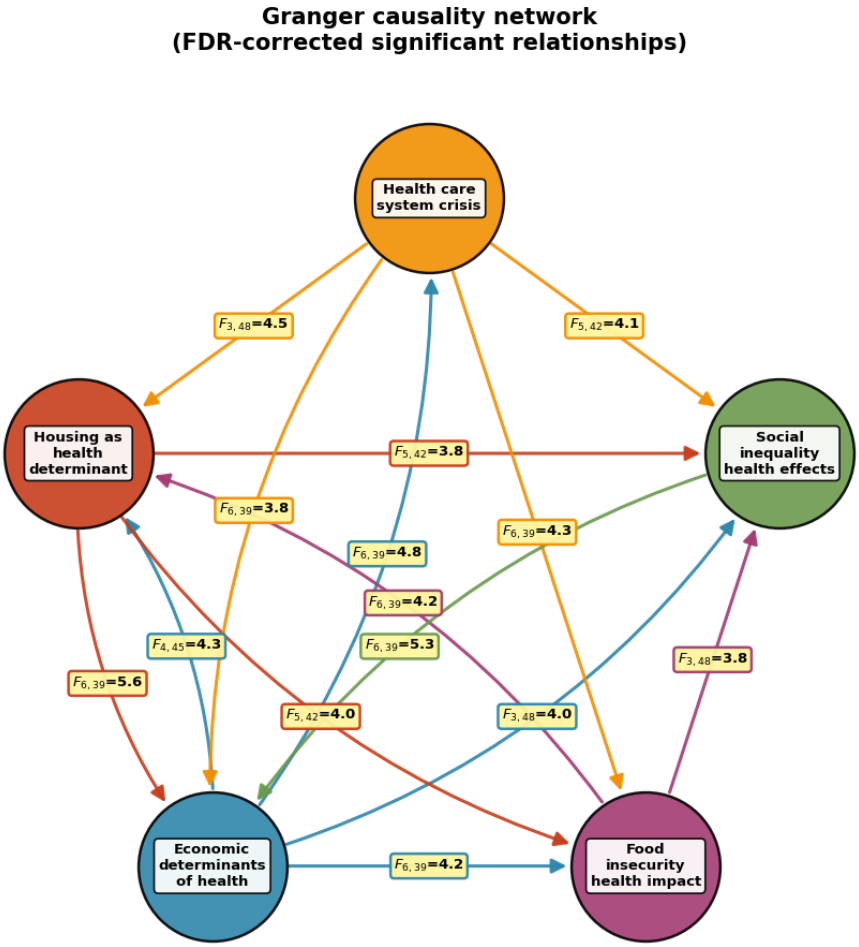
Table 3. FDR[a]-corrected Granger causality results (lag 1). All listed relationships are significant at α_{FDR}=.05.

Cause theme	Effect theme	F test (df)	FDR P value
Food insecurity health impact	Housing as health determinant	17.361 (1, 54)	.002
Food insecurity health impact	Social inequality health effects	9.380 (1, 54)	.015
Economic determinants of health	Food insecurity health impact	7.601 (1, 54)	.024
Health care system crisis	Social inequality health effects	6.164 (1, 54)	.043
Health care system crisis	Housing as health determinant	5.954 (1, 54)	.047

[a]FDR: false discovery rate.

XSL•FO
RenderX

Figure 5. The Granger causality network represents all of the relationships between each of the themes identified within the Canadian health care polycrisis. Arrows point from the Granger-causing theme to the predicted theme; each label gives the F statistic with its degrees of freedom at the highest significant lag. In contrast to the results reported in Table 3 (which only included results for lag=1), Figure 5 includes every possible relationship between 2 themes that were statistically significant using an FDR-corrected F-test at some level of lag from 1 to 6. Therefore, while Figure 5 shows the directional relationships at specific levels of lag, it does represent a summary of the entire network over time rather than being limited to just one single level of lag. The node with the most edges connecting other nodes in Figure 5 is economic determinants of health (7), followed by housing as a health determinant (6); the other 3 themes have 5 each. FDR: false discovery rate.



Policy Intervention Analysis

No intervention achieved statistical significance after applying FDR adjustment, with all interventions resulting in very small effect sizes (Cohen d <0.2). Economic support interventions resulted in the greatest absolute change (–0.056), with a small effect size (–0.128), and the most long-lasting impact (28 days). The only intervention type to cause an increase in positive sentiment was health care privatization. It produced a small absolute change (0.028) and a small effect size (0.064). It lasted only 1 day before returning to the baseline level. Workforce expansion produced almost no change (–0.004), with a negligible effect size (–0.010), despite having a higher implementation score (0.60) than Economic Support Interventions (0.20). The results are summarized in Table 4.

This can also be observed in Figure 6, where each colored vertical line represents a given policy intervention date overlaid on the continuous VADER and BERT sentiment scores. There is no observable change in trajectory in either sentiment score at any of these points. The baseline sentiment was continuously negative (mean VADER compound score –0.19, SD 0.43, across all 6757 headlines in the full sample period).

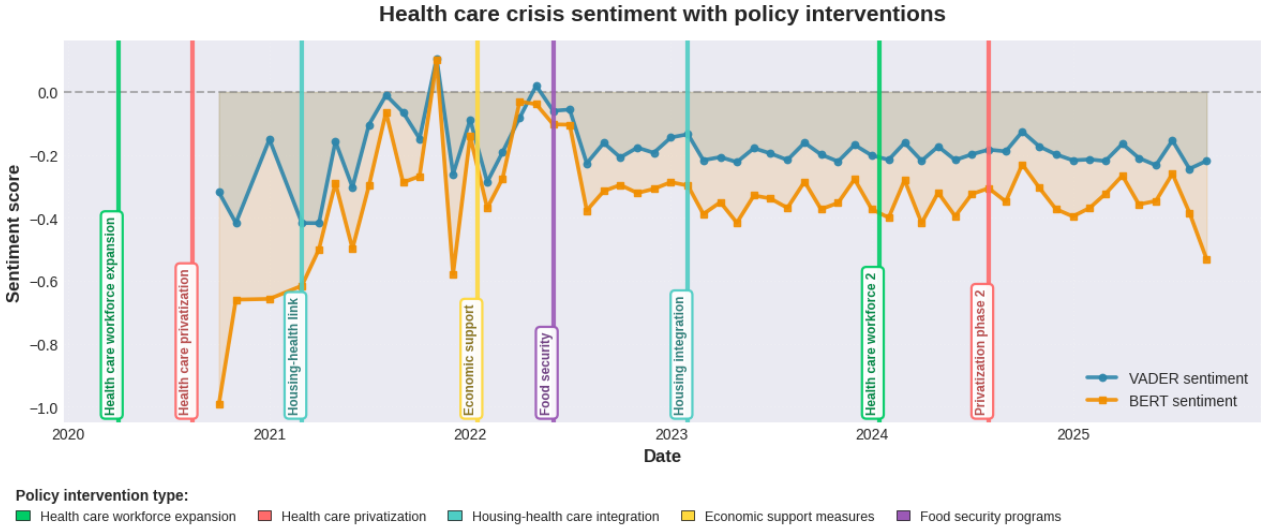
Both VADER and BERT scores remained negative in almost all months and fluctuated erratically, with no sustained trend, regardless of any intervention events. The BERT scores were lower than the VADER scores (by 0.15 on average), reflecting the negative nuances captured by the transformer model in the headlines. This further confirmed that the dates of policy interventions, as implemented and communicated to the public, do not appear to have made any significant impact on the dominant crisis narrative reflected in Canadian health news headlines.

XSL•FO
RenderX

Table 4. The impact of each type of policy intervention.

Policy type	Delta sentiment	Cohen *d*	Sustainability (days)	Implementation score
Economic support	–0.056	–0.128	28	0.20
Food security	–0.056	–0.122	10	0.40
Privatization	+0.028	+0.064	1	0.80
Workforce expansion	–0.004	–0.010	3	0.60
Housing integration	–0.017	–0.039	5	0.55

Figure 6. Time series data representing monthly BERT (orange) and VADER (blue) sentiment scores from October 2020 through September 2025 with an overlay of each of the 8 policy interventions assessed using a vertical color-coded line. The colors are used to designate the type of intervention. As can be observed by visual comparison of the 2 models in this figure, neither model demonstrated any significant shift resulting from any of the policy interventions. Given its sensitivity to contextual considerations, BERT was found to consistently produce lower (ie, more negative) values compared to VADER. Shaded areas extend from each series to zero. A legend that identifies each of the 5 types of policy intervention (health care workforce expansion, health care privatization, housing–health care integration, economic support measures, and food security programs) is provided at the bottom of the figure. BERT: bidirectional encoder representations from transformers; VADER: Valence Aware Dictionary and Sentiment Reasoner.



Predictive Feature Analysis

The primary factor that predicted the expression of a negative sentiment about health care issues was the text length, that is, how much text there was in the news article (0.521). The second largest set of predictors were "calendar month" (0.244) and "year" (0.124). Based on topic indicators related to health care, the top indicators for new articles with high levels of information included "economic determinants of health" (0.038); "social inequality health effects" (0.022); "housing as health determinant" (0.019); "food insecurity health impact" (0.017); and "health care system crisis" (0.015). To test the sensitivity of our results, we also calculated SHAP values [24] for the same Random Forest Model; text length remained the dominant predictor under both methods. However, SHAP ranked the Economic Determinants of Health topic indicator above the calendar-based features (month and year), which is contrary to those found using Gini-based ranking. This difference is expected given the previously cited propensity of Gini Impurity to preferentially rank high cardinality continuous variables over less frequent or rare binary variables [25]. The full comparison is provided in Multimedia Appendix 2.

Integrated Surveillance Dashboard

The results of all the analyses reported in the preceding Results subsections (Sentiment Surveillance through Predictive Feature Analysis) were synthesized into a prototype health care crisis monitoring dashboard (Figure 3). This dashboard demonstrates the operational capability of this study for performing near-real-time health system infoveillance. This prototype dashboard has 4 components.

At the top left is a donut chart illustrating the relative proportion of media content related to crisis themes in the last 3 months. This allows for a quick determination of what thematic area is currently the most prominent in the media environment.

At the top right is a multiseries line chart showing the time trend of temporal sentiments for each theme separately over the last 3 months. This will allow an analyst to determine if there are any themes that are diverging from the overall trend.

At the bottom left is a gauge showing the rolling mean sentiment score (–0.197, SD 0.410; n=142 headlines) of coverage of health care issues over the last 3 months. The gauge provides a single-number summary of the current state of health system infoveillance, as well as an immediate visual signal that the system has been negatively oriented over the past 3 months.

XSL•FO
RenderX

At the bottom right is a grouped bar chart showing topic volume by month over the past 6 months. This will enable an analyst to determine whether any themes are increasing their volumes faster than other themes in near real-time.

Together, these charts demonstrate how the proposed infoveillance can be turned into a decision-support tool for health policymakers. It does so by integrating all dimensions of analytical results—sentiment analysis, topic prevalence, and temporal dynamics into 1 display: infoveillance. This would enable policymakers to monitor ongoing developments without repeatedly conducting manual analysis. The current reading of –0.197 (SD 0.410) for sentiment, indicating that coverage of Canadian health care is well into negative territory, is consistent with findings presented in the Sentiment Surveillance subsection of the Results section.

Discussion

Principal Findings

The integration of the described multidisciplinary NLP pipeline, including techniques to perform sentiment analysis, topic modeling, causal network analysis, and forecasting, to monitor health system failures in multiple health themes was demonstrated in this study to be feasible.

A rolling sentiment indicator for the last 3 months; a trendline for each health theme; and topic volume indicators were displayed on the same dashboard for the purpose of providing a decision support system for stakeholders. Furthermore, the dashboard will continue to be dynamic as new headline data continues to automatically feed into it. As such, this study extends the infodemiology framework by Eysenbach [10] and the digital disease detection paradigm by Brownstein et al [13], originally developed for detecting outbreaks of infectious diseases, to detecting structural failures in health care systems. A polycrisis represents a great opportunity to test the proposed surveillance tool to identify cross-theme dynamics and to assess policy impacts across different health themes.

Additionally, the dual model sentiment approach achieved high levels of accuracy (accuracy=0.97), which are consistent with other research findings that indicate that short crisis-theme–specific texts can provide a relatively clear lexical valence for both rule-based and transformer models [26]. It is also possible that the high degree of agreement between the 2 models was influenced to some extent by label inheritance because the training data used for BERT were labeled using VADER-derived labels [27]. Therefore, future surveillance systems would benefit from using human-annotated datasets to address the circularity issue. When there are resource constraints, the complementary strengths of the VADER (computationally efficient) and BERT (contextually sensitive) models may be used together [11]. This circularity is an acknowledged weakness in pipelines using lexical methods to develop training labels for subsequent transformation of a model, so the good correspondence between VADER and BERT can reasonably be interpreted as demonstrating internal consistency rather than independently validating the accuracy of either model against true values [28,29].

Empirical Evidence of Polycrisis Dynamics

To date, there have been no studies that examine polycrisis systems in Canada using the Granger causality technique within a broad statistical framework to understand how and why various crises become linked. The results from the Granger causality analysis provide the first empirical, quantifiable description of polycrisis topology for discussions related to health care in Canada's public discourses. One major finding relates to the strength of the relationship between food insecurity health impact discourse influencing coverage of housing as a health determinant discourse (P_{FDR}=.002). As such, an increase in Canadian news coverage of food insecurity generally leads to an increase in coverage of housing as a health determinant in the subsequent month. This causal connection directly aligns with the polycrisis mechanism of causal entanglement by Lawrence et al [3] and supports the Cascade Institute's assertion that cascading from "upstream" to "downstream" represents the distinguishing feature of a polycrisis [5].

In addition to this relationship, there also exists a relationship between coverage of economic determinants of health and food insecurity health impact discourse ($F_{1,54}$=7.601, P_{FDR}=.024), which suggests that an upward cascade occurs in the media coverage: coverage of economic hardship occurs before and predicts coverage of food insecurity, which then causes an increase in coverage of housing and social inequality issues. This does not indicate a causal chain among the underlying social conditions themselves—instead, intuitively monitoring economic discourse could provide some degree of early warning regarding whether downstream food-insecurity, housing, and social-inequality discourse will occur.

Health care system crisis discourse constitutes a separate causal branch and independently Granger-causes coverage of both housing as a health determinant and social inequality as a health issue (Table 3). Thus, rather than occurring along a single linear chain, Canada's health care polycrisis discourse appears to propagate via 2 independent, parallel upstream pathways.

While food insecurity health impact was not the largest singular theme (Table 1), it functioned as the central node of the lag-1 causal network (Table 3): food insecurity health impact received a Granger signal from economic determinants of health ($F_{1,54}$=7.601) and, in turn, Granger-caused coverage of both housing as a health determinant and social inequality as a health issue—the 2 strongest directional connections noted in this study. This affirms Dietitians Canada's perspective on food insecurity as a cross-cutting determinant of health [30]; additionally, it indicates that food insecurity acts as a discourse pathway through which economic hardship is transformed into coverage of housing and social-inequality health concerns.

Health care system crisis discourse occupied an upstream position in the lag-1 causal network: it independently Granger-caused coverage of housing and social inequality, and it did not receive any significant input signal from any other theme (Table 3). At longer lags (Figure 5), it received one significant input, from economic determinants of health at lag 6. Overall, collectively, these findings demonstrate an empirical implementation of polycrisis theory in relation to nation-states

XSL•FO
RenderX

in a way that has not occurred previously either theoretically or worldwide [31].

Policy Surveillance: Null Signal With Implications

The fact that we did not observe a statistically significant effect from any of our policy interventions on the media sentiment variable (combined with consistently low effect sizes for those that did exhibit some type of impact) indicates that these results could be valuable in terms of assessing the efficacy of health policies. In summary, while the absence of statistically significant effects indicates that none of our policy interventions produced a sufficient shift in the media's prevailing negative discourse to produce observable changes in media surveillance signals, our results are consistent with evidence suggesting that during periods where there are sequential, multiple crises occurring at the same time, the cumulative nature of discrete, policy-based announcements continues to fail to alter the aggregate context of the discourse space [32]. Further, it is worth noting that the only policy initiative—health care privatization—that has the largest implementation level (0.8) exhibits a relatively short time frame (1 day) for retaining any type of influence. This short time frame may be attributed to ideologically polarizing announcements generating positive and negative coverage simultaneously, thereby cancelling each other out in total media sentiment. Additionally, the fact that text length and calendar year were stronger predictors of negative sentiment than any topic indicator suggests that the polycrisis itself represented the primary surveillance signal throughout the examined time frame and subject matter area. However, it should also be noted that text length is a stylistic or editorial aspect of headline writing and not a substantive measure of crisis severity. Therefore, the dominance of text length as a predictor of negative sentiment may reflect differences in headline lengths across outlets or over time instead of being an actual representation of a Surveillance signal.

Future Directions

The focus of this study on news headlines only limits the analysis to the professional media layer of public discourse about Canada's health care polycrisis and does not include the immediate, direct, and personal reactions of the public. Additional data layers through social media platforms such as Reddit or X could expand upon the methodology described in this study to survey public opinion directly during a comparable public health crisis [33]. Developing alternative methodologies to compare the use of the LDA topic model used throughout this study to other embedding-based models such as BERTopic and Top2Vec, which are capable of extracting local and coherent topics from short informal text, is another potential area of inquiry for subsequent studies.

Further possibilities exist for enhancing this analytical pipeline using large language models (LLMs). For example, a general-purpose LLM such as ChatGPT (OpenAI) may be trained to identify sentiment, determine themes, or provide summaries regarding the relevance of policy statements to specific stories with relatively minimal coding. A comparative study examining the feasibility of applying LLMs for identifying stance in public health discourse identified that all LLMs carry a significant amount of risk associated with misclassifying content relative to human-annotated expert standards. Only a few well-prompted models produced results comparable to those obtained via human annotation [34]. Therefore, if LLMs are used in place of the machine learning classifier used in the current study, they will most likely sacrifice precision and reproducibility (ie, prompt sensitivity, model version drift, and lack of a fixed audited lexicon similar to VADER's or an audited accuracy similar to the BERT model's) for increased ease-of-use and speed. These trade-offs should be evaluated empirically instead of being assumed. Since the topic and sentiment labels developed through LDA and VADER were used in the current study, substituting or supplementing them with an LLM-based classifier provides a clear direction for a future, focused-methodology-oriented study. We chose not to do so because it would alter the entire pipeline of analyses supporting the conclusions drawn from this study as opposed to improving upon it.

Limitations

Headlines in news articles are an example of how a particular news organization can construct its interpretation of what is important or "newsworthy." Therefore, while the data collected through our analysis may reflect a consensus of what was considered most newsworthy at a given time, it does not necessarily reflect directly stated opinions from members of the general population. Although we cannot reproduce all instances of news article collection (ie, we could not replicate every single news headline retrieved using the Google News API), the fact that we have access to archives of all previously retrieved headlines can allow us to analytically reproduce our results. We also recognize that by relying on keyword-based methods to determine which themes were represented in our dataset, we may have artificially increased the number of occurrences of certain categories of topics (eg, "health care system crisis"). As such, there may be some degree of bias introduced into our analyses due to overcounting of specific topic areas. BERT relies upon VADER-derived labels as part of its model development process. This reliance limits the extent to which independent models can be compared to one another. Given that we only analyzed 8 separate policy interventions in this study, there may be limitations associated with detecting even relatively small effects of these policies based solely on statistical analysis alone. Our corpus only includes English-language sources; therefore, if French-language coverage exists in Quebec or Acadia, it would not be included in our analysis. Future studies using this methodology could expand the types of languages represented in their corpora, extend the length of the policy window examined in each case study, and use human-annotated training labels instead of VADER-labeled training labels.

Although we were able to track Canada's polycrisis discourse signal across 259 different news outlets, there is evidence that the discourse signal has been concentrated among a few large national news outlets. Specifically, the top 5 news outlets accounted for 26.8% (1809/6757) of all headlines, and the top 10 news outlets accounted for 37.8% (2554/6757) of all headlines. Additionally, the news outlet with the most contributions to the discourse signal was the CBC, which contributed to 13.3% (897/6757) of the total. Editorial

XSL•FO
RenderX

positioning and regional focus of the larger news outlets may have had a greater influence on the discourse signal compared to a more even distribution of samples from smaller news outlets. However, we did not classify each news organization into categories based on political leanings, credibility as determined by third parties, or audience size. Descriptively, the news outlet pool included a variety of institutional positions, such as national broadcast networks (eg, CBC, CTV News, and Global News), financial-oriented publications (eg, Financial Post), policy institutes focused on markets (ie, Fraser Institute), and labor- or progressive-oriented organizations (eg, Canadian Union of Public Employees and Canadian Centre for Policy Alternatives). As such, we cannot exclude the possibility that varying amounts of reporting will occur from differing types of news outlets or orientations. Future studies should include classification and weighting of outlet-level characteristics.

Similarly, we did not compare the media-derived surveillance signal to an independent measure of system stress. A valid method of assessing whether the media-derived surveillance signal reflects true levels of system stress would be through comparison to real-time measures of health care usage metrics (such as Emergency Department Wait Times, Hospital Occupancy Rates, etc). Therefore, creating a link between the media-derived surveillance signal and an objective measure of system stress will be an important next step before using the model as an operational tool.

Lastly, because our media-derived surveillance signal represents only the frames chosen by media organizations, we capture what they choose to report about but do not capture what Canadians think about the polycrisis. Therefore, integrating this media-based signal with either publicly available social media or survey data representing Canadian public opinion would provide insight into how closely media framing aligns with the perceptions of Canadians regarding the polycrisis.

Conclusions

This study introduces a multimethod NLP-based surveillance pipeline which represents the first computational characterization of Canada's health care polycrisis as a causally entangled discourse system. Through the analysis of 6757 news headlines in conjunction with 5 interdependent health crisis themes, the pipeline found that negative sentiment has been consistently represented in media coverage throughout this study sample period. The study also demonstrated significant Granger-causality pathways among the health crisis themes consistent with polycrisis dynamics and indicated that none of the current policy interventions have produced a detectable footprint in the overall discourse signal.

These findings provide three principles to apply to health informatics in terms of designing surveillance systems for polycrisis: (1) health informatics pipelines designed to collect data from multiple health themes will be required to monitor cross-theme causal relationships missed by single-topic systems, (2) conducting Granger-causality tests among discourse themes will allow for identifying potential leading-indicators within each health theme to support development of early-warning alerting capabilities, and (3) assessing the impact of policies through health informatics surveillance systems should use FDR-corrected statistical inference and report effect sizes to preclude misinterpretation of policy effects as effective.

Although the 12-month forecasts are uncertain (all 95% CIs include zero) and show a rise only for health care system crisis; the cross-theme dynamics observed over 5 years of coverage indicate that health informatics infrastructure capable of analyzing polycrisis dynamics at a large scale is now a functional necessity rather than merely a methodological aspiration.

Acknowledgments

The authors would like to express their gratitude to the reviewers for their helpful comments.

Funding

No funding was received by the researchers for the conduct of this study. The article processing charge was covered by a Discovery Grant (RGPIN-2026-06675) from the Natural Sciences and Engineering Research Council of Canada.

Data Availability

The raw data set containing the 6757 news headline titles analyzed within this study was obtained through the use of the GoogleNews Python library (v1.6.14), using the Google News API. We report in detail in the methods section all of our search terms, date ranges, and exclusion criteria, which can be used to replicate how we collected the dataset. A detailed description of all 12 of the search query terms we used to create the overall corpus of news headlines is available in the Methods (Data Collection) section, so that readers may independently reproduce our dataset. The same CSV file we have used to examine the data in this study, which includes the crisis type, headline text, publication date, and source outlet information for all 6757 headlines, has been made available as Multimedia Appendix 3 to accompany this submission. The Python analysis scripts we used to analyze the data (VADER [Valence Aware Dictionary and Sentiment Reasoner], BERT [Bidirectional Encoder Representations from Transformers] fine-tuning, latent Dirichlet allocation, ARIMA [autoregressive integrated moving average], ETS [error, trend, and seasonality], and Granger Causality), along with the table outputs produced by each script, have been included in Multimedia Appendix 4 accompanying this submission. Therefore, all numbers reported in this paper can be recalculated directly.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Policy interventions assessed for impact on media sentiment. All 8 policy interventions considered in this study, including the 3 that predated sufficient headline volume for statistical testing (marked with an asterisk) and were therefore excluded from the results reported in Table 4.
[DOCX File , 16 KB - infodemiology_v6i1e99655_app1.docx]

Multimedia Appendix 2
Random forest feature importance: Gini Impurity vs SHAP (Shapley additive explanations). Comparison of feature importance rankings for the random forest classifier (negative-sentiment prediction) using the default Gini impurity measure versus mean absolute SHAP values.
[DOCX File , 15 KB - infodemiology_v6i1e99655_app2.docx]

Multimedia Appendix 3
Raw dataset of news headlines. The complete dataset of 6757 news headlines analyzed in this study, with crisis type, headline text, publication date, and source outlet recorded for each headline.
[XLSX File (Microsoft Excel File), 206 KB - infodemiology_v6i1e99655_app3.xlsx]

Multimedia Appendix 4
Analysis code and reproducibility package. The complete Python analysis code (data collection, preprocessing, topic modelling, sentiment analysis, time-series forecasting, and Granger causality testing) used to produce every table and figure in this manuscript, together with the topic-label correction script applied during manuscript revision and the raw output files each stage produces.
[DOCX File , 53 KB - infodemiology_v6i1e99655_app4.docx]

References

1. Solar O, Irwin A. A conceptual framework for action on the social determinants of health. World Health Organization. 2010 Jul 13. URL: https://www.who.int/publications/i/item/9789241500852 [accessed 2026-10-07]
2. Marmot M. Social determinants of health inequalities. Lancet 2005;365(9464):1099-1104 [FREE Full text] [doi: 10.1016/S0140-6736(05)71146-6] [Medline: 15781105]
3. Lawrence M, Homer-Dixon T, Janzwood S, Rockström J, Renn O, Donges JF. Global polycrisis: the causal mechanisms of crisis entanglement. Global Sustainability 2024 Jan 17;7:e6. [doi: 10.1017/sus.2024.1]
4. Tooze A. Welcome to the world of the polycrisis. Financial Times. 2022. URL: https://www.ft.com/content/498398e7-11b1-494b-9cd3-6d669dc3de33 [accessed 2026-10-06]
5. Lawrence M, Shipman M, Janzwood S, Arnscheidt C. Polycrisis research and action roadmap. Cascade Institute. 2024 Aug 16. URL: https://cascadeinstitute.org/technical-paper/polycrisisroadmap/ [accessed 2026-10-07]
6. Addressing Canada's health workforce crisis. Report of the Standing Committee on Health. 2023 Mar. URL: https://www.ourcommons.ca/DocumentViewer/en/44-1/HESA/report-10/ [accessed 2026-10-07]
7. New data on household food insecurity in 2023. PROOF. 2024 Apr 26. URL: https://proof.utoronto.ca/2024/new-data-on-household-food-insecurity-in-2023/ [accessed 2026-10-07]
8. Changes in households' vulnerability to food insecurity in Canada before and after the COVID-19 pandemic. Statistics Canada. 2025 Dec 17. URL: https://www150.statcan.gc.ca/n1/pub/82-003-x/2025012/article/00001-eng.htm [accessed 2026-10-07]
9. Sergeant M, de Souza R, Razak F. The end-of-life care, environment and health system ‘polycrisis’ needs a new approach. Policy Options. 2024 Aug 12. URL: https://policyoptions.irpp.org/2024/08/canadian-polycrisis [accessed 2026-10-07]
10. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
11. Cho HN, Jun TJ, Kim YH, Kang H, Ahn I, Gwon H, et al. Task-specific transformer-based language models in health care: scoping review. JMIR Med Inform 2024;12:e49724 [FREE Full text] [doi: 10.2196/49724] [Medline: 39556827]
12. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
13. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. N Engl J Med 2009;360(21):2153-2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
14. Budak C, Goel S, Rao JM. Fair and balanced? Quantifying media bias through crowdsourced content analysis. Public Opin Q 2016;80(S1):250-271. [doi: 10.1093/poq/nfw007]

XSL•FO
RenderX

15. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proc Int AAAI Conf Web Soc Media 2014;8(1):216-225. [doi: 10.1609/icwsm.v8i1.14550]
16. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); June 2 to June 7, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/n19-1423]
17. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022. [doi: 10.7551/mitpress/1120.003.0082]
18. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. 2015 Presented at: WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining; February 2-6, 2015; Shanghai China p. 399-408. [doi: 10.1145/2684822.2685324]
19. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv Preprint posted online on March 11, 2022. [doi: 10.48550/arXiv.2203.05794]
20. Angelov D. Top2Vec: distributed representations of topics. arXiv Preprint posted online on August 19, 2020. [doi: 10.48550/arXiv.2008.09470]
21. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 1969;37(3):424-438. [doi: 10.2307/1912791]
22. Shojaie A, Fox EB. Granger causality: a review and recent advances. Annu Rev Stat Appl 2022;9(1):289-319 [FREE Full text] [doi: 10.1146/annurev-statistics-040120-010930] [Medline: 37840549]
23. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
24. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 2017 Dec 4 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA p. 4768-4777 URL: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
25. Strobl C, Boulesteix A, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 2007;8(1):25 [FREE Full text] [doi: 10.1186/1471-2105-8-25] [Medline: 17254353]
26. Aggarwal I, Joseph S, Jaganathan N, Patel A, Kumar V, Devarapalli M. Sentiment analysis in healthcare: a comparison of VADER, BERT, and Flair NLP models on patient reviews of pain management physicians. Cureus 2025;17(7):e88902. [doi: 10.7759/cureus.88902] [Medline: 40881543]
27. Villanueva-Miranda I, Xie Y, Xiao G. Sentiment analysis in public health: a systematic review of the current state, challenges, and future directions. Front Public Health 2025;13:1609749 [FREE Full text] [doi: 10.3389/fpubh.2025.1609749] [Medline: 40620557]
28. Wagholikar KB, Estiri H, Murphy M, Murphy S. Polar labeling: silver standard algorithm for training disease classifiers. Bioinformatics 2020;36(10):3200-3206 [FREE Full text] [doi: 10.1093/bioinformatics/btaa088] [Medline: 32049335]
29. Ratner A, Bach SH, Ehrenberg H, Fries J, Wu S, Ré C. Snorkel: rapid training data creation with weak supervision. VLDB J 2020;29(2):709-730 [FREE Full text] [doi: 10.1007/s00778-019-00552-1] [Medline: 32214778]
30. Position statement on household food insecurity in Canada. Dietitians of Canada. 2024 Mar. URL: https://www.dietitians.ca/DietitiansOfCanada/media/Images/DC-Household-Food-Insecurity-Position-Statement_2024_ENG.pdf [accessed 2026-10-07]
31. Henig D, Knight DM. Polycrisis: prompts for an emerging worldview. Anthropol Today 2023;39(2):3-6. [doi: 10.1111/1467-8322.12793]
32. Zhu H. Interplay between discrete emotions and preventive behavior in health crises: big data analysis of COVID-19. Int J Environ Res Public Health 2022;19(24):16407 [FREE Full text] [doi: 10.3390/ijerph192416407] [Medline: 36554288]
33. Pang O, Movahedi Nia Z, Gillies M, Leung D, Bragazzi N, Gizo I, et al. Analyzing Reddit social media content in the United States related to H5N1: sentiment and topic modeling study. J Med Internet Res 2025;27:e70746 [FREE Full text] [doi: 10.2196/70746] [Medline: 40925599]
34. Espinosa L, Salathé M. Use of large language models as a scalable approach to understanding public health discourse. PLOS Digit Health 2024;3(10):e0000631 [FREE Full text] [doi: 10.1371/journal.pdig.0000631] [Medline: 39401186]

Abbreviations

ARIMA: autoregressive integrated moving average
BERT: Bidirectional Encoder Representations from Transformers
CBC: Canadian Broadcasting Corporation
ETS: error, trend, and seasonality
FDR: false discovery rate
LDA: latent Dirichlet allocation
LLM: large language model
NLP: natural language processing
NLTK: Natural Language Toolkit

XSL•FO
RenderX

SHAP: Shapley additive explanations
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by M Haupt; submitted 28.Apr.2026; peer-reviewed by A Singh, K Morris; comments to author 13.Aug.2026; revised version received 30.Sep.2026; accepted 01.Oct.2026; published 29.Oct.2026.

Please cite as:
Muteba Mwamba JW, Mago V
Monitoring Canada's Health Care Polycrisis Through News Headline Analysis: Infodemiology Study
JMIR Infodemiology 2026;6:e99655
URL: https://infodemiology.jmir.org/2026/1/e99655
doi:10.2196/99655
PMID:

©John Weirstrass Muteba Mwamba, Vijay Mago. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 29.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Parents' Multiperspective Views and Misinformation About COVID-19 Mitigation Measures During Tennessee School Board Meetings: Qualitative Content Analysis Using YouTube

Olufunto A Olusanya[1], MPH, MD, PhD; Brianna M White[1], MPH; Brenda Amuchi[2], MPH, MBBS; Chad Melton[1], PhD; Arash Shaban-Nejad[1], MSc, MPH, PhD

[1]UTHSC-ORNL Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, University of Tennessee Health Science Center, 50 North Dunlap Street, 492R, Memphis, TN, United States

[2]Boston University School of Public Health, Boston, MA, United States

Corresponding Author:
Arash Shaban-Nejad, MSc, MPH, PhD
UTHSC-ORNL Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, University of Tennessee Health Science Center, 50 North Dunlap Street, 492R, Memphis, TN, United States

Abstract

Background: In fall 2021, Tennessee school districts faced heightened debates over COVID-19 mitigation amid rising cases, limited vaccination availability, and widespread misinformation. School board meetings (SBMs) served as pivotal decision-making forums influencing district policies. This study investigated perceptions and misinformation regarding the COVID-19 mask mandate at SBMs held within 6 of Tennessee's largest school districts. With widespread debate over pandemic measures, including mask use in schools, understanding community sentiments is crucial for guiding public health policies.

Objective: This study aimed to investigate the viewpoints of parents or caregivers and teachers regarding COVID-19 safety protocols, particularly the mask mandate, and to identify the misinformation circulating within SBMs.

Methods: Participants' commentaries were extracted from 6 SBM recordings that were publicly uploaded to YouTube from August through September 2021. The data were examined qualitatively to capture themes related to concerns, support, and misinformation. Inductive thematic analysis was conducted using transcripts generated via Microsoft Azure speech-to-text and manually verified.

Results: Many parents or caregivers gave personal accounts of how the pandemic had impacted them, their children, and their communities, describing significant comorbidities, adverse psychosocial impacts, mental health disorders, learning difficulties, and worsening socioeconomic and educational disparities. Six thematic domains emerged: (1) perceived effects of the COVID-19 pandemic on children, teachers, and parents or caregivers, including psychosocial distress, learning disruptions, and burnout; (2) perceived effects of mask mandates on children, particularly concerns regarding physical health and psychosocial well-being; (3) perceived government overreach and legal objections to COVID-19 mitigation mandates; (4) tensions between personal liberty, religious beliefs, and collective responsibility in masking decisions; (5) circulation of misinformation and conflicting guidance regarding mask safety and effectiveness; and (6) institutional strain, social tensions, and hostility directed toward school officials alongside educator burnout.

Conclusions: Perspectives on COVID-19 mitigation varied widely across meeting participants, highlighting the need for health officials and policymakers to engage in proactive health promotion strategies. Strengthening public health communication, misinformation mitigation, and institutional support for teachers will be essential to ensuring safe and effective learning environments during future public health crises.

(JMIR Infodemiology 2026;6:e75691) doi:10.2196/75691

KEYWORDS

misinformation; mask mandate; vaccines; schools; in-person learning; COVID-19 pandemic

Introduction

The COVID-19 pandemic, caused by SARS-CoV-2, has significantly impacted public health, education, and society. Emerging in late 2019 in Wuhan, China [1], it continues to negatively affect health care delivery, community well-being, and social equity. By 2023, the United States had recorded more than 104 million COVID-19 cases and 1.1 million deaths [2], underscoring a major public health crisis that strained hospitals, health care workers, and public health infrastructure [3-6]. The pandemic led to school closures and a transition to online and

XSL•FO
RenderX

hybrid learning, exacerbating disparities in digital access, educational settings, and academic support [7]. Students from racial and ethnic minorities, low-income households, and with limited internet faced barriers to participation, academic progress, and mental health. Overall, the pandemic worsened long-standing racial and ethnic disparities across the continuum of health care, from testing and vaccination to hospitalizations and mortality, mirroring educational inequities [8,9].

During the COVID-19 pandemic, widespread school closures were implemented across the United States to slow transmission as infection rates increased. As schools reopened through phased or hybrid models, school districts encountered barriers to implementing evidence-based mitigation measures, such as mask use, social distancing, and hand hygiene, due to significant opposition from some parents or caregivers [10-12], despite strong public health recommendations. Before the Pfizer-BioNTech COVID-19 vaccine was authorized for children ages 5 - 11 years in October 2021, experts emphasized the importance of maintaining preventive measures [13,14]. By the 2021 - 2022 academic year, 18 states and the District of Columbia had mask mandates. However, by May 2022, all states except Hawaii had lifted these requirements, despite evidence that such measures reduced infection and mortality rates [15].

Tennessee, the 15th most populous US state in the southeastern United States, faced COVID-19 unique challenges [16]. In fall 2021, school board meetings (SBMs) were held statewide to disseminate safety information and support in-person learning. These meetings also served as public forums where parents or caregivers could express support or opposition to the mask mandates [17]. Many SBM recordings were uploaded to YouTube, a social networking platform widely known for accessible health education [18,19]. Unlike typical YouTube content, these recordings document civic proceedings without being algorithmically promoted. However, the lack of content oversight, combined with the widespread dissemination of misleading information and conspiracy theories, posed a significant public health challenge [20]. This misinformation, amplified by politicization and antiscience narratives, contributed to mistrust of public health institutions and hindered acceptance of COVID-19 mitigation measures [21]. Unlike popular YouTube health content, SBM recordings tend to reach smaller, local audiences and may not engage the broader public.

Despite rising COVID-19 cases, Tennessee lifted its mask mandate in April 2021 [22], and by July 2022, only 56.54% of residents were fully vaccinated [23]. While COVID-19 mandates have ended, misinformation, mistrust, and polarized health communication persist in debates about other pathogens. Thus, the implications extend beyond COVID-19 specifically. This study was conducted in fall 2021 amid active debates about COVID-19 mitigation policies and limited pediatric vaccination. Our study identifies overarching best-practice themes and offers actionable recommendations to improve preparedness and response in K-12 educational settings during future public health crises. We anticipate that these findings will inform health promotion, policymaking, and program implementation.

While these trends demonstrate resistance to public health recommendations, few studies have qualitatively examined the social and contextual factors influencing public beliefs, behaviors, and hesitance toward mitigation measures and vaccination. This research fills that gap by exploring parents' and caregivers' perceptions of the mask mandate and other public health interventions through an analysis of Tennessee's SBM recordings uploaded to YouTube.

Methods

Overview

YouTube provides publicly accessible web-based data used in this study [24]. SBMs, the unit of analysis for the study, were held in person across 6 of Tennessee's largest counties and were subsequently uploaded to YouTube. In the fall of 2021, we examined YouTube recordings of SBMs conducted between August and September 2021 (ie, before school districts reopened for the 2021 - 2022 school year) to gain insight into parent or caregiver perceptions of the universal mask mandate. Fall 2021 was selected as the analytical time point because it coincided with the start of the academic year after statewide COVID-19 mandates were lifted in Tennessee. During this time, school districts were actively deliberating mitigation strategies for the upcoming school year, making it a critical window for examining public testimony and policy debate. While additional time points could provide longitudinal insight, this study was intentionally designed to capture discourse during and immediately after this specific policy transition period. A qualitative analysis was conducted on the commentaries and testimonials delivered during these in-person meetings.

Study Area and Population

We used a purposive sampling strategy to select 6 large counties based on population sizes using data from the 2021 American Community Survey. Recordings for the SBM were selected based on the following eligibility criteria: (1) SBMs that were held within 6 of Tennessee's largest counties, sorted from largest to smallest: Shelby, Davidson, Knox, Hamilton, Williamson, and Montgomery, and (2) SBMs that were officially uploaded on the YouTube home pages of school districts. We acknowledge that speakers at SBMs may only represent a self-selected subset of community members; however, because SBMs serve as formal policymaking forums where decisions are directly shaped by public testimony, the views expressed remain central to understanding the local public health decision context.

Data Collection and Management

The framework that guided our data collection and storage was based on the following stages: (1) extraction and download of video recordings, (2) transcription of video recordings, (3) dual independent review of video recordings, and (4) data storage within a password-protected, Health Insurance Portability and Accountability Act (HIPAA)–compliant drive [25]. Each SBM recording was downloaded directly from each county's school district YouTube home page to reflect unedited, accurate content. The 6 downloaded video recordings were converted to audio files for transcription. Using Microsoft Azure's speech-to-text software, audio files were converted to verbatim written transcripts, which were then subsequently deidentified

XSL•FO
RenderX

to protect the confidentiality and privacy of SBM attendees. Transcripts were manually reviewed and corrected by members of the research team to ensure accuracy and data quality. Methodological rigor and trustworthiness were supported for this study through the following strategies: (1) credibility, confirmability, and accuracy were enhanced by having 2 independent research teams review all transcripts and conduct data analysis; (2) data saturation was determined once all relevant themes had been identified and no new information emerged [26]; (3) the use of SBM speakers' verbatim quotes throughout the analytic process to ensure accurate representation of their perspectives; and (4) comparing accounts across a diverse range of speakers strengthened the dependability and completeness of the findings by ensuring that multiple viewpoints were captured [27].

Ethical Considerations

Growing evidence highlights ethical concerns associated with conducting research using data from social media websites [28]. Accordingly, several ethical considerations were observed in this study to minimize concerns and ensure speakers' privacy and confidentiality. YouTube video and audio files were uploaded and stored in a password-protected, HIPAA-compliant data storage drive accessible only to research team members. Participants' personal information (ie, names and demographic characteristics) was deidentified to safeguard confidentiality. Only SBMs publicly uploaded to YouTube and readily accessible to any internet user were used in our research. In addition, speakers' quotes were reported verbatim in our Results section. Quotations were selected to reduce identifiability by removing personal details and focusing on thematic content. The University of Tennessee Health Science Center Institutional Review Board reviewed the study and determined that it met criteria for exemption from full review (23 - 09294-NHSR) [29]. Because the study analyzed publicly available online content and did not involve direct interaction with human participants, informed consent was not required. Participants did not receive compensation for their contributions, as the data analyzed were publicly available recordings of school board meetings.

Analytical Approach

Qualitative Analysis

Using the traditional qualitative method, our study was guided by thematic and inductive analytic approaches [30]. Transcript files were exported into the qualitative software program ATLAS.ti (version 9.0; Scientific Software Development GmbH) for thematic analysis [31]. ATLAS.ti's artificial intelligence–assisted functions were not used; all coding was conducted manually by the research team. Before analysis, transcribed data were repeatedly read and reviewed in their entirety to ensure an in-depth understanding of the data. Our team was divided into 2 groups, which independently conducted detailed coding of the transcripts. The coding process entailed systematically searching, identifying, collating, categorizing, reviewing, and reporting emerging patterns (themes) embedded throughout the transcript data [30]. Thereafter, coding discrepancies were resolved through consensus within the research team. Codes were subsequently condensed and recategorized to represent broad themes. A detailed codebook and code definitions were developed to elicit essential details from the transcripts. Codes were reassessed, compared, and regrouped into final thematic classifications to ensure consistency. This analytic approach has been previously adopted in another study that qualitatively assessed an online data source [32].

Identification of Misinformation

For this study, we defined misinformation as statements that are presented as facts but contradict public health guidance in effect during August-September 2021, evidence-based public health guidance from the Centers for Disease Control and Prevention, American Academy of Pediatrics, and World Health Organization. During coding, 2 researchers (OO and BW) independently flagged such statements and resolved disagreements through discussion, referring to these sources as needed. Value-based opinions were not coded as misinformation unless they contained empirically false factual claims.

Results

Overview

Speakers at the SBMs included parents or caregivers, health care professionals, and local government officials. The number of speakers at the SBM includes Davidson (n=32), Knox (n=28), Williamson (n=27), Montgomery (n=25), Hamilton (n=17), and Shelby (n=16). Many parents or caregivers gave personal accounts of how the pandemic had impacted them, their children, and their communities. Six major domains emerged from thematic analysis, summarized in Table 1 with representative participant quotations.

XSL•FO
RenderX

Table . Summary of key findings and representative quotes from Tennessee school board meetings (August-September 2021).

Domain	Key findings	Representative quotes
Perceived effects of the COVID-19 pandemic on children, teachers, and parents or caregivers	• Participants described wide-ranging psychosocial and educational disruptions. • Parents reported mental health issues (anxiety and depression), and teachers voiced concerns about safety, staffing shortages, and burnout.	• "It's causing an increased sense of uncertainty, apathy, and anxiety." • "Many of them were very anxious and depressed…one of our students confessed he was thinking about taking his life." • "Will parents be forced to leave work to watch their kids because their schools don't have sufficient staffing?"
Perceived effects of the mask mandate on children, teachers, and parents or caregivers	• Parents frequently opposed mask mandates, citing physical discomfort, respiratory problems, and learning loss. • Emotional effects included frustration and fear of stigmatization among children.	• "My son has a very difficult time breathing while wearing a mask." • "Do you know what sort of psychological damage we do to children by forcing them to cover their faces?" • "It is not healthy for a child to think that they are a disease vector."
Perceived government overreach and legal objections to COVID-19 mandates	• Participants debated the legality and ethics of mask mandates. • Some argued school boards lacked authority; others viewed mandates as ethical acts for collective safety.	• "This board does not have the legal right to issue a mask mandate." • "Making choices for the collective good is exactly what I want to model." • "You have the responsibility to take care of the school and the current health environment."
Perceptions of liberty and religious freedom as core to American identity	• Many framed opposition around parental rights and religious beliefs, equating mandates with loss of freedom. • Others stressed balancing personal liberty with social responsibility and community health.	• "You are not God ... To impose your will over free individuals is un-American." • "Our beliefs guarantee my freedom, and my children's, to breathe oxygen." • "To be a part of any organized, healthy community requires some abandonment of personal freedoms."
Misinformation and conflicting guidance on mask use	• Misinformation and pseudoscientific claims circulated widely. • Some cited unverified medical opinions asserting masks were harmful or ineffective, illustrating the influence of misinformation during the pandemic.	• "There is no evidence that masks decrease the spread of COVID-19." • "Masks cause issues for children ... microbes are inhaled deeply into the lungs." • "People engaging in cosmetic theater thinking they are making a difference."
Institutional strain, conflict, and social tensions in school-based COVID-19 mitigation	• Meetings revealed polarization and hostility toward school officials. • Personal attacks and threats were common. • Educators described exhaustion, lack of support, and inconsistent policy guidance.	• "You have decided that our kids should go to school every day wearing muzzles like rabid dogs." • "If you vote for this, we will vote you out." • "Deaths of teachers—no change."

Major Themes and Findings

Perceived Effects of the COVID-19 Pandemic on Children, Teachers, and Parents or Caregivers

Concerns of the COVID-19 Pandemic on Health and Health Care Systems

The COVID-19 pandemic caused significant morbidity or mortality and placed an enormous burden on health personnel. One health care professional shared firsthand accounts of the severity of COVID-19 within clinical settings and expressed concern for decreasing COVID-19 preventive measures:

> *... kids are getting sick ... Pediatric ICUs and ERs across the country, across the South are being stretched to capacity ...*

Another echoed this concern:

> *Our pediatric hospitals are already full of RSV and parainfluenza. There will not be staff to care for the children when they get sick.*

One teacher shared:

> *Once I got sick ... I was in the hospital twice. I've been diagnosed with ... multiple life-threatening illnesses as a consequence of COVID, including several heart conditions ...*

Some conveyed their frustrations at others' disregard for COVID-19 safety measures:

> *It is like you were throwing a large Chicken Pox party with our children—only the stakes are incredibly higher. How can you not prioritize the life of a child?*

XSL•FO
RenderX

Concerns About the COVID-19 Pandemic on Psychosocial Capabilities, Staff Shortages, and Student Learning

Some SBM speakers expressed that significant psychosocial impact and adverse health effects had created devastating effects. While some expressed their support for COVID-19 mitigation measures, many conveyed feelings of pandemic fatigue and burnout. Some parents or caregivers were concerned about school staffing shortages:

> *Will we need to shut down the schools, and will parents be forced to leave work to watch their kids because their schools don't have sufficient staffing ...?*

Referring to psychosocial impacts and increased frequency of mental health disorders among students, SBM speakers shared:

> *... the threat of reinstating restrictions is very upsetting to students. It's causing an increased sense of uncertainty, apathy, and anxiety ...*

and

> *... many of them were very anxious and depressed ... It has affected their mental health so poorly one of our students ... confessed that he was thinking about taking his life ...*

Significant disruptions to learning and in-school activities were discussed. One teacher articulated concerns about non-English–speaking students:

> *... for students with exceptionalities and English language learners, that is still a big deficit.*

Moreover, some parents or caregivers expressed the following criticism:

> *... contact tracing needs to stop. The kids that went home, they sat there and stared at the walls for two weeks and then went back to school ...*

and

> *... last year he was contact traced three times ... so, he missed eight weeks out of 18. The weeks that he was home, he did literally nothing ...*

Perceived Effects of the Mask Mandate on Children, Teachers, and Parents or Caregivers

Concerns About the Use of Masks on Children's Health

The sentiments on the mask mandate were negative, with several parents or caregivers opposed to their implementation in schools. One parent shared:

> *My son has a very difficult time breathing while wearing a mask. ... he has had croup six times, RSV twice, and pneumonia, which has compromised his lungs and made wearing a mask for long periods extremely difficult.*

Other parents or caregivers echoed this sentiment:

> *My four children went with a mask ... They hated it daily. We had headaches, stomachaches, tears about not wanting to go to school.*

Another shared:

> *... I have two vaccine-injured children ... still I would never put them in a mask because their brain needs oxygen to grow ...*

Concerns Regarding the Use of Masks on Children's Psychosocial Capabilities

While discussing the emotional and adverse impact of the masks, one parent shared:

> *Now, do any of you know what sort of psychological damage we do to children by forcing them to cover their faces? ... depriving them of the ability to see each other's faces ... Have you wondered about the health effects of forcing kids to breathe through sweat, spit, and dirt-soaked rags every single day?*

Moreover, some parents or caregivers expressed their fears about masks leading to learning loss:

> *But what we're not talking about is ... the learning loss and the behavioral issues that have popped up from wearing these masks.*

Similarly, one parent noted the undue social burden on children who were sometimes viewed as a public threat for not wearing masks:

> *... consider the profound mental health conditions that you are potentially instilling in your children. It is not healthy for a child to think that they are a disease vector, going around killing their grandparents and their neighbors.*

Perceived Government Overreach and Legal Objections to COVID-19 Mandates

Distrust of Laws Regarding the Implementation of Masks

During the SBM, some parents or caregivers referred to the state government's role, or perceived lack of it, in implementing COVID-19 safety measures, arguing that the proposed mask mandate was unlawful. One parent asserted:

> *Somebody brought up the governor. ... he's not an interpreter of the law. We have an interpreter of the law who told us that this board does not have the legal right to issue a mask mandate.*

Another added:

> *Masks and vaccines are emergency use authorizations. They are not to be mandated under federal law. The law is on our side, the science is on our side.*

One parent threatened to take legal action, "If this board exceeds their legal authority and votes to mask our kids, how many people here will sign on to a class action …?" However, others acknowledged that school districts did have the authority to impose mask requirements, noting, "The governor said the way laws are set up in Tennessee, school districts have the authority to institute mask requirements."

XSL•FO
RenderX

Perceptions of Collective Responsibility Versus Individual Choice in Masking Decisions

While some parents or caregivers opposed the implementation of school mask mandates, others viewed them as a collective effort and a selfless choice. One parent shared, "How do I explain to my first-grade daughter that she needs to wear a mask ... but other kids don't have to wear masks because their parents want them to have a normal experience, and personal choice is more important than the health of a community." A teacher emphasized the importance of modeling community-oriented behavior, stating, "... I try to facilitate with my kids ... making choices and taking actions for the collective good is exactly what I want to model." Several individuals urged school board members to prioritize student safety: "As a county school board, you have the responsibility to take care of the school and the current health environment." However, some parents or caregivers interpreted these expectations differently, arguing, "It's also a violation of ethics for this board to vote in a partisan manner and not represent their constituents ..."

Perceptions of Liberty and Religious Freedom as Core to American Identity

Perceived Parental Right to Decide the Child's Masking Behavior

Some parents or caregivers emphasized their right to determine what was best for their children, expressing concern that mask requirements threatened their personal freedoms. One parent stated, "You are not God ... To impose your will over the power of free individuals is not only morally reprehensible, it's un-American." Another parent added, "... the constitution and the Declaration of Independence, the Bill of Rights, and the Federalist Papers, and also the Bible ... these guarantee my freedom and yours, and our children to breathe oxygen." One parent explained that the reluctance to wear masks was motivated by their religious beliefs: "Here's the email that I got when I secured my son's religious exemption that he would not have to wear a mask anymore because it goes against our beliefs."

Nevertheless, several parents or caregivers supported universal masking: "To be a part of any organized, healthy community requires some abandonment of personal freedoms." Another parent criticized what was perceived as misplaced priorities on dress code policies, "Or should I tell her (daughter) that she has to wear certain clothes—but don't worry, you don't have to wear masks because you're not responsible if you give your friends a virus ..."

Prioritizing Students' Right to In-Person Learning or Education and Safety

Discussions also centered on barriers to implementing a typical back-to-school experience. In contrast to arguments focused on parental rights and personal freedoms, several parents or caregivers emphasized students' rights to safety and uninterrupted education. One parent argued: "This decision is not about parents' rights—this is about students' rights to safe learning environments ...." Another highlighted concerns about harassment and harmful rhetoric:

> *It is not okay for students to pull off the masks of other students who choose the right to wear masks. It is not okay for a teacher to make fun of students that choose to wear masks. It is not okay to compare wearing a mask to child enslavement or the Holocaust.*

For these parents or caregivers, returning to school with a mask mandate was seen as one of the most effective strategies to ensure a safe, stable, and consistent in-person learning environment.

Misinformation and Conflicting Guidance on Mask Use

The COVID-19 pandemic appeared to exacerbate the infodemic, with misinformation amplified through politicized debates surrounding mask mandates, vaccinations, and other mitigation measures. While the majority of health care providers advocated for mask use, a minority of health care providers offered contradictory guidance. One physician stated:

> *I'm a physician—I believe in statistics—I follow the science. There is no evidence that masks decrease the spread of COVID-19. Masks cause issues for children ...*

While a parent cited recommendations received from their health care provider, claiming that masking could worsen illness severity:

> *Dr. [Insert name] states that asymptomatic or mild cases of COVID become more severe when the infected is masked .... The oxygen-lowering effects of masks cause these microbes to be inhaled deeply into the lungs, where respiratory disease far worse than COVID can result.*

Some parents or caregivers expressed skepticism toward public health measures more broadly, with one noting, "I feel bad for all these people engaging in cosmetic theater thinking that they are making a difference against COVID." Others highlighted knowledge gaps and long-term safety concerns, stating, "... we don't know enough about it, and I'm not willing to risk the long-term health of my child breathing in from a mask that we are meant to exhale."

Institutional Strain, Conflict, and Social Tensions in School-Based COVID-19 Mitigation

Personal Attacks on School Officials

Some parents or caregivers expressed appreciation for the schools' efforts to implement COVID-19 mitigation measures. However, others directed sharp criticism at school board members, with one stating, "You and the school board have decided that our kids should go to school every day wearing muzzles like rabid dogs." In some cases, opposition escalated into explicit threats: "If you vote for this ... we will vote you out—if you own a business, we will boycott your business ..., you become a domestic enemy, and we will come for you politically and financially." These attacks were occasionally followed by loud clapping, whistling, and cheering from some attendees.

XSL•FO
RenderX

Teacher Burnout and Lack of Institutional Support

Teachers attending the SBM also expressed fatigue and burnout. Many highlighted the pressures of navigating virtual and in-person learning while facing limited support, including insufficient sick leave, minimal pay increases, and low staffing levels. Moreover, conflicting health guidelines and policies from school districts further contributed to a sense that teachers' health was not a priority. "Last year, when my spouse was showing signs of COVID, I was told to come in because I was vaccinated…" Another added, "Now the data shows different and deadly strains, yet no change .... Deaths of teachers—no change."

Privacy Concerns and Social Shaming

During the SBM, the COVID-19 vaccination had been approved for children aged 12 years and older. Accordingly, some parents or caregivers expressed concerns about disclosing students' COVID-19 vaccination status. A few parents or caregivers noted that teachers publicly asked students to share this information, stating:

> *... there have been middle school teachers who have asked the students ... to raise their hand if they've been vaccinated. I have heard administrators at assemblies this week who have praised and thanked children who are wearing masks.*

One parent criticized this practice as social shaming:

> *They are asking minor students to share protected health information in a public arena, which violates their privacy. Calling them out in front of others is social shaming.*

Echoing this concern, another parent remarked:

> *... we're making things even more divisive for our children ... this student is wearing a mask, and the student isn't ...*

Discussion

Principal Findings

Taken together, the 6 thematic domains highlight recurring patterns in SBM discourse, including contested authority over decisions related to children's health, the use of identity- and value-driven framing that facilitated the spread of misinformation, and increasing institutional strain as they navigated their role in public health decision-making. These patterns reflect broader challenges described in prior works on public trust, health communication, and the governance of public health interventions in educational and community settings [21,33-36].

Public schools are a central site where decisions about child health, education, and public welfare intersect. In fall 2021, SBMs across Tennessee became a key forum for considering COVID-19 mitigation strategies, including mask mandate, amid heightened concern, uncertainty, and debate. To examine how authority over child health decisions was negotiated in these settings, we examined comments and testimonials from SBM recordings uploaded to YouTube by school districts. Our analyses were conducted using qualitative thematic analysis of extracted data. Our findings revealed a significant divide among speakers. A subset of parents or caregivers strongly supported the mask mandate, emphasizing the importance of ensuring students' safety and maintaining the continuity of in-person learning during the pandemic, while others raised concerns grounded in personal values, perceived rights, and skepticism toward public health guidance. These findings underscore the complexity of balancing public health measures with community values and the challenges of addressing misinformation that may shape public opinion. These themes resonate with the principles of precision health [37], which position community perceptions, local sociopolitical environments, and real-world behavioral contexts as integral components of effective prevention.

However, some parents or guardians opposed local mask mandates in educational settings, citing violations of parental rights and infringement on personal liberties. Several SBM speakers described significant fears and concerns about the impacts of the COVID-19 pandemic, its mitigation measures, and other unfavorable sequelae (eg, disruptions to learning and mental health impacts). According to SBM commentaries, the COVID-19 pandemic and public health prevention measures had negatively affected students, parents or caregivers, and teachers. As in other reports, these SBM open forum sessions were occasionally characterized by highly politicized debates among some attendees who expressed negative sentiments, often expressed by other influential people [33] toward COVID-19–related policies and mitigation measures [17]. Scientific misinformation and conspiracy theories were circulated by some parents or caregivers and a few health care professionals who spoke outside their areas of expertise. During these meetings, debates emerged over whether vaccine and mask mandates were constitutional or infringed on personal liberty. In contrast, parents or caregivers who supported mask mandates emphasized students' rights to a safe learning environment and access to quality education. Religious objections primarily centered on perceived encroachments on parental authority rather than doctrinal prohibitions, which is consistent with available literature, indicating that religious resistance to masking reflects broader ideological and identity-based concerns rather than theological mandates [34,38-41].

The proliferation of misinformation posed a significant challenge to disease prevention and containment measures during the COVID-19 pandemic. The widespread availability of unverified user-generated content on social media platforms, combined with unrestricted discourse that amplified incomplete or false narratives, undermined vaccine acceptance and deterred adherence to COVID-19 mitigation measures [35,36]. Likewise, parental fears and concerns about vaccine side effects, poor health literacy and information-seeking behavior, and lack of knowledge have eroded public trust in vaccines and public institutions [42]. Using four overarching themes, we offer recommendations to facilitate decision-making for the K-12 school system and bolster health promotion preparedness strategies and response for future outbreaks or pandemics: (1) implement evidence-based safety measures; (2) address misinformation using culturally appropriate, trustworthy, clear,

XSL•FO
RenderX

and timely messages; (3) prioritize in-person learning and student support services; and (4) facilitate continued collaboration and engagement of relevant stakeholders. Our findings can help health officials and policymakers design policies to support school decision-making in districts with similar sociopolitical and demographic contexts.

Themes from the SBMs, particularly concerns about student safety, teacher burnout, and confusion over public health guidance, suggest that maintaining in-person learning during outbreaks requires locally trusted, clearly communicated, and context-specific safety practices. To reduce transmission during periods of high COVID-19 community spread, or during future outbreaks, school districts should collaborate with public health authorities and health care providers to promote vaccinations, booster doses, and routine childhood immunizations, including catch-up doses for those who missed them. SBM speakers frequently described uncertainty and conflicting guidance. To address this, school districts should collaborate with local public health authorities to ensure that mitigation strategies, whether masking during periods of high transmission, improving ventilation, or offering accessible testing options, are communicated in a consistent, evidence-informed manner that reflects evolving guidance. These measures align with stakeholder concerns raised in SBMs and help reduce confusion among parents, caregivers, and teachers. Continued government-supported programs are essential to ensure equitable access to vaccines, therapeutics, and testing, particularly for high-risk populations and underserved communities.

The World Health Organization refers to the surge of misinformation during the COVID-19 pandemic as an "infodemic," highlighting the rapid spread of false or misleading information that undermines public health efforts [43]. Misinformation poses a serious threat to the public because it prolongs outbreaks and pandemics when individuals are confused and distrustful about preventive measures. Evidence-based strategies are recommended to tackle the spread of false online information, including providing training to recognize fake news, dispel pseudoscientific health beliefs and practices, and increase the visibility of accurate, science-based information across media platforms [44]. Strengthening automated systems and reliable algorithms to identify and flag false online content is crucial [45]. In addition, targeted public health approaches that include health education and promotion, sensitivity to local social norms, and meaningful community engagement are essential for rebuilding trust and supporting informed decision-making [46,47]. Public health information should be trustworthy, timely, and tailored to key audiences, that is, parents or caregivers, teachers, and students. Addressing misinformation remains essential, particularly as schools continue to navigate COVID-19 alongside other respiratory illnesses such as influenza and respiratory syncytial virus. Current guidance from leading pediatric and public health organizations emphasizes that masks are safe for children, including those with most medical conditions, and are not associated with breathing problems, learning difficulties, or anxiety [13]. Multiple studies continue to demonstrate that high-quality masks reduce the emission and spread of respiratory particles and pathogens, including coronaviruses and influenza viruses [48,49]. Information should be shared in community languages to build trust and reach non-English–speaking parents or caregivers. Surveillance and real-time reporting systems could improve access to updates for all parents or caregivers and support timely health decisions.

In-person learning is essential for tackling educational, racial, social, and nutritional disparities. Moreover, in-person learning could promote students' well-being and create a safe working environment for teachers. Conversely, school closures and disruptions to in-person learning can be associated with lost future earnings [50,51] and unfavorable economic implications [52] over the life course of affected children. The benefits of keeping schools open are profound because in-person learning facilitates real-time, personalized interactions with teachers or peers; enables healthy, supportive social relationships; and builds students' identity and sense of self [13]. After COVID-19, teachers have continued to navigate an evolving educational landscape by balancing in-person instruction with support for students recovering from disruptions. To improve future readiness for respiratory illness surges, districts can invest in staffing, provide paid leave after exposure or illness, and streamline contact tracing and illness notification in line with public health guidance [53,54].

Finally, active collaboration and engagement across all stakeholders (ie, the government, researchers, school administrators, and policymakers) remain essential. School authorities can benefit from establishing coalitions that bring together public health experts, mental health specialists, communication experts, health care providers, parents or caregivers, and teachers to offer psychological support to students who continue to face challenges related to grief, academic disruption, social isolation, and ongoing uncertainty. These partnerships can help deliver timely, locally relevant, and age-appropriate information on respiratory illnesses and mitigation practices, addressing the confusion and misinformation that speakers commonly raised during SBMs. SBM speakers frequently described burnout, limited sick leave, and insufficient support. Addressing these concerns, such as by expanding access to mental health resources, ensuring adequate paid sick leave, and promoting work-life balance policies, may strengthen teacher well-being and support school resilience during future public health challenges [55].

Strengths and Limitations

Several strengths and limitations should be considered alongside this study. One notable strength is the use of publicly available YouTube recordings of SBMs, which provide an accessible and diverse range of parental perspectives on COVID-19 mask mandates. This approach enabled the capture of real-time, spontaneous reactions to pandemic-related policies, which may not have been possible with other research methods. Additionally, the public nature of these forums offered a broad range of opinions, reflecting varying levels of support or opposition to health measures across Tennessee communities. However, this also presents a limitation, as the data were sourced from a social media platform, raising concerns about potential risks to participant privacy and confidentiality. Online forums make it hard to ensure users understand data sharing, possibly

XSL•FO
RenderX

affecting their disclosures. The lack of face-to-face interaction limits the ability to ask immediate questions, seek clarifications, and probe, which may reduce data depth. Additionally, the study examined 1 SBM per district, providing a snapshot of community discourse at a given point in time. Moreover, given that public testimony at SBMs is voluntary, speakers may represent individuals with stronger or more polarized opinions, while those with moderate views may be less inclined to speak publicly. Therefore, the perspectives captured may not fully reflect the distribution of attitudes within the broader community.

Data were collected about a year into the pandemic, around the initial vaccine rollout, which may have influenced attitudes, concerns, and debates. Public perceptions then were shaped by evolving science, policy changes, and vaccine access and acceptance. Finally, the results' transferability is limited. Tennessee's political climate, demographic makeup, community values, and local mitigation policies may differ greatly from those in other states or regions. Therefore, while themes provide insights into broader national trends, caution is needed when applying findings elsewhere without considering local sociopolitical and cultural factors. Despite these issues, the study offers valuable insights into public discussions about school mask mandates and highlights areas for improving health messaging and policies. Findings may be particularly relevant to regions characterized by political polarization, strong local control over school governance, or communities where public health measures intersect with ideological debates.

Conclusions

Future research could examine how perspectives expressed in SBMs change over time, particularly as public health guidance, vaccine availability, and community transmission patterns evolve. Longitudinal studies may reveal shifts in parental concerns, trust in health authorities, and attitudes toward mitigation during different pandemic phases. Expanding research beyond Tennessee would assess if themes were consistent across regions. Comparative analyses could identify regional differences in misinformation, policy debates, and stakeholder engagement. Further research could examine how SBM discussions affect local policies, school implementation of health recommendations, and community adherence to mitigation strategies.

This study provides real-time insight into how parents or caregivers, teachers, and community members articulated concerns about COVID-19 mitigation in public school settings during a period of active uncertainty. The themes explored illustrate how misinformation, political polarization, and competing moral frameworks shaped local decision-making. These findings are relevant and useful for preparation against future outbreaks and highlight a need for transparent communication, early engagement with caregivers, and collaboration across stakeholder groups. Recognizing diverse perspectives helps health officials and policymakers develop more effective, inclusive strategies that address public concerns and build trust. A proactive approach to policy and communication is essential to keeping educational systems adaptable, resilient, and able to safeguard health and maintain in-person learning during outbreaks.

Funding

This study was partially supported by the National Cancer Institute.

Disclaimer

Generative AI was used to assist with grammar and formatting specifications of the manuscript. No AI tools were used for data analysis, interpretation, or generation of scientific content. All content was reviewed and approved by the authors.

Data Availability

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

Authors' Contributions

Conceptualization: OO, BW, CM
Data curation: CM
Formal analysis: OO, BW, BA
Funding acquisition: AS-N
Methodology: OO, BW, CM
Project administration: AS-N
Software: OO, BW, CM
Supervision: OO, AS-N
Writing—original draft: OO, BW, BA, CM, AS-N
Writing—review and editing: OO, BW, BA, CM, AS-N

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Sharma A, Tiwari S, Deb MK, Marty JL. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): a global pandemic and treatment strategies. Int J Antimicrob Agents 2020 Aug;56(2):106054. [doi: 10.1016/j.ijantimicag.2020.106054]
2. Silk BJ, Scobie HM, Duck WM, et al. COVID-19 surveillance after expiration of the public health emergency declaration—United States, May 11, 2023. MMWR Morb Mortal Wkly Rep 2023 May 12;72(19):523-528. [doi: 10.15585/mmwr.mm7219e1] [Medline: 37167154]
3. Ripp J, Peccoralo L, Charney D. Attending to the emotional well-being of the health care workforce in a New York City health system during the COVID-19 pandemic. Acad Med 2020 Aug;95(8):1136-1139. [doi: 10.1097/ACM.0000000000003414] [Medline: 32282344]
4. Uppal A, Silvestri DM, Siegler M, et al. Critical care and emergency department response at the epicenter of the COVID-19 pandemic. Health Aff (Millwood) 2020 Aug;39(8):1443-1449. [doi: 10.1377/hlthaff.2020.00901] [Medline: 32525713]
5. Khalafallah AM, Lam S, Gami A, et al. A national survey on the impact of the COVID-19 pandemic upon burnout and career satisfaction among neurosurgery residents. J Clin Neurosci 2020 Oct;80:137-142. [doi: 10.1016/j.jocn.2020.08.012]
6. Ranney ML, Griffeth V, Jha AK. Critical supply shortages—the need for ventilators and personal protective equipment during the Covid-19 pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/NEJMp2006141] [Medline: 32212516]
7. Dvorsky MR, Shroff D, Larkin Bonds WB, Steinberg A, Breaux R, Becker SP. Impacts of COVID-19 on the school experience of children and adolescents with special educational needs and disabilities. Curr Opin Psychol 2023 Aug;52:101635. [doi: 10.1016/j.copsyc.2023.101635] [Medline: 37451025]
8. Brakefield WS, Olusanya OA, White B, Shaban-Nejad A. Social determinants and indicators of COVID-19 among marginalized communities: a scientific review and call to action for pandemic response and recovery. Disaster Med Public Health Prep 2022 May 2;17:1-10. [doi: 10.1017/dmp.2022.104] [Medline: 35492024]
9. McCarthy G, Shore S, Ozdenerol E, Stewart A, Shaban-Nejad A, Schwartz DL. History repeating—how pandemics collide with health disparities in the United States. J Racial Ethn Health Disparities 2023 Jun;10(3):1455-1465. [doi: 10.1007/s40615-022-01331-5] [Medline: 35595916]
10. Sombetzki M, Lücker P, Ehmke M, et al. Impact of changes in infection control measures on the dynamics of COVID-19 infections in schools and pre-schools. Front Public Health 2021;9:780039. [doi: 10.3389/fpubh.2021.780039] [Medline: 34988054]
11. Martin S, Vanderslott S. “Any idea how fast ‘It’s just a mask!’ can turn into ‘It’s just a vaccine!’”: from mask mandates to vaccine mandates during the COVID-19 pandemic. Vaccine (Auckl) 2022 Dec 5;40(51):7488-7499. [doi: 10.1016/j.vaccine.2021.10.031] [Medline: 34823912]
12. He L, He C, Reynolds TL, et al. Why do people oppose mask wearing? A comprehensive analysis of U.S. tweets during the COVID-19 pandemic. J Am Med Inform Assoc 2021 Jul 14;28(7):1564-1573. [doi: 10.1093/jamia/ocab047] [Medline: 33690794]
13. COVID-19 guidance for safe schools and promotion of in-person learning. American Academy of Pediatrics. 2022. URL: https://www.communitycommons.org/entities/05608ab0-b309-4a8d-a63f-9370ff6d8966 [accessed 2026-03-30]
14. Operational guidance for K-12 schools and early care and education programs to support safe in-person learning. Centers for Disease Control and Prevention. 2022. URL: https://archive.cdc.gov/#/details?url=https://www.cdc.gov/coronavirus/2019-ncov/community/schools-childcare/k-12-childcare-guidance.html [accessed 2023-03-03]
15. Decker S. Which states banned mask mandates in schools, and which required masks. Education Week. 2022. URL: https://www.edweek.org/policy-politics/which-states-ban-mask-mandates-in-schools-and-which-require-masks/2021/08 [accessed 2022-08-01]
16. Chavez-Lindell TL, Cahill KA, Kintziger KW, Odoi A. Perceptions of the impact of COVID-19 in Tennessee, USA: a retrospective study. PeerJ 2023;11:e15473. [doi: 10.7717/peerj.15473] [Medline: 37456880]
17. Lenthang M. How school board meetings have become emotional battlegrounds for debating mask mandates. ABC News Network. 2021. URL: https://abcnews.go.com/US/school-board-meetings-emotional-battlegrounds-debating-mask-mandates/story?id=79657733 [accessed 2023-03-03]
18. Ortiz-Martínez Y, Galvis-Cataño LM, Arias-Rodríguez D, Romero-Dager C, Bonilla-Aldana DK, Rodriguez-Morales AJ. YouTube and the 2022 Monkeypox outbreak: opportunities for awareness and infection control. J Hosp Infect 2024 Oct;152:173-174. [doi: 10.1016/j.jhin.2022.07.007] [Medline: 35835284]
19. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/04/PI_2021.04.07_Social-Media-Use_FINAL.pdf [accessed 2023-03-03]
20. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85. [doi: 10.2196/jmir.1933] [Medline: 23615206]
21. Larson HJ, Gakidou E, Murray CJL. The vaccine-hesitant moment. N Engl J Med 2022 Jul 7;387(1):58-65. [doi: 10.1056/NEJMra2106441] [Medline: 35767527]
22. Hubbard K. These states have COVID-19 mask mandates. US News & World Report. 2022. URL: https://www.usnews.com/news/best-states/articles/these-are-the-states-with-mask-mandates [accessed 2023-03-03]

XSL•FO
RenderX

23. Tennessee COVID-19 vaccine tracker. The Tennessean. 2022. URL: https://data.tennessean.com/covid-19-vaccine-tracker/tennessee/47/ [accessed 2023-03-03]
24. Burke SC, Snyder S, Rager RC. An assessment of faculty usage of YouTube as a teaching resource. Internet J Allied Health Sci Pract 2009;7(1):8. [doi: 10.46743/1540-580X/2009.1227]
25. Olusanya OA, White B, Malik F, et al. Healthcare professionals' perceptions and recommendations regarding adolescent vaccinations in Georgia and Tennessee during the COVID-19 pandemic: a qualitative research. PLoS One 2022;17(11):e0277748. [doi: 10.1371/journal.pone.0277748] [Medline: 36399477]
26. Johnson JL, Adkins D, Chauvin S. A review of the quality indicators of rigor in qualitative research. Am J Pharm Educ 2020 Jan;84(1):7120. [doi: 10.5688/ajpe7120] [Medline: 32292186]
27. Houghton C, Casey D, Shaw D, Murphy K. Rigour in qualitative case-study research. Nurse Res 2013 Mar;20(4):12-17. [doi: 10.7748/nr2013.03.20.4.12.e326] [Medline: 23520707]
28. Flicker S, Haans D, Skinner H. Ethical dilemmas in research on Internet communities. Qual Health Res 2004 Jan;14(1):124-134. [doi: 10.1177/1049732303259842] [Medline: 14725180]
29. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):708-713. [doi: 10.1089/cyber.2012.0334] [Medline: 23679571]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Soratto J, Pires DD, Friese S. Thematic content analysis using ATLAS.ti software: potentialities for researchs in health. Rev Bras Enferm 2020;73(3):e20190250. [doi: 10.1590/0034-7167-2019-0250] [Medline: 32321144]
32. Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube. PLoS One 2014;9(10):e110171. [doi: 10.1371/journal.pone.0110171] [Medline: 25333470]
33. White BM, Melton C, Zareie P, Davis RL, Bednarczyk RA, Shaban-Nejad A. Exploring celebrity influence on public attitude towards the COVID-19 pandemic: social media shared sentiment analysis. BMJ Health Care Inform 2023 Jan;30(1):e100665. [doi: 10.1136/bmjhci-2022-100665] [Medline: 36810135]
34. DeMora SL, Merolla JL, Newman B, Zechmeister EJ. Reducing mask resistance among White evangelical Christians with value-consistent messages. Proc Natl Acad Sci U S A 2021 May 25;118(21):e2101723118. [doi: 10.1073/pnas.2101723118] [Medline: 33975906]
35. Kim HK, Tandoc EC. Consequences of online misinformation on COVID-19: two potential pathways and disparity by eHealth literacy. Front Psychol 2022;13:783909. [doi: 10.3389/fpsyg.2022.783909] [Medline: 35237207]
36. Scales D, Gorman J, Jamieson KH. The COVID-19 infodemic—applying the epidemiologic model to counter misinformation. N Engl J Med 2021 Aug 19;385(8):678-681. [doi: 10.1056/NEJMp2103798] [Medline: 33979506]
37. Shaban-Nejad A, Michalowski M. From precision medicine to precision health: a full angle from diagnosis to treatment and prevention. In: Shaban-Nejad A, Michalowski M, editors. Precision Health and Medicine. W3PHAI 2019. Studies in Computational Intelligence: Springer; 2020, Vol. 843. [doi: 10.1007/978-3-030-24409-5_1]
38. Perry SL, Whitehead AL, Grubbs JB. Prejudice and pandemic in the promised land: how White Christian nationalism shapes Americans' racist and xenophobic views of COVID-19. In: Race and Ethnicity in Pandemic Times: Routledge, Vol. 2021:41-54. [doi: 10.4324/9781003206521-4]
39. Smothers H, Burge R, Djupe P. The gendered religious response to state action on the coronavirus pandemic. Pol Gen 2020 Dec;16(4):1063-1074. [doi: 10.1017/S1743923X20000306]
40. Bloemraad I, Silva F, Voss K. Rights, economics, or family? Frame resonance, political ideology, and the immigrant rights movement. Social Forces 2016 Jun;94(4):1647-1674. [doi: 10.1093/sf/sov123]
41. Djupe P. God Talk: Experimenting With the Religious Causes of Public Opinion: Temple University Press; 2014. [doi: 10.2307/j.ctvrf8960]
42. Nicomedes CJC, Avila RMA. An analysis on the panic during COVID-19 pandemic through an online form. J Affect Disord 2020 Nov 1;276:14-22. [doi: 10.1016/j.jad.2020.06.046] [Medline: 32697692]
43. Munich security conference. World Health Organization. 2020. URL: https://www.who.int/news-room/speeches/item/munich-security-conference [accessed 2022-08-01]
44. Naeem SB, Bhatti R, Khan A. An exploration of how fake news is taking over social media and putting public health at risk. Health Info Libraries J 2021 Jun;38(2):143-149. [doi: 10.1111/hir.12320]
45. Molina MD, Sundar SS, Le T, Lee D. "Fake news" is not simply false information: a concept explication and taxonomy of online content. Am Behav Sci 2021 Feb;65(2):180-212. [doi: 10.1177/0002764219878224]
46. Vergara RJD, Sarmiento PJD, Lagman JDN. Building public trust: a response to COVID-19 vaccine hesitancy predicament. J Public Health (Oxf) 2021 Jun 7;43(2):e291-e292. [doi: 10.1093/pubmed/fdaa282]
47. Infodemic. World Health Organization. 2022. URL: https://www.who.int/health-topics/infodemic#tab=tab_1 [accessed 2022-08-01]
48. Howard J, Huang A, Li Z, et al. An evidence review of face masks against COVID-19. Proc Natl Acad Sci USA 2021 Jan 26;118(4):e2014564118. [doi: 10.1073/pnas.2014564118]

XSL•FO
RenderX

49. Leung NHL, Chu DKW, Shiu EYC, et al. Author correction: Respiratory virus shedding in exhaled breath and efficacy of face masks. Nat Med 2020 Jun;26(6):981-981. [doi: 10.1038/s41591-020-0946-9] [Medline: 32461699]
50. Jaume D, Willén A. The long-run effects of teacher strikes: evidence from Argentina. J Labor Econ 2019 Oct;37(4):1097-1139. [doi: 10.1086/703134]
51. Psacharopoulos G, Collis V, Patrinos HA, Vegas E. Lost wages: the COVID-19 cost of school closures. SSRN J 2020 Aug 27. [doi: 10.2139/ssrn.3682955]
52. Lempel H, Epstein JM, Hammond RA. Economic cost and health care workforce effects of school closures in the U.S. PLoS Curr 2009 Oct 5;1:RRN1051. [doi: 10.1371/currents.rrn1051] [Medline: 20025205]
53. Godber KA, Atkins DR. COVID-19 impacts on teaching and learning: a collaborative autoethnography by two higher education lecturers. Front Educ 2021 Jul 29;6. [doi: 10.3389/feduc.2021.647524]
54. Donohue JM, Miller E. COVID-19 and school closures. JAMA 2020 Sep 1;324(9):845-847. [doi: 10.1001/jama.2020.13092] [Medline: 32745182]
55. Ju C, Lan J, Li Y, Feng W, You X. The mediating role of workplace social support on the relationship between trait emotional intelligence and teacher burnout. Teaching and Teacher Education 2015 Jun;51(1):58-67. [doi: 10.1016/j.tate.2015.06.001]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
SBM: school board meeting

Edited by E Lee; submitted 08.Apr.2025; peer-reviewed by AWK Yeung, A Afzalian, C Garrett; revised version received 09.Feb.2026; accepted 23.Feb.2026; published 20.Apr.2026.

Please cite as:
Olusanya OA, White BM, Amuchi B, Melton C, Shaban-Nejad A
Parents' Multiperspective Views and Misinformation About COVID-19 Mitigation Measures During Tennessee School Board Meetings: Qualitative Content Analysis Using YouTube
JMIR Infodemiology 2026;6:e75691
URL: https://infodemiology.jmir.org/2026/1/e75691
doi:10.2196/75691

© Olufunto A Olusanya, Brianna M White, Brenda Amuchi, Chad Melton, Arash Shaban-Nejad. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 20.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Implementation and User Evaluation of the SANGYAN Digital Health Platform to Enhance Knowledge About COVID-19 and Other Health Conditions: Quasi-Experimental Study

Ashish Joshi[1], PhD; Ashruti Bhatt[2], MPH; Surapaneni Krishna Mohan[3], PhD; Gunjan Malhotra[2], MBBS; Doilyn Oliveira[2], MPH; Saravanavel Kalpana Revathi[2], MPH; Ashoo Grover[4], MD

[1]School of Public Health, University of Memphis, Robison Hall 3825 DeSoto Avenue, Memphis, TN, United States

[2]Foundation of Healthcare Technologies Society, Delhi, India

[3]Panimalar Medical College Hospital & Research Institute, Chennai, India

[4]Indian Council of Medical Research, Delhi, India

Corresponding Author:
Ashish Joshi, PhD
School of Public Health, University of Memphis, Robison Hall 3825 DeSoto Avenue, Memphis, TN, United States

Abstract

Background: The spread of misinformation during the COVID-19 pandemic highlighted the importance of evidence-based information. The SANGYAN podcast promotes evidence-based knowledge on health-related issues in multiple languages in a simple, cost-effective, and concise manner. This provides individuals access to the appropriate information in an accessible manner.

Objective: The study's goal is to assess user preferences for health information on a digital health platform designed to address COVID-19 misinformation.

Methods: SANGYAN was developed by integrating the principles of social cognitive theory and information processing theory. The SANGYAN podcast was created to promote the importance of evidence-based information in order to address the spread of misinformation. The study design was a quasi-experimental study; prior to introducing the SANGYAN podcast, participants' sociodemographic information was collected, and health literacy was assessed using the Rapid Estimate of Adult Literacy in Medicine, Revised scale. After listening to the podcast, participants were interviewed about its usability, and they completed the System Usability Scale and the Client Satisfaction Questionnaire - 8. Data were collected from a total of 500 participants, 250 each from the Rural Health Training Center and Panimalar Medical College Hospital & Research Institute. The participants were older than 18 years when they were included. Descriptive and bivariate analyses were performed.

Results: A total of 500 participants were enrolled in the study, 50% (250/500) from rural areas and 50% (250/500) from urban areas. The majority of the participants were 45 years to 64 years old (155/500, 31%), were women (289/500, 57.8%), had poor health literacy (384/500, 76.8%), and had a high school education or less than a high school certification (241/500, 48.2%). The mean overall System Usability System score was 70.9 (SD 17.73), with those aged 18 years to 24 years having the highest mean score (81.2, SD 15.48). High user satisfaction was present, with 97.6% (487/499) obtaining the desired information from the platform.

Conclusions: The study revealed that the SANGYAN podcast provides information to diverse individuals, as it is multilingual, and was found useful by the participants.

(JMIR Infodemiology 2026;6:e67504) doi:10.2196/67504

KEYWORDS

podcast; usability system; client satisfaction; misinformation; COVID-19; SARS-CoV-2; pandemic; SANGYAN podcast; evidence-based; digital health intervention; usability; health literacy; REALM scale; System Usability Scale; SUS; client satisfaction questionnaire

Introduction

Background

SARS-CoV-2, a new coronavirus, spread rapidly throughout the world. More than 200 countries and territories have been impacted by COVID-19, which started in 2019 and is still having an impact [1]. Low- and middle-income countries (LMIC) were severely impacted during the pandemic in comparison with higher-income countries [2]. Along with the spread of the virus,

XSL•FO
RenderX

there was arapid circulation of misleading information on social media platforms [3].

In 1918, the influenza pandemic affected one-third of the global population and resulted in 50 million deaths worldwide. During that time, medical therapies, countermeasures, and information exchange or public health platforms were primarily accomplished via telephone, mail, or person-to-person interactions [4]. Currently, with more than 2.9 billion people using social media globally on a regular basis, information sharing has increased significantly [5,6]. Digital media was the primary means of disseminating COVID-19 information [7]. At the same time, digital media has resulted in the both knowing and unknowing spread of misleading and false information, which can influence public opinions and behaviors and have serious consequences by positively or negatively manipulating the common person's perspective [8]. Unfortunately, this has led to people trying harmful remedies and the creation of fear and mistrust [7,9].

LMIC in particular faced unique challenges due to limited resources, poor infrastructure, unregulated social media, and a lack of crisis communication, all of which contributed to obscuring knowledge of COVID-19, panic, and confusion in these regions [10,11]. Misinformation was widespread through social media platforms such as Facebook, Twitter, and online newspapers [12]. Related literature showed that misinformation can induce anxiety, fear, and mistrust in reliable organizations through the sharing of critical and negative opinions [13]. There is evidence that misleading information is likely to spread more rapidly when finding reliable knowledge is hard and when people have low trust on the information sources accessible to them. It is important to understand that COVID-19–related information and myths are not simple enough for the general public to distinguish [14].

The World Health Organization (WHO) addressed digital health as a global strategy and defined digital health as "the field of knowledge and practice associated with the development and use of digital technologies to improve health" [15]. Although digital technologies have amplified misinformation, they can also be a source of solutions [14]. One such promising approach is the use of a targeted podcast that is contextually relevant and regionally tailored.

Digital communication technologies play a vital role in health communications and can transmit health promotion into knowledge translation in communities for public health and global health practice [16,17]. Podcasts are an effective and practical strategy for delivering and disseminating quality health-related information to the public [18]. They are a valuable resource because they provide participants with unrestricted access to necessary information in a simple, effective, and concise format [19]. Several studies have investigated the acceptability or feasibility of this mode of learning, but limited research has been conducted on the use of podcasts as a knowledge-sharing medium or platform [20]. The COVID-19 pandemic increased the number of podcast users in India, which reached up to 29.3% [21]. Podcasts enable the rapid dissemination of up-to-date information and provide listeners with a wide range of perspectives (eg, patient perspective) [22]. The development of audio or visual material for a listenership that wants to hear what they want, when they want, where they want, and how they want is at the core of podcasting [23].

In response to the increasing need for efficient health communication strategies, our study introduces SANGYAN (Multimedia Appendices 1 and 2) as a multilingual, interactive, decision-making, educational audio (IDEA) podcast designed to deliver and disseminate evidence-based COVID-19 and other health-related information in a format that is easy to understand. It takes into consideration the health literacy and cultural and contextual relevance of the populations living in urban and rural regions of Tamil Nadu, in the southern state of India. The study assessed the usefulness of this digital health platform.

SANGYAN Podcast as a Digital Health Platform to Deliver Evidence-Based Health Information

The importance of leveraging digital technologies to fight against infodemics has been well recognized. Recommendations include creating educational materials, increasing the dissemination of evidence-based science, correcting behaviors, and encouraging healthy practices to address current misconceptions. SANGYAN is one such podcast-based digital platform that is evidence-based and encourages healthy practices (Figure 1) [24]. Previous studies have also recognized podcasts as a suitable source to disseminate information to improve the knowledge of their participants; similarly, SANGYAN aims to debunk misinformation at a large scale [25-27]. The SANGYAN podcast study engaged 500 study participants, 250 from rural and 250 from urban areas. The initiative aimed to promote accurate information dissemination to the public to prevent misinterpretation. The podcast does not require internet access, making it suitable for use in rural areas with limited connectivity. An open booth was established on the hospital premises, ensuring visibility to patients, relatives, and health care professionals visiting the facility. The study was self-funded and conducted in collaboration with Panimalar Medical College Hospital & Research Institute (PMCHRI).

XSL•FO
RenderX

Figure 1. A multilingual, interactive, decision-making, educational audio (IDEA) podcast called SANGYAN[28].



Design and development of SANGYAN [28] involved creating a multilingual, IDEA podcast aiming to deliver evidence-based health information to individuals and the community in a clear, reliable, easy-to-understand, and culturally relevant manner. Additionally, listeners can access it easily, as it offers a free subscription and free audio download service [28] as an innovative multilingual platform designed to prevent the spread of misinformation and provide accurate information to the public. The platform was developed by a multidisciplinary team including experienced health care professionals such as nutritionists, public health professionals, and medical professionals who contributed comprehensive knowledge on various topics tailored to the needs and requirements of the community. Additionally, the communication team was responsible for collecting and disseminating health care information as well as editing audio content, while the informatics team ensured technical compliance and quality standards. Some episodes feature expert-to-expert discussions, while most include audience questions, making the SANGYAN podcasts an engaging and interactive educational tool.

SANGYAN was developed combining the principles of social cognitive theory (SCT) and information processing theory. SCT is a contemporary learning theory positing interactions among behavioral, social, and environmental factors that collectively influence human behavior. It provides a framework for understanding how individuals respond to and engage with learning opportunities, motivation, and self-regulation. In our context, we applied this theory to address societal misconceptions about various noncommunicable diseases (NCDs), COVID-19, and related health issues, with the goal of increasing public awareness. By leveraging the principles of SCT, we aimed to improve environmental factors by providing evidence-based information and enhancing knowledge, thereby counteracting misinformation [29-31]. Information processing theory aims to elucidate the mechanisms underlying human information processing, rather than solely describing behavioral responses to environmental stimuli. It posits that cognitive development is not merely a sequential progression toward a specific outcome but is also modulated by internal thoughts, prior experiences, sensory input, and perceptual processes. Additionally, it highlights the significant role of social and cultural factors in shaping cognition and learning paradigms. In addition, using the principles of information processing theory, it uses the fact that listening to podcasts helps process the information as a sensory input, which is further processed into short-term memory then saved as a long-term memory in the brain due to the fact that the information is important [32-34].

SANGYAN consists of podcasts broadly on the categories of COVID-19, NCDs, and menstrual practices. SANGYAN provides evidence-based information created by public health experts for all health topics such as diabetes, hypertension, stress, and obesity in an interactive, multilingual way. Currently, SANGYAN is available in English, Tamil, Telugu, Hindi, Gujarati, Kannada, and Malayalam.

This study represents a novel approach with the aim of addressing the critical impact of health misinformation, especially during the COVID-19 pandemic, by establishing the effectiveness of the multilingual podcast SANGYAN as a digital health tool for disseminating evidence-based information across various educational backgrounds and socioeconomic groups. Unlike traditional methods or internet-based methods, our study aimed to use podcasts to increase access to health care information without the use of the internet, even in most rural areas, in addition to being user-friendly and culturally appropriate. The quasi-experimental design addresses three key challenges: health equity, knowledge misinterpretation, and accessibility barriers. This study uniquely highlights the potential of audio-based learning systems, especially among a population with poor literacy and who is digitally naive, for enhancing equitable access to reliable health information.

XSL•FO
RenderX

Methods

Study Design

This pre-post, quasi-experimental study was conducted from June 2022 through August 2022 in the PMCHRI and Rural Health Training Centre (RHTC) in Chennai. Individuals present at these locations were approached and provided with a briefing of the study, and those who consented were included as participants. Each participant enrolled in the study was oriented on how to use the platform, after which the participants self-navigated the platform, selected any topic of interest, and listened to the audio podcast in a multilingual setting, based on their interest, relevance, and preferences including preferred language.

For this study, using the intervention was defined as accessing and listening to the SANGYAN podcast at least once as well as completing the postuse questionnaire. The SANGYAN platform is an open digital health platform. The podcast covered several COVID-19 topics, with all evidence-based health information created by public health experts in an interactive, multilingual format. The goal was to deliver and disseminate COVID-19–related health information to individuals and communities living in urban and rural settings in the most accurate and reliable manner and address any misinformation they might have regarding COVID-19. Satisfaction and usefulness of the SANGYAN podcast were assessed using the Client Satisfaction Questionnaire - 8 (CSQ-8) [35] and System Usability Scale (SUS) [36].

Study Setting

The study was conducted in the RHTC, Thiruvallur District, Tamil Nadu, and PMCHRI, Chennai, Tamil Nadu.

Sampling and Sample Size

A total sample of 500 individuals (n=250 each at the PMCHRI and RHTC) aged at least 18 years were selected using a convenience sampling method.

Data Collection Process

The data collectors received 1 day of intensive training on the data collection process. The data collectors received a brief orientation on how to obtain informed consent and use the health literacy scale (Rapid Estimate of Adult Literacy in Medicine, Revised [REALM-R] examiner record) [37,38], SUS [36], and CSQ-8 [35]. To ensure the study's consistency, one episode of the SANGYAN podcast series was shared with the research team. Furthermore, the team was trained in the local Indian dialect Tamil as well as English.

Ethical Considerations

Institutional ethical clearance was provided by the PMCHRI Institutional Human Ethics Committee (CDSCO registration number ECR/1399/Inst/TN/2020 with approval number PMCH&RI/IHEC/2021/037 dated January 13, 2021). The study was conducted according to the guidelines set forth in the Declaration of Helsinki [39].

Written informed consent was obtained from each participant. For cases of illiteracy, the consent form was read to the participant, then written consent was sought. Complete confidentiality of information collected from participants was ensured. Data security was maintained through regular backups of all computers, and data files were kept password-protected. Participants were not given special incentives in cash or kind for participation in the study. As the study contributed to knowledge, there was no risk involved in participating in the study at any time.

Variables and Measurements

Data were gathered in 2 stages. The first stage occurred prior to listening to the podcast and included data collection regarding sociodemographic characteristics, health literacy, and any prior use of podcasts or other digital medium for acquiring health information. The second stage included the collection of information after the participants listened to the podcast. This included the SUS to assess the usability and acceptance of the SANGYAN podcast and CSQ-8 to assess satisfaction with using the SANGYAN podcast as a digital health intervention to acquire evidence-based, accurate, and reliable health information in a format that is easy to understand. Individuals were asked to provide their feedback after listening to the same episode of the SANGYAN podcast to ensure consistency. Each of these variables are described in detail in the following paragraphs and in Multimedia Appendix 3.

Sociodemographic information included age, gender, education level, religion, region of residence, and employment status.

The REALM-R examiner record was also used to evaluate participants' health literacy. It is an updated tool for estimating adult literacy in medicine. It is an 11-item word recognition test that is used to identify people who are at risk of having poor literacy skills [37,38].

To assess the use of podcasts or other media for health information, participants were asked about their prior use of podcasts to acquire health information.

The SUS measures the acceptance and usability of a system using a subjective assessment of 10 usability statements. Responses consist of 5 options, from strong agreement to strong disagreement [36]. The SUS scores were calculated with reference to the guidance outlined by Sauro [40]. The scores for each question were converted to a new number (ie, the questionnaire consisted of 10 questions that are answered on a 5-point Likert scale ranging from strongly disagree to strongly agree). On the SUS, odd-numbered statements (1, 3, 5, 7, 9) are positively expressed, and even-numbered statements (2, 4, 6, 8, 10) are negatively expressed. The responses were summed then multiplied by 2.5 to convert the original scores, which range from 0 to 40, to a range of 0 to 100. We interpreted the SUS scores using a cutoff of 68, with a score greater than 68 indicating average performance and a score less than 68 considered to be below average.

User satisfaction was assessed using the CSQ-8, which consists of 8 items and assessed general satisfaction of individuals after using the podcast [35]. The questions collect respondents' opinions and conclusions about services they have received. Response options differ by item, but they are all based on a 4-point scale.

XSL•FO
RenderX

Data Analysis

Descriptive analysis was conducted to outline the percentages and frequencies of the variables. Exploratory bivariate analysis was conducted to understand the correlations between sociodemographic features and their perceptions, the usefulness of the podcast, and acceptance of the podcast. No further adjusted nor multivariate analyses were conducted. A *P* value <.05 was considered significant, with 95% CIs. also considered All analyses were performed using SPSS v.24 (IBM Corp).

Results

Sociodemographic Details of the Study Participants

The study involved 500 participants, with 250 each from the PMCHRI, Chennai, Tamil Nadu (urban setting), and the RHTC, Thiruvallur District (rural setting). The majority of the participants (155/500, 31%) were in the age group of 45 years to 64 years old and were women (289/500, 57.8%). In addition, 11.8% (59/500) had no formal education, 44.6% (223/500) were employed, and more than one-half (384/500, 76.8%) had poor health literacy (Table 1). Rural and urban comparisons showed that a higher percentage of participants from rural settings had no formal education than those living in urban settings (45/250, 18% versus 14/250, 5.6%). A similar disparity in health literacy status was observed, with a higher proportion of participants living in rural settings having poor health literacy than those living in urban settings (217/250, 86.8% versus 167/250, 66.8%; Table 1).

XSL•FO
RenderX

Table . Sociodemographic characteristic of the participants according to region of residence.

Variables	Total sample (N=500), n (%)	Rural-dwelling participants (n=250), n (%)	Urban-dwelling participants (n=250), n (%)
Age group (years)			
18 - 24	76 (15.2)	26 (10.4)	50 (20)
25 - 35	138 (27.6)	76 (30.4)	62 (24.8)
36 - 44	105 (21)	49 (19.6)	56 (22.4)
45 - 64	155 (31)	87 (34.8)	68 (27.2)
≥65	26 (5.2)	12 (4.8)	14 (5.6)
Gender			
Male	211 (42.2)	97 (38.8)	114 (45.6)
Female	289 (57.8)	153 (61.2)	136 (54.4)
Educational level			
No formal education	59 (11.8)	45 (18)	14 (5.6)
High school or less than high school certification	241 (48.2)	148 (59.2)	93 (37.2)
Graduate and higher	168 (33.6)	43 (17.2)	125 (50)
Intermediate or diploma	32 (6.4)	14 (5.6)	18 (7.2)
Employment status			
Employed	253 (50.6)	110 (44)	143 (57.2)
Homemaker	110 (22)	78 (31.2)	32 (12.8)
Unemployed	98 (19.6)	50 (20)	48 (19.2)
Other (retired, n=14; student, n=24; lost job due to COVID-19, n=1)	39 (7.8)	12 (4.8)	27 (10.8)
Religion			
Christian	50 (10)	21 (8.4)	29 (11.6)
Hindu	434 (86.8)	223 (89.2)	211 (84.4)
Muslim	10 (2)	4 (1.6)	6 (2.4)
Unwilling to disclose	6 (1.2)	2 (0.8)	4 (1.6)
Health literacy (REALM-R[a])			
Adequate health literacy	116 (23.2)	33 (13.2)	83 (33.2)
Poor health literacy	384 (76.8)	217 (86.8)	167 (66.8)

[a]REALM-R: Rapid Estimate of Adult Literacy in Medicine, Revised.

Assessment of Health Literacy Using the REALM-R

The REALM-R was used to assess participants' familiarity with medical jargon and to identify those at risk of poor health literacy. According to REALM-R grading, those who obtained a score ≤6 were considered to have poor health literacy, whereas those with a score >6 were considered to have adequate health literacy [38]. Table 1 shows that, overall, 76.8% (384/500) of the respondents found it challenging to read the REALM-R chart's medical terminologies, while 23.2% (116/500) of respondents had adequate health literacy levels. The table also shows differences in health literacy levels between rural and urban settings; 86.8% (217/250) of rural residents and 66.8% (167/250) of urban residents had poor health literacy.

Difference Between the Use of the Podcast or Other Medium to Access Relevant Health Information

Approximately 89.6% (443/500) of the total sample reported that they had never used nor heard of a podcast, with 81.2% (203/250) of urban residents and 96% (240/250) of rural residents indicating the same. Only 17.6% (44/250) of urban-dwelling participants and 3.2% (8/250) of rural-dwelling participants used podcasts to access relevant health information. Around 90.4% (24/250) of participants from rural areas and 80% (200/250) of urban-dwelling participants said they did not use any app to receive pertinent health-related information. However, 97.4% (487/500) of the study participants believed that SANGYAN podcasts gave them the necessary, pertinent information (Table 2), and 97.7% (375/379) of the participants

XSL•FO
RenderX

with poor literacy felt that the podcast provided relevant information.

Table . Sociodemographic characteristics and their relationship with health information access.

Sociodemographic variables	SANGYAN podcasts provided you with relevant information (yes), n (%)	Used podcasts before (yes), n (%)	Use podcasts for health-related information (yes), n (%)	Use SMS for health-related information (yes), n (%)	Use WhatsApp for health-related information (yes), n (%)	Use any app for health-related information (yes), n (%)
Total sample (N=500)	487 (97.4)	52 (10.4)	152 (30.4)	76 (15.2)	108 (21.6)	74 (14.8)
Geographic setting						
Urban (n=250)	241 (96.4)	44 (17.6)	29 (11.6)	68 (27.2)	78 (31.2)	50 (20)
Rural (n=250)	246 (98.4)	8 (3.2)	123 (49.2)	8 (3.2)	30 (12)	24 (9.6)
Age group (years)						
18 - 24 (n=76)	73 (96.1)	13 (17.1)	15 (19.7)	16 (21.1)	16 (21.1)	13 (17.1)
25 - 35 (n=138)	133 (96.4)	8 (5.8)	55 (39.9)	15 (10.9)	34 (24.6)	30 (21.7)
36 - 44 (n=105)	103 (98.1)	13 (12.4)	41 (39)	15 (14.3)	27 (25.7)	17 (16.2)
45 - 64 (n=155)	155 (100)	17 (11)	39 (25.2)	27 (17.4)	28 (18.1)	12 (7.7)
≥65 (n=26)	23 (88.5)	1 (3.8)	2 (7.7)	3 (11.5)	3 (11.5)	2 (7.7)
Gender						
Female (n=289)	283 (97.9)	28 (69.7)	105 (36.3)	46 (15.9)	59 (20.4)	36 (12.5)
Male (n=211)	204 (96.7)	24 (11.4)	47 (22.3)	30 (14.2)	49 (23.2)	38 (18)
Educational level						
No formal education (n=59)	57 (96.6)	3 (5.1)	13 (22)	1 (1.7)	2 (3.4)	0 (0)
High school or less than high school certification (n=241)	236 (97.9)	8 (3.3)	118 (49)	29 (12)	79 (32.8)	15 (6.2)
Graduate and higher (n=168)	163 (97)	39 (23.2)	40 (23.8)	41 (24.4)	66 (39.3)	53 (31.5)
Intermediate or diploma (n=32)	31 (96.9)	2 (6.3)	4 (12.5)	5 (15.6)	7 (21.9)	6 (18.8)
REALM-R[a]						
Poor health literacy (n=384)	375 (97.7)	23 (6)	127 (33.1)	51 (13.3)	61 (15.9)	29 (7.6)
Adequate health literacy (n=116)	112 (96.6)	29 (25)	25 (21.6)	25 (21.6)	47 (40.5)	45 (38.8)
Employment status						
Employed (n=253)	246 (97.2)	34 (13.4)	67 (26.5)	45 (17.8)	69 (27.3)	55 (21.7)
Homemaker (n=110)	107 (97.3)	6 (5.5)	59 (53.6)	6 (5.5)	16 (14.5)	9 (8.2)
Unemployed (n=98)	97 (99)	7 (7.1)	24 (24.5)	16 (163)	19 (19.4)	5 (5.1)
Other (n=39: retired, n=14; student, n=24; lost job due to COVID-19, n=1)	37 (94.9)	5 (12.8)	2 (5.1)	9 (23.1)	4 (10.3)	5 (12.8)

[a]REALM-R: Rapid Estimate of Adult Literacy in Medicine, Revised.

Usability of and Satisfaction With the SANGYAN Podcast

Table 3 shows the mean SUS scores with respect to sociodemographic variables. The overall mean SUS score was 70.9 (SD 17.73). By age group, participants aged from 18 years to 24 years had the highest mean score of 81.5 (SD 15.6). When compared by health literacy levels, participants with poor or adequate health literacy both demonstrated above average mean SUS scores (68.5, SD 17.2 and 78.7, SD 17.2, respectively).

XSL•FO
RenderX

Table . Mapping of sociodemographic variables with usability and satisfaction scores for the SANGYAN podcast (N=500).

Sociodemographic variables	System Usability Scale (SUS) score, mean (SD)	Score category, n (%)	
		<68 points	≥68 points
Age group (years)			
18 - 24 (n=76)	81.5 (15.6)	18 (23.7)	58 (76.3)
25 - 35 (n=138)	73.8 (17.6)	64 (46.4)	74 (53.6)
36 - 44 (n=105)	72.1 (16)	50 (47.6)	55 (52.4)
45 - 64 (n=155)	63.7 (16.5)	101 (65.2)	54 (34.8)
≥65 (n=26)	62.1 (18)	17 (65.4)	9 (34.6)
Gender			
Female (n=289)	69.3 (16.3)	151 (52.2)	138 (47.8)
Male (n=211)	73 (19.4)	99 (46.9)	112 (53.1)
Educational level			
No formal education (n=59)	60.7 (18.5)	41 (69.5)	18 (430.5)
High school or less than high school certification (n=241)	66.7 (16.6)	149 (61.8)	92 (38.2)
Graduate and higher (n=168)	80.2 (16.9)	47 (28)	121 (72)
Intermediate or diploma (n=32)	77.3 (16.1)	13 (40.6)	19 (59.4)
Health literacy			
Poor literacy (n=384)	68.5 (17.2)	216 (56.3)	168 (43.8)
Adequate literacy (n=116)	78.7 (17.2)	34 (29.3)	82 (70.7)
Region of residence			
Rural (n=250)	67.8 (17.5)	148 (59.2)	102 (40.8)
Urban (n=250)	74 (17.5)	102 (40.8)	148 (59.2)
Employment status			
Employed (n=253)	72.2 (16.2)	112 (44.3)	141 (55.7)
Homemaker (n=110)	67.1 (15.2)	67 (60.9)	43 (39.1)
Unemployed (n=98)	60.5 (17.9)	61 (62.2)	38 (38.8)
Others (n=39: retired, n=14; student, n=24; lost job due to COVID-19, n=1)	75.6 (15.8)	10 (25.6)	28 (71.8)

Participant Satisfaction With SANGYAN: CSQ-8

Table 4 shows the percentages of participants who were satisfied with SANGYAN, where scores of 1 or 2 indicate dissatisfaction and scores of 3 or 4 indicate satisfaction. Of the participants, 94.8% (474/500) were satisfied with the quality of SANGYAN, and 97.4% (487/500) received the kind of information they desired from SANGYAN. In addition, 93% (465/500) of the participants were likely to use SANGYAN again, and 97% (485/500) of the participants would recommend SANGYAN to their friends.

XSL•FO
RenderX

Table . Client Satisfaction Questionnaire - 8 (CSQ-8) scores for satisfaction levels when using the SANGYAN podcast (N=500).

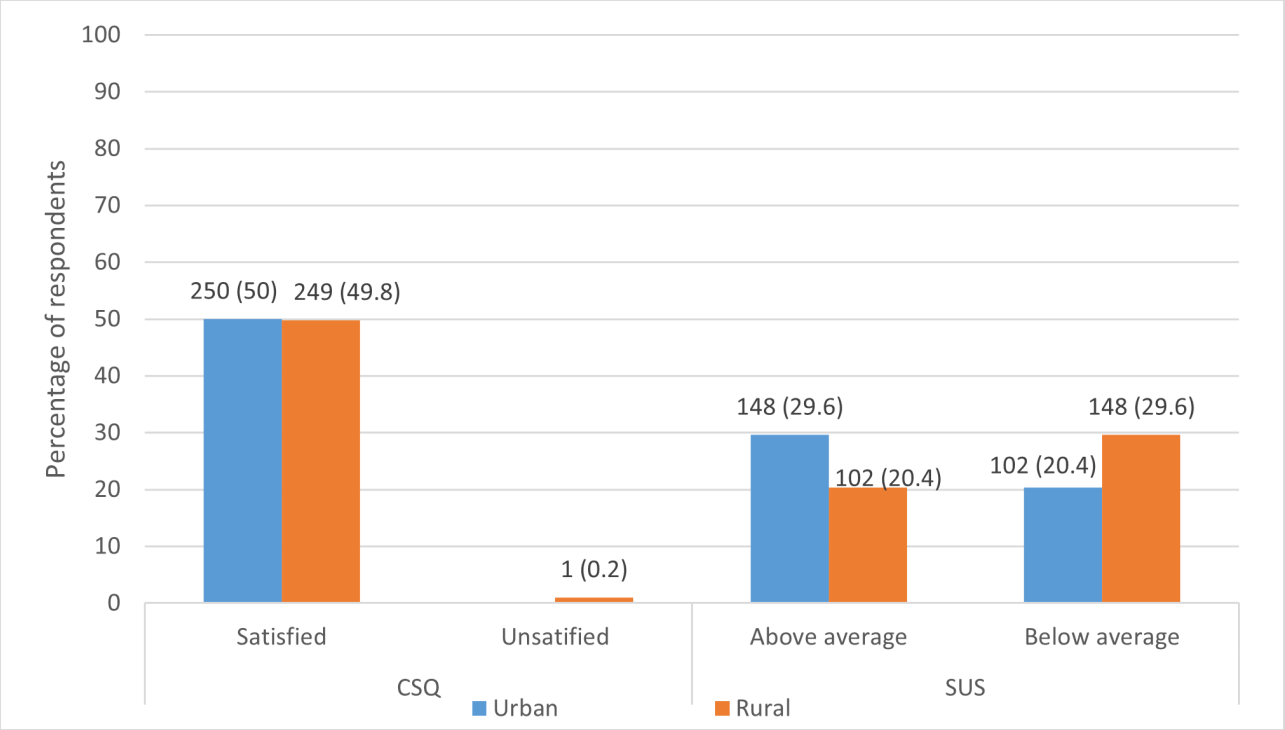
CSQ-8 questions	Score, n (%)			
	1	2	3	4
How would you rate the quality of service you received?	7 (1.4)	16 (3.2)	237 (47.4)	237 (47.4)
Did you get the kind of service you wanted?	0 (0)	12 (2.4)	211 (42.2)	276 (55.2)
To what extent has our service met your needs?	1 (0.2)	37 (7.4)	219 (43.8)	242 (48.4)
If a friend were in need of similar help, would you recommend our service to him or her?	5 (1)	9 (1.8)	174 (34.8)	311 (62.2)
How satisfied are you with the amount of help you received?	49 (9.8)	19 (3.8)	233 (46.6)	198 (39.6)
Have the services you received helped you to deal more effectively with your problems?	0 (0)	11 (2.2)	179 (35.8)	309 (61.8)
In an overall, general sense, how satisfied are you with the service you received?	5 (1)	14 (2.8)	203 (40.6)	271 (54.2)
If you were to seek help again, would you come back to our service?	9 (1.8)	18 (3.6)	138 (27.6)	327 (65.4)

Usability of and Satisfaction With the SANGYAN Podcast as a Digital Health Platform

Figure 2 illustrates the usability of and satisfaction with the SANGYAN podcast as a digital health platform to obtain evidence-based health information in a multilingual, interactive format that is easy to understand. Of the individuals who were asked to listen to an episode of the SANGYAN podcast, 99.8% (499/500) expressed their satisfaction with it as a platform to receive appropriate health information at the right time.

XSL•FO
RenderX

Figure 2. Satisfaction with and usability of the SANGYAN podcast based on results from the Client Satisfaction Questionnaire (CSQ) and System Usability Scale (SUS), respectively, compared between urban and rural residents.



Associations Between Sociodemographic Characteristics and SUS and CSQ-8 Scores

Table 5 shows the associations between the sociodemographic variables and SUS and CSQ-8 scores, with respect to region of residence. Above average system use was more prevalent in respondents aged 25 years to 35 years (74/250, 29.6%) than in those aged 18 years to 24 years (58/250, 23.2%), and the relationship with age groups was statistically significant (χ^2_4=38.7, *P*<.001). Above average ratings for system usability were more common among women (138/250, 55.2%) than men (112/250, 44.8%; χ^2_1=16.9, *P*<.001). In addition, statistically significant relationships between system usability ratings educational level were present (χ^2_6=64, *P*=.001), with a greater proportion (106/250, 42.4%) of those having a graduate degree rating it above average. In addition, most of the respondents in the different age groups were highly satisfied with the SANGYAN podcast. Differences in ratings by employment status were also present (χ^2_6=28.07, *P*<.001). There were no significant differences in SUS scores by religion (χ^2_3=6.69, *P*=.08).

XSL•FO
RenderX

Table . Differences in System Usability Scale (SUS) and Client Satisfaction Questionnaire - 8 (CSQ-8) scores by sociodemographic characteristics of the study respondents (N=500).

Variables	SUS						CSQ-8			
	Above average, n (%)		Below average, n (%)		χ^2 (*df*) [a]	*P* value	High satisfaction, n (%)		Low satisfaction, n (%)	
	Rural (n=250)	Urban (n=250)	Rural (n=250)	Urban (n=250)			Rural (n=250)	Urban (n=250)	Rural (n=250)	Urban (n=250)
Age group (years)					38.7 (4)	<.001				
18 - 24 (n=76)	20 (26.3)	38 (50)	6 (7.9)	12 (15.8)			26 (34.2)	50 (65.8)	0 (0)	0 (0)
25 - 35 (n=138)	37 (26.8)	37 (26.8)	39 (28.3)	25 (18.1)			76 (55.1)	62 (44.9)	0 (0)	0 (0)
36 - 44 (n=105)	14 (13.3)	41 (39)	35 (33.3)	15 (14.3)			49 (46.7)	56 (53.3)	0 (0)	0 (0)
45 - 64 (n=155)	27 (17.4)	27 (17.4)	60 (38.7)	41 (26.5)			87 (56.1)	68 (43.9)	0 (0)	0 (0)
≥65 (n=26)	4 (15.4)	5 (19.2)	8 (30.8)	9 (34.6)			11 (42.3)	14 (53.8)	1 (3.8)	0 (0)
Gender					16.9 (2)	<.001				
Female (n=289)	62 (21.5)	76 (26.3)	60 (20.8)	151 (52.2)			152 (52.6)	136 (47.1)	1 (0.3)	0 (0)
Male (n=211)	40 (19)	72 (34.1)	57 (27)	42 (19.9)			97 (46)	114 (54)	0 (0)	0 (0)
Educational level					64 (6)	<.001				
No formal education (n=59)	8 (13.6)	10 (16.9)	37 (62.7)	4 (6.8)			44 (74.6)	14 (23.7)	1 (1.7)	0 (0)
High school or less than high school certification (n=241)	50 (20.7)	42 (17.4)	98 (40.7)	51 (21.2)			148 (61.4)	83 (34.4)	0 (0)	0 (0)
Intermediate or diploma (n=32)	8 (25)	11 (34.4)	6 (18.8)	7 (21.9)			14 (43.8)	18 (56.3)	0 (0)	0 (0)
Graduate and higher (n=168)	36 (21.4)	85 (50.6)	7 (4.2)	40 (23.8)			43 (25.6)	125 (74.4)	0 (0)	0 (0)
Religion [b]					6.69 (3)	.08				
Christian (n=50)	14 (28)	17 (34)	7 (14)	12 (24)			21 (42)	29 (58)	0 (0)	0 (0)
Hindu (n=434)	85 (19.6)	123 (28.3)	138 (31.8)	88 (20.3)			222 (51.2)	211 (48.6)	1 (0.2)	0 (0)

[a] χ^2 analyses were conducted using the original category groupings, but the categories were merged in the table for clarity and readability.

[b] Muslim participants were not reported because religion was not significantly associated with SUS scores (*P*=.08) and the number of Muslim participants in the sample was relatively small (n=10).

Associations Between Sociodemographic Characteristics and Health Literacy (REALM-R)

Of the 155 (155/500, 31%) participants in this study who were aged from 45 years to 64 years, 87.7% (136/155) had poor health literacy (Table 6), representing the largest proportion even though the younger age groups (ie, 18 - 24 years and 25 - 35 years) had more limited health literacy regarding medical terminologies than the older population. There were significant associations between levels of health literacy and age groups (χ^2_4=25.4, *P*<.001). According to the evaluation using the REALM-R examiner record, 66.8% (167/250) of individuals from urban regions had poor health literacy, as did 86.8% (217/250) of those in rural areas. In addition, there was a significant association between health literacy (REALM-R) and regions of residence (χ^2_1=28.1, *P*<.001). However, the difference in the REALM-R score between genders was not statistically significant (χ^2_1=0.401, *P*=.53). The association between REALM-R grades and education status was statistically significant (χ^2_6=167, *P*<.001).

XSL•FO
RenderX

Table . Associations between the sociodemographic variables and Rapid Estimate of Adult Literacy in Medicine, Revised (REALM-R) grades.

Variables	REALM-R grade, n (%)		χ^2 (*df*)[a]	*P* value
	Adequate health literacy (n=116)	Poor health literacy (n=384)		
Age group (years)			25.4 (4)	<.001
18 - 24 (n=76)	27 (35.5)	49 (64.5)		
25 - 35 (n=138)	45 (32.6)	93 (67.4)		
36 - 44 (n=105)	21 (20)	84 (80)		
45 - 64 (n=155)	19 (12.3)	136 (87.7)		
≥65 (n=26)	4 (15.4)	22 (84.6)		
Gender			0.401 (1)	.53
Female (n=289)	70 (24.2)	219 (75.8)		
Male (n=211)	46 (21.8)	165 (78.2)		
Region of residence			28.1 (1)	<.001
Rural (n=250)	33 (13.2)	217 (86.8)		
Urban (n=250)	83 (33.2)	167 (66.8)		
Education level			167 (6)	<.001
No formal education (n=59)	0 (0)	59 (100)		
High school or less than high school certification (n=241)	15 (6.2)	226 (93.8)		
Graduate and above (n=168)	95 (56.5)	73 (43.5)		
Intermediate or diploma (n=32)	6 (18.8)	26 (81.3)		
Employment status			35.3 (6)	<.001
Employed (n=253)	78 (30.8)	175 (69.2)		
Homemaker (n=110)	14 (12.7)	96 (87.3)		
Unemployed (n=99)	10 (10.2)	89 (89.9)		
Other (n=39: retired, n=14; student, n=24; unempoyed due to COVID-19, n=38)	14 (36.8)	24 (63.2)		

[a]χ^2 analyses were conducted using the original category groupings, but the categories were merged in the table for clarity and readability.

Discussion

Principal Findings

In this study, 50% (250/500) of respondents were from rural areas, and 50% (250/500) were from urban areas. Interestingly, a large portion of the participants reported neither prior use (420/494, 85%) nor familiarity with podcasts (443/495, 89.5%) for gathering any kind of health-related information before this study, highlighting a considerable gap in awareness and exposure to digital learning tools across various demographics. The greatest proportion of participants was among the age group of 45 years to 64 years (155/500, 31%), followed closely by the age group of 25 years to 34 years (138/500, 27.6%). Comparatively more female participation (289/500, 57.8%) was observed, reflecting greater availability among the demographic segments. Additionally, 76.8% (384/500) of the population had a poor literacy rate, and 48.2% (241/500) of the participants had an education level of high school or less, while 11.8% (59/500) had no formal education. Our podcast effectiveness was assessed using the SUS, which revealed an overall good usability, with a mean score of 70.9. Participants aged 18 years to 24 years had the highest mean SUS score of 81.2, suggesting higher acceptance among younger users, who perceived the podcasts as more user-friendly and engaging. The CSQ evaluation showed that 94.8% of the users reported satisfaction with the quality of service received, 97.4% got the information they were seeking, 93% of the participants were likely to use SANGYAN again, and 97% would recommend SANGYAN to their friends and family, highlighting high user trust and community-oriented impact. Holistically, the SANGYAN app was well-received and perceived as a useful and credible source of health information. Similar findings were reported in recent studies that also demonstrated acceptability and effectiveness of mobile and app-based interventions for enhancing and improving health-related behaviors and user engagement among individuals with chronic conditions [41,42]. The exploratory subgroup analyses suggested that younger, female, and more educated participants found the platform more usable and satisfactory, highlighting the influence of demographic and literacy factors on digital health engagement.

XSL•FO
RenderX

In the twenty-first century, the SARS-CoV-2 pandemic affected millions of individuals globally. During this age of global communication, misleading information spreads rapidly, and this can potentially lead to unreliable sources of information overwhelming the general population [43]. A previous study in Ghana showed that podcast or voice messages are reliable sources of information against misinformation [44].

In our study, poor health literacy was observed in 76.8% (384/500) of the participants, among which 56.5% (217/384) were from rural regions and 43.5% (167/384) were from urban regions. A previous study in India also demonstrated poor health literacy, at a rate of 77% [45]. Previous studies have also reported poor health literacy in the following countries: Bulgaria (62%), Spain (58%), Suriname (95%), Saudi Arabia (54%), and South Africa (68%) [40,43-45]. In contrast, adequate health literacy was reported for the Netherlands (71%), Ireland (59%), Greece (55%), and Poland (56%). In Suriname, better health literacy was demonstrated in urban settings [46].

Further, health literacy was significantly associated with employment status, showing that unemployed people and retired people were more likely to demonstrate higher levels of poor literacy. Similarly, a previous study in Denmark found poor health literacy levels among unemployed people [47]. Health literacy also had a significant association with education levels; participants demonstrating adequate literacy were likely graduates or professionals. Similar to this, a study in Jordan showed that participants demonstrating adequate literacy were likely highly educated [48].

SANGYAN was considered a source of relevant health information by 98% of the participants, and 99.8% of the participants were satisfied with the usage of the SANGYAN podcast. Similarly, previous studies in India and the United States found that podcasts were relevant and provided the necessary information [25,26]. Studies in the United States also found that its users were satisfied with the use of podcasts [49,50].

This study showed a significant association between health literacy and SUS scores, such that a person demonstrating poor health literacy was likely to have a below average SUS score. In contrast, a previous study in the United States reported no significant association between health literacy and SUS scores for digital health–based platforms [50]. Our study showed that participants in rural settings were likely to score below average, whereas in a previous study conducted in rural settings of the United States, participants scored an average of 81 points, which indicates above average usability [51].

The participants from rural settings found the SANGYAN podcast to be a satisfactory source of health information. Similarly, a study in rural Senegal recorded satisfactory responses for voice message–based platforms [27]. The poor health literacy demonstrated among participants and their satisfactory response to the SANGYAN podcast emphasize the use of podcasts to deliver health information to participants who experience difficulty reading. Based on previous studies, podcasts are practical, easy to use, and educational [48,52].

Further, in contrast to existing literature from studies conducted in India, Uganda, and Malaysia that showed podcasts' potential to be used as educational tools predominantly with students [53-57], this study showed that podcasts are a viable tool for information delivery across all age groups. It demonstrated that acceptance of and satisfaction with podcast use are not limited to any particular age group.

Existing studies of methods to tackle infodemics used the traditional approach of debunking myths. Although this is effective, it predominantly uses graphics. As a tool against infodemics, SANGYAN is not only evidence-based but also available in various languages. Because it is delivered in an audio format, it caters to the population with poor literacy levels [58,59].

A study reviewed various digital tools used to curb misinformation through mobile apps and official government and mass media websites [60]. Participants in another study indicated that the provision of health-related information in languages other than English would have been used [61]. With podcasting having a large audience in India, this can be one digital tool to combat misinformation. Use of podcasts for educational purposes in the global and Indian contexts has increased, making SANGYAN an ideal platform [62-67]. The results of this study demonstrated high participant satisfaction and acceptance, suggesting that SANGYAN presents an ideal multilingual approach. With the results of this study showing high satisfaction and acceptance from participants, SANGYAN could be an ideal multilingual tool to combat the rapid spread of misinformation and disinformation. Further, it helps connect communities to evidence-based information that enhances the good health and well-being of individuals, families, and the communities in which they live.

The strength of this study is that it used a digital platform as a medium to impart information, and it shared information in the form of audio rather than text, catering to a wider range of people. The study also provided podcasts through existing kiosks in rural and urban health centers, making it accessible to people without the internet or a smartphone. The information the podcast provides is evidence-based and from a reliable source, hence enabling it to tackle misinformation. Before starting the data gathering procedure, the data collectors underwent a training session. The findings of the study add to the global evidence that well-designed, user-centric, multilingual tools can improve public health engagement and enhance self-caring practices in limited-resource settings [41,42,68].

Limitations

Our study had a few limitations. First, since we used convenience sampling, participants were recruited based on availability rather than a random selection method. Moreover, the study used a single-arm design and therefore lacked a control group to compare the findings. The research was limited to the Tamil Nadu area, limiting the external validity and generalizability of the findings to a broader population.

Last, the study did not evaluate the long-term retention of knowledge or sustained behavioral impact delivered through the podcasts. Future research studies in this domain should

XSL•FO
RenderX

bridge these gaps in knowledge to strengthen the scientific evidence. This would be beneficial for observing the impact of a podcast-based digital health platform on the issue of misinformation and disinformation.

Conclusions

This study established the acceptability of and satisfaction with the SANGYAN podcast platform among rural and urban areas. The findings emphasize the significance of creating an adaptable platform that enables people with low health literacy to connect to topics relevant to their health, by listening to clear, dependable, and brief podcast-delivered information. SANGYAN podcast could be an important source of information that can be used to more effectively educate people. Additionally, it can be used to reduce misinformation and help with the self-management of COVID-19, communicable diseases, and NCDs. This would also help to enhance the acceptability of podcasts as a source of health-related information.

Authors' Contributions

Conceptualization: AJ
Data curation: AB, DO, SKR
Formal analysis: AB
Methodology: AB and AJ
Project administration: AB, SKM
Resources: AB, DO, GM, SKR
Software: AB
Supervision: AJ
Validation: AB, DO, GM, SKR, SKM
Visualization: AB, AJ
Writing - original draft: AB, AJ
Writing - review and editing: AB, AG, AJ

Conflicts of Interest

The authors were part of the team involved in the development and heuristic evaluation of the SANGYAN platform, primarily through the Foundation of Healthcare Technologies Society. No financial nor personal conflicts of interest are declared.

Multimedia Appendix 1
Podcast image showing topics and languages.
[PNG File, 85 KB - infodemiology_v6i1e67504_app1.png]

Multimedia Appendix 2
Podcast image showing episodes and filters.
[PNG File, 236 KB - infodemiology_v6i1e67504_app2.png]

Multimedia Appendix 3
SANGYAN tool.
[DOCX File, 29 KB - infodemiology_v6i1e67504_app3.docx]

References

1. Weekly epidemiological update on COVID-19 - 8 February 2022, edition 78. World Health Organization. 2022. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---8-february-2022 [accessed 2026-02-15]
2. Verhagen LM, de Groot R, Lawrence CA, Taljaard J, Cotton MF, Rabie H. COVID-19 response in low- and middle-income countries: don't overlook the role of mobile phone communication. Int J Infect Dis 2020 Oct;99:334-337. [doi: 10.1016/j.ijid.2020.07.069] [Medline: 32763447]
3. van der Linden S, Roozenbeek J, Compton J. Inoculating against fake news about COVID-19. Front Psychol 2020;11:566790. [doi: 10.3389/fpsyg.2020.566790] [Medline: 33192844]
4. Merchant RM, Lurie N. Social media and emergency preparedness in response to novel coronavirus. JAMA 2020 May 26;323(20):2011-2012. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]
5. Clement J. Number of global social media users 2010-2021. Statista. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2019-08-16]
6. Paules CI, Marston HD, Fauci AS. Coronavirus infections-more than just the common cold. JAMA 2020 Feb 25;323(8):707-708. [doi: 10.1001/jama.2020.0757] [Medline: 31971553]

XSL•FO
RenderX

7. Rastogi A, Syed S, Sharma T, et al. Enhancing the health coverage in India by empowering the corona warriors through educational intervention. Int J Community Med Public Health 2021;8(7):3305. [doi: 10.18203/2394-6040.ijcmph20212325]
8. R J, D B, waran K. Social media reigned by information or misinformation about COVID-19: a phenomenological study. SSRN Journal 2020. [doi: 10.2139/ssrn.3596058]
9. Evanega S, Lynas M, Adams J, Smolenyak K, Insights CG. Coronavirus misinformation: quantifying sources and themes in the COVID-19 'infodemic' (preprint). Journal of Medical Internet Research. Preprint posted online on 2020. [doi: 10.2196/preprints.25143]
10. Lau LL, Hung N, Go DJ, et al. Knowledge, attitudes and practices of COVID-19 among income-poor households in the Philippines: a cross-sectional study. J Glob Health 2020 Jun;10(1):011007. [doi: 10.7189/jogh.10.011007] [Medline: 32566169]
11. World Health Organization. Protecting yourself and others from the spread COVID-19. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2022-09-20]
12. Islam MS, Sarkar T, Khan SH, et al. COVID-19-related infodemic and its impact on public health: a global social media analysis. Am J Trop Med Hyg 2020 Oct;103(4):1621-1629. [doi: 10.4269/ajtmh.20-0812] [Medline: 32783794]
13. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552. [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
14. Bin Naeem S, Kamel Boulos MN. COVID-19 misinformation online and health literacy: a brief overview. Int J Environ Res Public Health 2021 Jul 30;18(15):8091. [doi: 10.3390/ijerph18158091] [Medline: 34360384]
15. Dhingra D, Dabas A. Global strategy on digital health. Indian Pediatr 2020 Apr 15;57(4):356-358. [doi: 10.1007/s13312-020-1789-7] [Medline: 32284477]
16. Dadaczynski K, Okan O, Messer M, et al. Digital health literacy and web-based information-seeking behaviors of university students in Germany during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2021 Jan 15;23(1):e24097. [doi: 10.2196/24097] [Medline: 33395396]
17. Benedict L, Umakanthan B, Thane G, Wang W, Alhalbouni S. Podcasting as a tool for health communication? The public health insight podcast and emergent opportunities. Global Health: Annual Review 2021(6) [FREE Full text]
18. Balls-Berry J, Sinicrope P, Valdez Soto M, Brockman T, Bock M, Patten C. Linking podcasts with social media to promote community health and medical research: feasibility study. JMIR Form Res 2018 Oct 24;2(2):e10025. [doi: 10.2196/10025] [Medline: 30684430]
19. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libraries J 2007 Mar;24(1):2-23 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00701.x]
20. Ifedayo AE, Ziden AA, Ismail AB. Podcast acceptance for pedagogy: the levels and significant influences. Heliyon 2021 Mar;7(3):e06442. [doi: 10.1016/j.heliyon.2021.e06442] [Medline: 33748494]
21. Eveland WP, Dunwoody S. User control and structural isomorphism or disorientation and cognitive load? Communic Res 2001 Feb;28(1):48-78. [doi: 10.1177/009365001028001002]
22. Newman J, Liew A, Bowles J, Soady K, Inglis S. Podcasts for the delivery of medical education and remote learning. J Med Internet Res 2021 Aug 27;23(8):e29168. [doi: 10.2196/29168] [Medline: 34448719]
23. Jham BC, Duraes GV, Strassler HE, Sensi LG. Joining the podcast revolution. J Dent Educ 2008 Mar;72(3):278-281. [doi: 10.1002/j.0022-0337.2008.72.3.tb04493.x] [Medline: 18316531]
24. Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. J Public Health Policy 2020 Dec;41(4):410-420. [doi: 10.1057/s41271-020-00247-w] [Medline: 32826935]
25. Militello L, Sezgin E, Huang Y, Lin S. Delivering perinatal health information via a voice interactive app (SMILE): mixed methods feasibility study. JMIR Form Res 2021 Mar 1;5(3):e18240. [doi: 10.2196/18240] [Medline: 33646136]
26. Murthy N, Chandrasekharan S, Prakash MP, et al. Effects of an mHealth voice message service (mMitra) on maternal health knowledge and practices of low-income women in India: findings from a pseudo-randomized controlled trial. BMC Public Health 2020 Jun 1;20(1):820. [doi: 10.1186/s12889-020-08965-2] [Medline: 32487065]
27. Downs SM, Sackey J, Kalaj J, Smith S, Fanzo J. An mHealth voice messaging intervention to improve infant and young child feeding practices in Senegal. Matern Child Nutr 2019 Oct;15(4):e12825. [doi: 10.1111/mcn.12825] [Medline: 30950190]
28. Sangyan. Foundation of Healthcare Technologies Society. URL: https://fhts.ac.in/sangyan/ [accessed 2022-08-18]
29. Connor M, Norman P. Predicting Health Behaviour: Open University Press; 2005.
30. ScienceDirect. Social cognitive theory. ScienceDirect. URL: https://www.sciencedirect.com/topics/social-sciences/social-cognitive-theory [accessed 2025-05-27]
31. Ab Rahim NR, Darwis N, Noh N, Asfar J, Ismail HB. Exploring learning strategies through the social cognitive theory. International Journal of Research and Innovation in Social Science 2025;18. [doi: 10.47772/IJRISS.2025.9010276]
32. Khan A, Khan S, Zia-Ul-Islam S, Khan M. Communication skills of a teacher and its role in the development of the students' academic success. Journal of Education and Practice 2017;8(1):18-21 [FREE Full text]
33. Bouchrika I. What is information processing theory? Stages, models & limitations for 2026. Research.com. 2025. URL: https://research.com/education/what-is-information-processing-theory [accessed 2025-05-22]

XSL•FO
RenderX

34. Swanson HL. Information processing theory and learning disabilities: an overview. J Learn Disabil 1987 Jan;20(1):3-7. [doi: 10.1177/002221948702000102] [Medline: 3805889]
35. Attkisson CC, Greenfield TK. Client Satisfaction Questionnaire-8 and Service Satisfaction Scale-30. In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcome Assessment: Lawrence Erlbaum Associates, Inc; 1994:402-420 URL: https://psycnet.apa.org/record/1994-97261-016
36. Lewis JR. The System Usability Scale: past, present, and future. International Journal of Human–Computer Interaction 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
37. Dumenci L, Matsuyama RK, Kuhn L, Perera RA, Siminoff LA. On the validity of the Rapid Estimate of Adult Literacy in Medicine (REALM) scale as a measure of health literacy. Commun Methods Meas 2013 Jun 18;7(2):134-143. [doi: 10.1080/19312458.2013.789839] [Medline: 23935770]
38. Assessment tools. Adult Meducation. URL: http://adultmeducation.com/AssessmentTools.html [accessed 2022-08-31]
39. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
40. Sauro J. A Practical Guide to Measuring Usability.: MeasuringU; 2010. URL: https://measuringu.com/wp-content/uploads/2017/05/QuantitativeUsabilityTestOnline.pdf [accessed 2026-02-15]
41. Chen D, Zhang H, Wu J, et al. Effects of an individualized mHealth-based intervention on health behavior change and cardiovascular risk among people with metabolic syndrome based on the behavior change wheel: quasi-experimental study. J Med Internet Res 2023 Nov 29;25:e49257. [doi: 10.2196/49257] [Medline: 38019579]
42. Kerr D, Ahn D, Waki K, Wang J, Breznen B, Klonoff DC. Digital interventions for self-management of type 2 diabetes mellitus: systematic literature review and meta-analysis. J Med Internet Res 2024 Jul 22;26:e55757. [doi: 10.2196/55757] [Medline: 39037772]
43. Nelson T, Kagan N, Critchlow C, Hillard A, Hsu A. The danger of misinformation in the COVID-19 crisis. Mo Med 2020;117(6):510-512. [Medline: 33311767]
44. Willcox M, Moorthy A, Mohan D, et al. Mobile technology for community health in Ghana: is maternal messaging and provider use of technology cost-effective in improving maternal and child health outcomes at scale? J Med Internet Res 2019 Feb 13;21(2):e11268. [doi: 10.2196/11268] [Medline: 30758296]
45. U P R, Belman M, Kamath A, B U, Shenoy K A, A L U. Evaluation of health literacy status among patients in a tertiary care hospital in coastal karnataka, India. J Clin Diagn Res 2013 Nov;7(11):2551-2554. [doi: 10.7860/JCDR/2013/6120.3608] [Medline: 24392398]
46. Diemer FS, Haan YC, Nannan Panday RV, van Montfrans GA, Oehlers GP, Brewster LM. Health literacy in Suriname. Soc Work Health Care 2017 Apr;56(4):283-293. [doi: 10.1080/00981389.2016.1277823] [Medline: 28271964]
47. Abdel-Latif MMM, Saad SY. Health literacy among Saudi population: a cross-sectional study. Health Promot Int 2019 Feb 1;34(1):60-70. [doi: 10.1093/heapro/dax043] [Medline: 28973389]
48. van Rensburg ZJ. Levels of health literacy and English comprehension in patients presenting to South African primary healthcare facilities. Afr J Prim Health Care Fam Med 2020 Jan 30;12(1):e1-e6. [doi: 10.4102/phcfm.v12i1.2047]
49. Sørensen K, Pelikan JM, Röthlin F, et al. Health literacy in Europe: comparative results of the European Health Literacy Survey (HLS-EU). Eur J Public Health 2015 Dec;25(6):1053-1058. [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
50. Svendsen IW, Damgaard MB, Bak CK, et al. Employment status and health literacy in Denmark: a population-based study. Int J Public Health 2021;66:598083. [doi: 10.3389/ijph.2021.598083] [Medline: 34744563]
51. Nobis S, Lehr D, Ebert DD, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. Diabetes Care 2015 May;38(5):776-783. [doi: 10.2337/dc14-1728] [Medline: 25710923]
52. Martinez W, Hackstadt AJ, Hickson GB, et al. The My Diabetes Care patient portal intervention: usability and pre-post assessment. Appl Clin Inform 2021 May;12(3):539-550. [doi: 10.1055/s-0041-1730324] [Medline: 34192774]
53. Wolpaw J, Toy S. Creation and evaluation of an anesthesiology and critical care podcast. J Educ Perioper Med 2018;20(1):E620. [doi: 10.46374/volxx-issue1-wolpaw] [Medline: 29928667]
54. Thirumalai M, Brown N, Niranjan S, et al. An interactive voice response system to increase physical activity and prevent cancer in the rural Alabama black belt: design and usability study. JMIR Hum Factors 2022 Jan 4;9(1):e29494. [doi: 10.2196/29494] [Medline: 34982714]
55. Fitzpatrick AL, van Pelt M, Heang H, et al. Using targeted mHealth messages to address hypertension and diabetes self-management in Cambodia: protocol for a clustered randomized controlled trial. JMIR Res Protoc 2019 Mar 19;8(3):e11614. [doi: 10.2196/11614] [Medline: 30888330]
56. Kaahwa M, Zhu C, Muhumuza M. Investigating ICT skills and the use of audio media in distance education among teachers and students: the case of Mountains of the Moon University in Uganda. AF 2018;31(1). [doi: 10.21825/af.v31i1.9035]
57. Kaahwa M, Zhu C, Muhumuza M, Karemera C. Differences between audio media and conventional methods regarding students' academic performance and the influence of audio media satisfaction on their academic scores at a Ugandan university. IJIL 2021;29(4):491. [doi: 10.1504/IJIL.2021.115500]

XSL•FO
RenderX

58. Mugahed W, Othman MS, Al-Rahmi WM. Evaluating student's satisfaction of using social media through collaborative learning in higher education. International Journal of Advances in Engineering & Technology 2013;6(4):1541-1551 [FREE Full text]
59. Dajani FK. Examining social presence influence on students' satisfaction with online learning environments. In: Scheg AG, editor. Critical Examinations of Distance Education Transformation across Disciplines: IGI Global; 2015:211-236. [doi: 10.4018/978-1-4666-6555-2.ch011]
60. Khoramian Tusi S, Sheikh Fathollahi M, Rahnamaye Tamrooyee F, Akbari Javar M. Study of the effect of podcasting on learning and satisfaction in dental students. Journal of Mashhad Dental School 2015;39(3):229-238 [FREE Full text]
61. van der Linden S. Misinformation: susceptibility, spread, and interventions to immunize the public. Nat Med 2022 Mar;28(3):460-467. [doi: 10.1038/s41591-022-01713-6] [Medline: 35273402]
62. Challenger A, Sumner P, Bott L. COVID-19 myth-busting: an experimental study. BMC Public Health 2022 Jan 19;22(1):131. [doi: 10.1186/s12889-021-12464-3] [Medline: 35045852]
63. Alonso-Galbán P, Alemañy-Castilla C. Curbing misinformation and disinformation in the COVID-19 era: a view from Cuba. MEDICC Rev 2020;22(2) [FREE Full text] [doi: 10.37757/MR2020.V22.N2.12]
64. Lockyer B, Islam S, Rahman A, et al. Understanding COVID-19 misinformation and vaccine hesitancy in context: findings from a qualitative study involving citizens in Bradford, UK. Health Expect 2021 Aug;24(4):1158-1167. [doi: 10.1111/hex.13240] [Medline: 33942948]
65. Cho D, Cosimini M, Espinoza J. Podcasting in medical education: a review of the literature. Korean J Med Educ 2017 Dec;29(4):229-239. [doi: 10.3946/kjme.2017.69] [Medline: 29207454]
66. Lee C, Zhou MS, Wang ER, Huber M, Lockwood KK, Parga J. Health care professional and caregiver attitudes toward and usage of medical podcasting: questionnaire study. JMIR Pediatr Parent 2022 Feb 1;5(1):e29857. [doi: 10.2196/29857] [Medline: 35103616]
67. MacKenzie LE. Science podcasts: analysis of global production and output from 2004 to 2018. R Soc Open Sci 2019;6(1):180932. [doi: 10.1098/rsos.180932] [Medline: 30800351]
68. Lee K, Chung Y, Kim JS. Research trends on metabolic syndrome in digital health care using topic modeling: systematic search of abstracts. J Med Internet Res 2024 Dec 12;26(26):e53873. [doi: 10.2196/53873] [Medline: 39666378]

Abbreviations

CSQ-8: Client Satisfaction Questionnaire - 8
IDEA: interactive, decision-making, educational audio
LMIC: low- and middle-income countries
NCD: noncommunicable disease
PMCHRI: Panimalar Medical College Hospital & Research Institute
REALM-R: Rapid Estimate of Adult Literacy in Medicine, Revised
RHTC: Rural Health Training Center
SCT: social cognitive theory
SUS: System Usability Scale
WHO: World Health Organization

Edited by T Mackey; submitted 13.Oct.2024; peer-reviewed by M Haupt, W Ahmed; revised version received 03.Nov.2025; accepted 23.Nov.2025; published 07.May.2026.

Please cite as:
Joshi A, Bhatt A, Mohan SK, Malhotra G, Oliveira D, Revathi SK, Grover A
Implementation and User Evaluation of the SANGYAN Digital Health Platform to Enhance Knowledge About COVID-19 and Other Health Conditions: Quasi-Experimental Study
JMIR Infodemiology 2026;6:e67504
URL: https://infodemiology.jmir.org/2026/1/e67504
doi:10.2196/67504

© Ashish Joshi, Ashruti Bhatt, Surapaneni Krishna Mohan, Gunjan Malhotra, Doilyn Oliveira, Saravanavel Kalpana Revathi, Ashoo Grover. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 7.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Analyzing Misinformation and Disinformation: Understanding Swiss COVID-19 Narratives Through Natural Language Processing Analysis

Federico Germani[1*], PhD; Giovanni Spitale[1*], PhD; Franc Fritschi[2], MA; Sonja Merten[2], MPH, MD, PhD, PD; Nikola Biller-Andorno[1], MA, MHBA, MD, PhD

[1]Institute of Biomedical Ethics and History of Medicine, University of Zurich, Winterthurerstrasse 30, Zürich, Switzerland

[2]Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland

[*]these authors contributed equally

Corresponding Author:
Nikola Biller-Andorno, MA, MHBA, MD, PhD
Institute of Biomedical Ethics and History of Medicine, University of Zurich, Winterthurerstrasse 30, Zürich, Switzerland

Abstract

Background: The COVID-19 pandemic has highlighted the challenges posed by the rapid spread of misinformation and disinformation, exacerbating societal polarization and institutional distrust. Understanding how misinformation and disinformation is understood and framed in public discourse is essential to developing strategies for building societal resilience and promoting informed decision-making during crises.

Objective: This study explores the use of the terms misinformation and disinformation across Swiss public discourse during the COVID-19 pandemic, examining their framing within newspaper articles and social media interactions. The findings aim to inform policymakers and journalists or communicators on mitigating the societal impact of misinformation and disinformation through the promotion of a common understanding of the terms misinformation and disinformation.

Methods: We analyzed 2 datasets using a natural language processing pipeline, including lemmatization, co-occurrence analysis, and semantic network mapping: media articles retrieved via Factiva and social media posts collected via CrowdTangle.

Results: The framing of misinformation and disinformation varied significantly across the datasets. News media highlighted its role in shaping public sentiment, often discussing the tension between journalistic integrity and the amplification of falsehoods. Social media exhibited polarized narratives, with discussions centered on conspiracy theories, distrust in institutions, and grassroots mobilization.

Conclusions: Diverging narratives on the very concepts of misinformation and disinformation across public discourse reflect broader societal tensions. Robust journalistic integrity in the media and resilience strategies against misinformation and disinformation involving empowering publics through information literacy approaches are critical to bridging divides and reducing polarization.

(JMIR Infodemiology 2026;6:e76441) doi:10.2196/76441

KEYWORDS

COVID-19; misinformation; disinformation; public discourse; natural language processing; Switzerland

Introduction

The COVID-19 pandemic has not only presented unprecedented public health challenges but has also exposed the susceptibility of public discourse to the rapid dissemination of misinformation and disinformation [1]. This phenomenon—which includes intentional falsehoods and unintentional inaccuracies [2]—has amplified societal divisions [3], diminished trust in institutions [4], and complicated efforts to manage the crisis effectively [5]. Misinformation refers to inaccuracies shared unknowingly by people, thinking the information is accurate, whereas disinformation refers to falsehoods disseminated deliberately to deceive with intent to harm, usually for political gains or economic reasons [6]. Despite being distinct concepts, here, we discuss both, as the focus of this work is on their framing and impact on public discourse, and because this distinction might not always be clear to lay publics using the terms. Misinformation and disinformation have been implicated in undermining vaccine uptake [7], conspiracy theories [8], and eroding confidence in evidence-based measures [9].

As a fundamental element of societal resilience, a transparent and reasoned public discourse is essential during crises [10,11]—for instance, in Italy, public communication during the COVID-19 pandemic combined institutional messages with

XSL•FO
RenderX

emotionally resonant media campaigns, highlighting collective responsibility and solidarity [12]. However, the fragility of such discourse is not only due to the circulating misinformation and disinformation but also due to the contentious and often polarized use of the terms "misinformation" and "disinformation" [13,14]. These terms have frequently been used in emotionally charged and politicized contexts, becoming subjects of debate themselves, rather than tools for clarity and common understanding [13-15]. Scholarly debates have abundantly explored how misinformation has been shaped and propagated during the COVID-19 pandemic [4,16,17], yet the specific framing of misinformation and disinformation in public discourse—whether in media narratives or on social media platforms—remains underexplored, especially in the Swiss context.

This paper contributes to addressing this gap by analyzing how misinformation and disinformation have been framed—referring to the process of selecting and emphasizing certain aspects of reality to promote specific interpretations, evaluations, or solutions to an issue [18]—in Swiss public discourse during the COVID-19 pandemic. Given the different languages and cultures coexisting in Switzerland, the country's unique sociocultural and political system presents a unique case study for examining the interplay between national narratives and regional dynamics. Our findings aim to inform strategies for promoting societal resilience and a shared understanding of the concepts of misinformation and disinformation across diverse audiences. This study is part of the broader research project titled "Boosting Public Discourse: A Targeted Evidence-Based Strategy to Improve Moral Reasoning" [19], which examines the framing of key moral terms, such as *freedom* [20], in Switzerland during the COVID-19 pandemic. Here, we focus on misinformation and disinformation, analyzing its conceptualizations in Swiss public discourse as reflected in news media and social media during the COVID-19 pandemic.

Methods

Ethical Considerations

This study analyzed secondary data consisting exclusively of publicly available news articles and publicly accessible social media posts from Facebook pages and groups. No human participants were recruited or contacted, and no identifiable private or sensitive personal data were collected. The analysis was conducted on aggregated textual data, and no attempt was made to identify individual users. According to the Swiss Federal Act on Research involving Human Beings (Human Research Act), research that does not involve human participants, identifiable health-related data, or biological material falls outside the scope of the Act and does not require review by a cantonal ethics committee [21]. As this study relied solely on publicly accessible, nonidentifiable data, formal ethics approval was not required. All procedures complied with applicable data protection regulations and institutional standards for responsible research.

Data Collection

The data used in this study include 2 comprehensive datasets, each curated to capture a wide range of public discourses surrounding the COVID-19 pandemic. The first dataset (news articles) consists of 209 articles sourced from the Dow Jones Factiva database [22], covering the period from September 2019 to April 2023. The query was specifically designed to extract articles discussing COVID-19, misinformation, and disinformation, focusing on media reporting on these topics in relation to Switzerland. It is important to note that the corpus does not consist solely of articles published by Swiss media. Due to the query logic in Factiva, which does not allow for country-level granularity in source definition, the dataset also includes articles about Switzerland that were published by foreign media outlets. Moreover, recognizing that Swiss audiences regularly consume foreign media, the query was designed to capture a diverse range of journalistic narratives. Consequently, the dataset encompasses news articles in German, French, Italian, and English that discuss Switzerland and issues related to misinformation and disinformation, thereby providing a more comprehensive insight into how these narratives were framed and disseminated in public discourse. The full details on the data collection, including the formulation of the different queries, are available in this study's OSF repository [23]. The second dataset (social media posts) comprises 580 posts collected via CrowdTangle [24], focusing on public interactions on Facebook pages and groups. These posts, gathered between September 2019 and April 2023, were drawn from Swiss Facebook pages and groups where discussions of COVID-19–related topics occurred. The query targeted the keywords "misinformation" (in German: *Fehlinformation*) and "disinformation" (in German: *Desinformation*), capturing a spectrum of user-generated content, including opinions, debates, and reactions to public health measures. Here, geographical filtering was performed using CrowdTangle's "local relevance" parameter, which limits the query to "posts from Pages or public groups that are predominantly local to that area" [25]. This analysis focuses primarily on German-language content. Further details on the data collection methodology and query formulation are available in the study's OSF repository [23].

Analysis

The analytical process in this study leverages natural language processing (NLP) tools and semantic network maps to extract meaningful insights from the data. The first step involved parsing and preprocessing the datasets to prepare them for analysis. For news articles, metadata, such as publication dates, sources, and article lengths, were extracted. As social media posts included some noise (posts not mentioning any of our COVID-19 descriptors), we filtered the corpus, removing posts not containing the following words: "covid," "corona," "virus," "covid-19," "coronavirus," "pandemic," "epidemic," "outbreak," "pandémie," "épidémie," "Pandemie," "Epidemie," "Seuche," "pandemia," and "epidemia." Lemmatization was carried out using the NLP library spaCy to reduce words to their base forms [26]. This step enabled uniform analysis across variations of the same word (eg, "misinform," "misinformation"). To capture and describe the semantic space of lemmas, we applied co-occurrence analysis, a method that examines the frequency and patterns of lemmas appearing together within a defined textual context—in our case, sentences. This approach allows us to identify semantic relationships between terms, uncovering

XSL•FO
RenderX

underlying structures and associations in the data. Semantic network maps were generated using Gephi, an open-source software for network data analysis [27]. These maps provided a high-level perspective on the terms most frequently co-occurring with "misinformation" and "disinformation" across the datasets. In these maps, nodes represent terms, with their size proportional to frequency, while edges indicate the strength of co-occurrence between them. The topology of the maps, shaped by the Circle Pack, Noverlap, and Label Adjust layouts, reflects the organization of terms within thematic clusters, which were determined by modularity class [28]. Color-coding based on modularity class further emphasizes these clusters, providing a clear and intuitive visual representation of the relationships and structure within the discourse. The modularity class analysis was conducted with a resolution parameter set to 0.5, which determines the granularity of the clusters detected [29]. A resolution of 0.5 was chosen empirically and informed by prior research [30] to ensure meaningful thematic groupings. In the analysis, we considered the largest and most prominent thematic clusters of each dataset (see the Results section).

To complement the quantitative analysis, we manually selected narrative examples to contextualize the findings. These examples illustrate specific uses of the key terms "misinformation" and "disinformation" within distinct contexts. This approach adds depth by showcasing how these terms are used and perceived in real-world scenarios. NLP techniques can efficiently process large volumes of text data (as in the case of our corpora), enabling us to systematically identify recurring patterns and lemmas that might be overlooked in manual analyses [31-33]. By integrating NLP analyses with narrative examples, we combine the computational efficiency of automated analysis with the context provided by individual narratives. This approach enables a more nuanced understanding of phenomena and enhances the understandability and richness of research findings. Detailed methodological procedures and aspects are available in this study's OSF repository [23].

Results

News Media (Factiva Dataset)

The entire retrieved dataset included 209 articles. From the analysis of the news dataset obtained via Factiva, the NLP analysis visualized through the co-occurrence analysis plot of the terms most commonly associated with "misinformation" and "disinformation" (Figure 1) and the semantic network map of the terms co-occurring in association with "misinformation" and "disinformation" (Figure 2) reveals 5 relevant themes.

The first theme emerging from the co-occurrence analysis (Figure 1) and from the analysis of the semantic network map (Figure 2) is centered on addressing disinformation through critical engagement, education, and political discourse. "Verschwörungstheorie" (conspiracy theory) and "verbreiten" (to spread), both present also among the most co-occurring terms with "misinformation" and "disinformation" (with either "*Desinformation*" or "*Fehlinformation*") (Figure 1), as well as the term "Mythos" (myth), point to narratives about the spread of false information and its role in fueling conspiracy theories, whereas "Bildung" (education), "entkräften" (to refute), "dekonstruieren" (to deconstruct), "proaktiv" (proactive), and "sachlich" (objective) indicate news media discussions about how to equip individuals with the tools to identify and counter misinformation and disinformation. These terms suggest a characterization of what traditional media consider misinformation and disinformation, complemented by a focus on promoting critical thinking and evidence-based discourse. In addition, the terms "klar" (clear) and "deutlich" (explicit), also among the most co-occurring terms with "misinformation" and "disinformation" (Figure 1), as well as the lemma "Sprache" (language), possibly reflect the importance of clear and transparent communication to provide information and counteract misinformation and disinformation effectively. Examples of articles' excerpts discussing these topics include the following:

> *Die teils abstrusen Vorstellungen können reale Folgen haben. "Wir haben in der Forschung gesehen, dass Verschwörungstheorien zu einem stärkeren Misstrauen in Politik und in andere Menschen führten," sagt Pummerer. "Verschwörungsanhänger haben auch weniger Vertrauen in andere Menschen und halten sich seltener an soziale Normen. Das hat Konsequenzen für den gesellschaftlichen Zusammenhalt." / These sometimes abstruse ideas can have real consequences. "In our research, we have seen that conspiracy theories lead to greater mistrust in politics and in other people," says Pummerer. "Conspiracy theorists also have less trust in other people and are less likely to adhere to social norms. This has consequences for social cohesion."* [Row 14 Factiva dataset, November 19, 2021]

> *[…] Grundsätzlich kommt das Forschungszentrum zu dem Schluss, dass professionelle Medien durch die Pandemie nochmals an Bedeutung gewonnen haben, indem sie zuverlässige Orientierung bieten und die Desinformation limitieren. Eine «ungünstige Entwicklung» sei vor diesem Hintergrund, dass sich die ökonomische Krise des Journalismus nochmals akzentuiert habe, schreiben die Forscher. Während die Werbeeinnahmen im Print schon länger zurückgehen, waren sie erstmals auch im Online-Werbemarkt rückläufig. / The research center concludes that professional media have gained further importance during the pandemic by providing reliable information and limiting disinformation. Against this backdrop, the economic crisis in journalism has been further exacerbated, the researchers write. While print advertising revenues have been declining for some time, they have now also fallen for the first time in the online advertising market.* [Row 48 Factiva dataset, October 26, 2021]

Terms including "politisch" (political), among the most co-occurring terms with "misinformation" and "disinformation" (Figure 1), as well as the terms "Grünen-Politikerin" (Green Party politician) and "Landeszentrale" (state center), linked to the terms "misinformation" and "disinformation", might suggest discussions about how political actors and institutions play a key role in shaping (the response to) misinformation and

XSL•FO
RenderX

disinformation. Further, terms such as “Information” (relevant as one of the most common co-occurring terms with “misinformation” and “disinformation”), “Verfügung” (available), and “stellen” (to provide) within this cluster might highlight the need for readily accessible, reliable information, again, as a countermeasure to misinformation and disinformation.

> *Der immer wieder erhobene Vorwurf, Medien würden im Zusammenhang mit der Corona-Pandemie «Panikmache» betreiben, wird durch die Studie entkräftet. Während der ersten Welle, als die Situation neu und vieles noch unbekannt war, berichteten Journalistinnen und Journalisten in 16 Prozent der untersuchten Artikel alarmistisch über das Virus. Während der zweiten Welle traf das gemäss der Inhaltsanalyse des FÖG nur noch auf 6 Prozent der Artikel zu. Das heisst also: Ein Grossteil der fast 18’700 untersuchten Beiträge aus 60 Schweizer Medien war nicht alarmistisch.*
>
> *The study refutes the repeatedly made accusation that the media engaged in ‘scaremongering’ in connection with the coronavirus pandemic. During the first wave, when the situation was new and much was still unknown, journalists reported alarmistically on the virus in 16 percent of the articles examined. During the second wave, according to the FÖG content analysis, this only applied to 6 percent of the articles. This means that a large proportion of the almost 18,700 articles examined from 60 Swiss media outlets were not alarmist.* [Row 30 Factiva dataset, October 25, 2021]

In general, this cluster highlights the interplay between misinformation and disinformation, political engagement, and education. It suggests that media portray the challenges of misinformation and disinformation as requiring action through educational initiatives and clear and transparent communication from institutions (and from the news media themselves). The terms point to a narrative in which facing the challenges posed by misinformation and disinformation involves empowering individuals through education and ensuring political, institutional, and journalistic accountability in providing factual information.

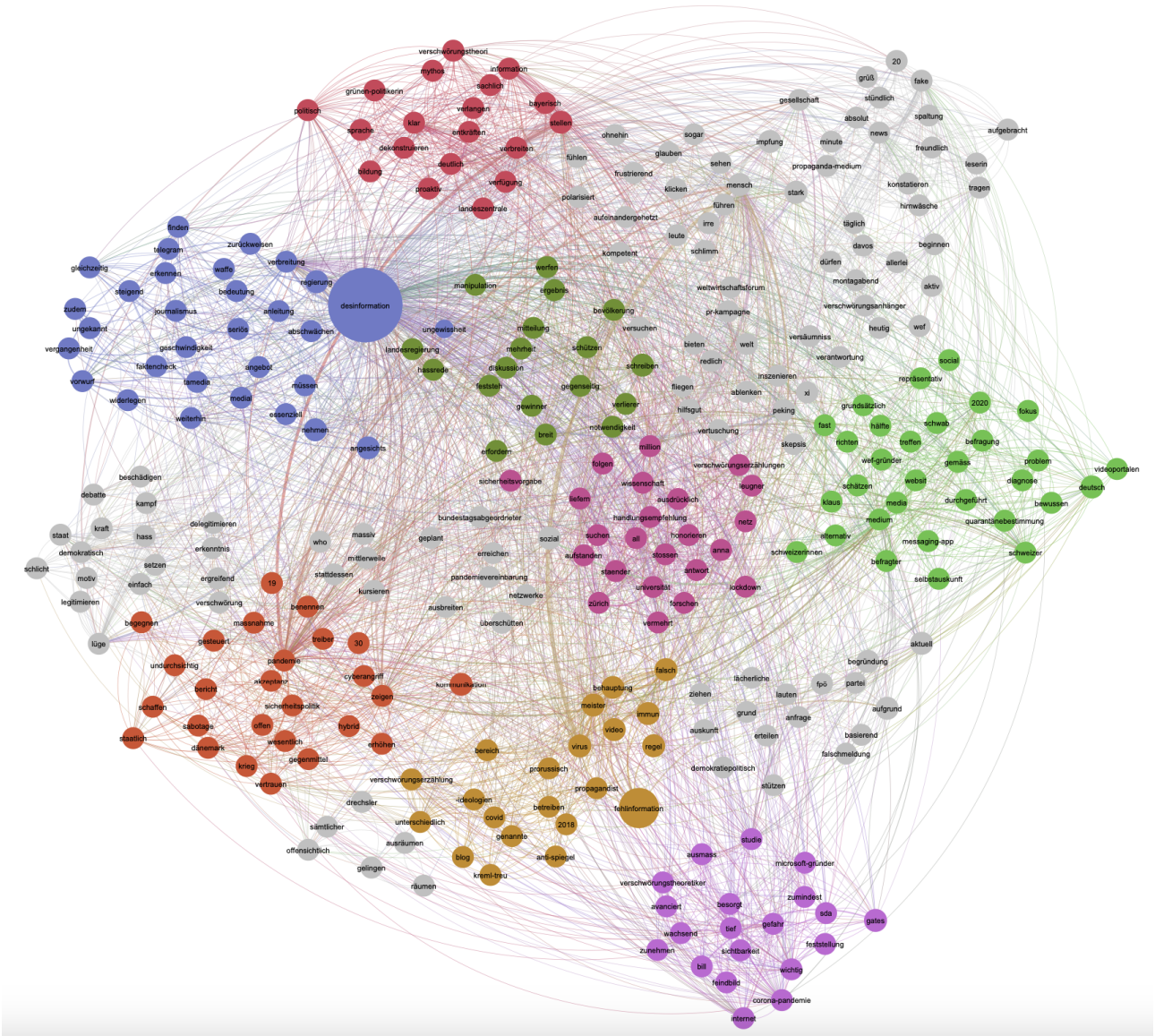
> *Gleichzeitig nimmt die Bedeutung von seriösem Journalismus angesichts der steigenden Verbreitung von Desinformation zu. Dies zeigt das am Montag veröffentlichte “Jahrbuch Qualität der Medien 2021” des Forschungszentrums Öffentlichkeit und Gesellschaft (Fög) der Universität Zürich. Abhilfe schaffen könnte allenfalls der Staat. “Es zeichnet sich immer mehr ab, dass qualitativ hochwertiger Journalismus nur durch eine direkte Medienförderung zu finanzieren ist,” wird Medien-Experte und Fög-Direktor Mark Eisenegger zitiert. / At the same time, the importance of serious journalism is increasing in the face of the spread of disinformation. This is shown in the “Jahrbuch Qualität der Medien 2021”(Yearbook of Quality of the Media 2021) published on Monday by the Research Center for the Public Sphere and Society (Fög) at the University of Zurich. The state could provide a remedy. “It is becoming increasingly apparent that high-quality journalism can only be financed through direct media funding,” said media expert and Fög director Mark Eisenegger.* [Row 8 Factiva dataset, October 25, 2021]

XSL•FO
RenderX

Figure 1. Co-occurrence plot of the terms most commonly associated with “misinformation” and ”disinformation” (*“Fehlinformation” and “Desinformation”*) in the Factiva dataset on news articles. The plot illustrates the strength of co-occurrence between lemmas, with darker blue colors indicating a higher frequency of co-occurrence. The frequency represents the number of times (in percentage, range 0 - 1) in which a lemma has been found in association with “misinformation” and ”disinformation” (either “*Desinformation”* or “*Fehlinformation”*), meaning they were found in the same sentence. Crossed cells indicate no co-occurrence between the listed lemmas and the terms “misinformation” and ”disinformation” (either “*Desinformation”* or “*Fehlinformation”*).

XSL•FO
RenderX

Figure 2. Semantic network map of the lemmas found in the dataset on news media articles (Factiva) associated with “misinformation” and ”disinformation.” Node size represents lemma frequency, edges represent co-occurrence, and color and topology represent modularity class.



A second cluster highlighted in the semantic network map (Figure 2) suggests a theme focused on the difficulties of recognizing, verifying, and mitigating misinformation and disinformation, particularly within the context of rapid media dissemination and institutional credibility. In this context, the terms “Regierung” (government), “Vorwurf” (accusation), and “Verbreitung” (spread), all 3 among the most common co-occurring terms with “misinformation” and “disinformation,” as well as the term “Waffe” (weapon), might indicate disinformation being framed as a strategic tool that undermines governance and public trust. The terms suggest narratives involving accusations against governments and institutions, highlighting the adversarial use of disinformation in political contexts.

> *Sowohl die Schweiz als auch Österreich haben mit der SVP und der FPÖ starke rechte Parteien, die den Corona-Kurs der Regierung torpedieren.*
>
> *In Österreich leben wir ja schon sehr lange mit der FPÖ, ebenso wie die Schweiz mit der SVP. Und es sollte eigentlich niemanden mehr verwundern, wie die beiden Parteien agieren. Aber in einer Krisensituation werden solche Kräfte nochmals zu einer besonderen Belastung. Wenn die Gesundheitssprecherin der FPÖ, die notabene eine Ärztin ist, sich hinstellt und wider alle Tatsachen behauptet, dass die Intensivstationen voller Opfer von Impfschäden seien, ist man schon mittendrin in der Verschwörungs- und Desinformationsszene. Nur sitzen diese Leute im Parlament.*
>
> *Was macht das mit dem politischen Klima im Land?*
>
> *Es schadet der Demokratie. Es geht dabei ja nicht mehr um unterschiedliche Meinungen. Es geht darum, dass politische Kräfte Desinformation und Lügen legitimieren. Wenn man das zulässt, lösen sich alle Parameter auf, nach denen Politik funktioniert. Dann ist alles erlaubt.*
>
> *Both Switzerland and Austria have strong right-wing parties, the SVP and the FPÖ, that are undermining the government’s approach to the coronavirus.*

XSL•FO
RenderX

> *In Austria, we have been living with the FPÖ for a very long time, as has Switzerland with the SVP. And no one should be surprised by the actions of either party. But in a crisis situation, such forces become a particular burden. When the FPÖ's health spokeswoman, who is a doctor, stands up and claims, contrary to all facts, that the intensive care units are full of victims of vaccine damage, you are right in the middle of the conspiracy and disinformation scene. Only these people are in parliament.*
>
> *What does that do to the political climate in the country?*
>
> *It damages democracy. It is no longer about different opinions. It is about political forces legitimizing disinformation and lies. If you allow that to happen, all the parameters by which politics works are dissolved. Then anything goes.* [Row 39 Factiva dataset, December 23, 2021]

Furthermore, the terms "Journalismus" (journalism), "Faktencheck" (fact-checking), "medial" (media-related), and "Tamedia" (a Swiss media company) reflect discussions about the role of media as a channel for countering misinformation and disinformation.

> *Gemäss Selbstauskunft treffen Schweizerinnen und Schweizer besonders auf Social Media, Videoportalen und Messaging-Apps sowie Websites alternativer Medien auf Desinformation. Als Hauptquellen von Falschinformationen nehmen sie Aktivisten, Bürgerinnen, Politiker und Social Bots wahr. Grundsätzlich gelten eher Individuen als Urheber und weniger Organisationen.*
>
> *According to the information they provided, the Swiss encounter disinformation particularly on social media, video portals and messaging apps, as well as on alternative media websites. They perceive activists, citizens, politicians and social bots as the main sources of misinformation. Individuals are generally considered to be the authors more often than organizations.* [Row 30 Factiva dataset, October 25, 2021]

Terms like "seriös" (serious) and "essenziell" (essential) within this cluster might point to the importance of reliable journalism in addressing the issue. The terms "Geschwindigkeit" (speed), "unbekannt" (unknown), and "Ungewissheit" (uncertainty) point to the challenges posed by the rapid and often untraceable spread of misinformation and disinformation, leading to heightened public uncertainty, and the term "zurückweisen" (reject), also among the most common co-occurring terms with "misinformation" and "disinformation", as well as the terms "widerlegen" (refute) and "abschwächen" (weaken), reflect media narratives on the efforts to counter misinformation and disinformation through rebuttal and fact-checking. This cluster seems to highlight the systemic and persistent challenges posed by misinformation and disinformation, suggesting the existence within Swiss media discourse of a dual narrative: on one side, the rapid dissemination of misinformation and disinformation through social media exacerbates public uncertainty, weakening trust in institutions. On the other, it proposes that journalism plays a critical role in fact-checking and mitigating misinformation and disinformation.

The third cluster identified in the semantic network map (Figure 2) reflects a theme centered on the divisive nature of disinformation in public discourse and its role in shaping perceptions of governance and democracy. The terms "Hassrede" (hate speech) and "Lüge" (lie)—a term frequently linked with "misinformation" or "disinformation"—as well as the terms "Verlierer" (loser) and "gegenseitig" (mutual) suggest that misinformation and disinformation are part of a hostile and adversarial public discourse, especially in connection with referendums, possibly indicating that disinformation exacerbates societal divisions through polarizing narratives.

> *Die Reaktionen der siegreichen Befürworter des Covid-Gesetzes und jene der unterlegenen Gegner deuten nicht darauf hin, dass sich die Gräben zwischen den Lagern so rasch wieder schliessen lassen. Die Gewinner sprechen von einem Vertrauensbeweis aus Vernunft, die Verlierer von einer Irreführung.*
>
> *So erwarten etwa die Grünen nach dem "pragmatischen und enorm wichtigen Ja" von den Gegnern des Gesetzes, dass diese das demokratische Ergebnis akzeptieren und "dazu beitragen, dass die Schweiz zu einer gesunden Diskussionskultur zurückfindet." Auch die FDP hofft, dass "künftig wieder konstruktive Ansätze die Debatte prägen."*
>
> *Es sei zu hoffen, dass diejenigen, die sich als selbsternannte Freunde der Verfassung in Szene gesetzt hätten, sich "nun auch als Freunde der direkten Demokratie erweisen und das Abstimmungsergebnis akzeptieren," hieben die Freidenker in die gleiche Kerbe. Sie hatten in Anlehnung an die gegnerischen "Freiheitstrychler" mit dem "Freiheitsimpfler" für ein Ja zum Gesetz geworben.*
>
> *The reactions of the victorious proponents of the Covid law and those of the defeated opponents do not suggest that the rifts between the camps can be closed so quickly. The winners speak of a vote of confidence based on reason, the losers of a deception.*
>
> *The Green Party, for example, expects the opponents of the law to accept the democratic result after the "pragmatic and enormously important yes" vote and to "help Switzerland to rediscover a healthy culture of discussion." The FDP also hopes that "constructive approaches will shape the debate again in the future."*
>
> *It is to be hoped that those who had presented themselves as self-proclaimed friends of the constitution would now also prove to be friends of direct democracy and accept the result of the vote, the freethinkers took the same line. They had campaigned for a yes vote in favor of the law with the "Freiheitsimpfler" (freedom imposter), in reference to the opposing "Freiheitstrychler"*

XSL•FO
RenderX

(freedom truches). [Row 47 Factiva dataset, November 11, 2021]

The presence within this cluster of the terms "Staat" (state) and "demokratisch" (democratic), also a relevant co-occurring term with "misinformation" and "disinformation" (Figure 1), highlights the interplay between misinformation and disinformation and perceptions of governance, suggesting that narratives within public discourse centered around disinformation undermine democratic principles and state legitimacy, as discussed before. Further, the terms "Manipulation" (manipulation), "legitimieren" (legitimize), and "Motiv" (motive) might reflect the intentional use of disinformation to influence public perception, delegitimize opponents, or justify certain actions. These terms point to the strategic framing of narratives to achieve specific political or social goals. The terms "Debatte" (debate), "Diskussion" (discussion), and "schreiben" (write) indicate that misinformation and disinformation often act as catalysts for contentious debates, suggesting that misinformation and disinformation drive public discourse but often in ways that deepen divisions rather than fostering understanding. Within this context, the terms "Gewinner" (winner), "Verlierer" (loser), and "Kraft" (force) appear to suggest that misinformation and disinformation further create competitive narratives where groups or ideas are framed as either victorious or defeated, which could lead to polarization and political or ideological struggles within the society. In sum, this cluster highlights how the media in Switzerland reported on the divisive and strategic nature of misinformation and disinformation in shaping public discourse and governance, reflecting a polarized environment where misinformation and disinformation fuels hostility and deepens societal divides while being used strategically to manipulate perceptions and outcomes. The terms in this cluster also point to the need for collective action to protect democracy.

A fourth cluster touches upon the dissemination of misinformation and disinformation through digital media channels, its ideological dimensions, and the resulting erosion of public trust. For example, the lemmas "prorussisch" (pro-Russian), "kreml-treu" (Kremlin-loyal), and "Propagandist" point to the role of propaganda and ideologically aligned media in spreading disinformation. These terms suggest a discourse and narrative concerning targeted disinformation campaigns originating from specific geopolitical actors or ideologies.

> *Zentrale Vorstellung ist, dass die Corona-Pandemie „seitens der Eliten seit Jahren vorbereitet wurde" – eine gängige Darstellung von jenen, die ein weltweites konspiratives, selbstverständliches jüdisches Netz vermuten, das letztlich am Untergang des Abendlandes und an einer „Öko-Diktatur" arbeitet. Und so mündet eine unsinnige Frage in die andere, immer davon ausgehend, dass die Pandemie von langer Hand – um nicht zu sagen von dunklen Mächten – vorbereitet wurde. Und da kommt dann die Frage – wie „diese Erkenntnisse bei Ihren Entscheidungen zu den Corona-Maßnahmen in der Zukunft" berücksichtigt werden. Desinformation und Fake News sind schlicht und ergreifend keine Erkenntnisse, sie werden einfach aus den unterschiedlichsten Motiven in die Welt gesetzt – oft und sehr oft um demokratische Staaten zu delegitimieren.*
>
> *The central idea is that the Corona pandemic "has been prepared by the elites for years"—a common belief of those who suspect a worldwide conspiratorial, self-evident Jewish network that is ultimately working on the downfall of the West and an "eco-dictatorship." And so one nonsensical question leads to another, always assuming that the pandemic was prepared long in advance—not to mention by dark forces. And then comes the question of how "these findings will be taken into account in your decisions on corona measures in the future". Disinformation and fake news are simply not insights; they are simply created for a wide variety of reasons—often and very often to delegitimize democratic states.* [Row 36 Factiva dataset, July 26, 2022]

The terms "Blog," "Video," "betreiben" (to operate), and "Bereich" (sector/field) within this cluster might underline the role of digital media platforms and blogs in amplifying misinformation and disinformation narratives. The terms "Covid," "Virus," "Immun" (immune), and "regel" (Rule), commonly co-occurring with the terms "misinformation" and "disinformation" (Figure 1), highlight that at least a part of the media discourse focused on health-related misinformation and disinformation and its societal implications. This cluster seems to be centered around discussions in Swiss media about the role of digital platforms and ideologically driven content, highlighting the intersection of conspiracy theories, health misinformation, and geopolitical propaganda, all contributing to a fragmented and polluted information landscape.

Finally, a fifth cluster touches upon themes around the influence of conspiracy theories, prominent individuals as targets of disinformation, and the evolving visibility of narratives in the context of the pandemic. For example, "Verschwörungstheoretiker" (conspiracy theorist), "Feindbild" (enemy image), and "Feststellung" (assertion) point to the centrality of conspiracy theories in misinformation and disinformation narratives. These terms suggest that conspiracy theorists shape the public's perception by presenting simplified or distorted explanations of events, often targeting specific individuals or groups. Terms such as "Microsoft-Gründer" (Microsoft founder), "Bill," and "Gates" indicate the focus on Bill Gates as a prominent figure in conspiracy theories related to COVID-19, where he is frequently portrayed as a central figure and used as an example of powerful individuals framed as scapegoats in disinformation narratives.

> *Der nicht erst in der Corona-Pandemie zum Feindbild von Verschwörungstheoretikern avancierte Microsoft-Gründer Bill Gates ist tief besorgt über die Verbreitung von Desinformation und Lügen im Internet. 'Diese verrückten Ideen verbreiten sich irgendwie schneller in den sozialen Medien als die Wahrheit. Ich bin überrascht, dass mein Name in diesen Verschwörungstheorien auftaucht […]. Ich finde, dass es irgendwie ironisch ist, dass ich*

XSL•FO
RenderX

> *anmahnte, auf diese Pandemie vorbereitet zu sein - und jetzt gibt es Leute, die sagen, ich sei dafür verantwortlich.' Der Multimilliardär appellierte an die Vernunft der Menschen: 'Wir befinden uns inmitten einer Pandemie, und es ist wichtiger als je zuvor, sich mit den Tatsachen und der Wahrheit auseinanderzusetzen.'*
>
> *Microsoft founder Bill Gates, who has been an enemy stereotype of conspiracy theorists since before the coronavirus pandemic, is deeply concerned about the spread of disinformation and lies on the internet. 'Somehow these crazy ideas spread faster on social media than the truth. I'm surprised my name comes up in these conspiracy theories […]. I think it's kind of ironic that I urged people to be prepared for this pandemic – and now there are people saying I'm responsible.' The multibillionaire appealed to people's common sense: 'We are in the midst of a pandemic, and it's more important than ever to deal with the facts and the truth.'* [Row 11 Factiva dataset, September 15, 2020]

And to be effective, conspiracy theories and disinformation are associated with an emotional manipulative intent: terms such as "besorgt" (worried) and "tief" (deep) reflect the emotional impact of disinformation, which can amplify fear and distrust among the public. Overall, this cluster highlights the presence, within Swiss media discourse, of a narrative focused on conspiracy theories and how they intersected with and shaped public perceptions and public discourse during the COVID-19 pandemic.

Social Media (CrowdTangle Dataset)

The entire retrieved dataset included 580 Facebook posts from pages and groups. From the analysis of the social media dataset obtained via CrowdTangle, the NLP analysis visualized through the co-occurrence analysis plot of the terms most commonly associated with "misinformation" and "disinformation" (Figure 3) and the semantic network map of the terms co-occurring in association with "misinformation" and disinformation" (Figure 4) reveals 4 relevant themes.

The first theme focuses on the proliferation of health-related misinformation during the pandemic, particularly on vaccinations and public health measures. Terms such as "Corona," "Corona-Pandemie" (coronavirus pandemic), and "Corona-Impfung" (coronavirus vaccination), all present among the most commonly co-occurring terms with "misinformation" and "disinformation" (Figure 3), as well as the term "Durchseuchung" (herd immunity), highlight discussions on the pandemic's progression and vaccine-related debates. Social media posts frequently questioned the effectiveness of vaccines, with claims often rooted in "Fehlinterpretation" (misinterpretation) or deliberate framing of scientific findings. Combined with terms such as "Schutz" (protection) and "Problem" (problem), this cluster seems to highlight a debate within public discourse around the protective benefits of vaccines and other interventions.

> *Ungeimpft. Eine Begegnung Eine Bekannte, die ich eigentlich recht gut mag, ist mit Ü50 noch nicht geimpft. Sie ist keine Massnahmen-Gegnerin und keine militante Schwurblerin. Sie hat einfach Angst vor möglichen Folgen einer Impfung. Auf die Frage, warum sie denn Angst habe, führt sie die ihr zugänglichen Informationen an. In den Medien seien ja immer wieder widersprüchliche Aussagen zu vernehmen. Studien und Gegenstudien, sogar Ärzte, die von der Impfung abraten. Meine Bekannte ist Künstlerin, Wissenschaft und Medien liegen ihr nicht nahe, haben in ihrem Leben selten eine Rolle gespielt. Jetzt sieht sie sich einem Wust aus Information ausgeliefert, die sie persönlich nicht einordnen kann. Sie ist überfordert. Sie anerkennt die Gefährlichkeit von Covid, und isoliert sich deshalb, trägt Maske, hält Distanz. Sie leidet unter ihrer Angst und auch an den Folgen ihrer Nicht-Impfung. Ihr Partner ist geimpft, sie streiten sich deswegen. Die Verantwortung für ihre Situation tragen Medien wie Nau, 20 Minuten oder die #SRFArena, die unwissenschaftlichen Scheissdreck gleichwertig neben wissenschaftlichen Erkenntnissen präsentieren, weil quoten- oder klickgeil. Meine Bekannte kann nicht zwischen einer MIT- oder ETH-Studie und einer Umfrage eines Füdli-Institutes unterscheiden, sie kann wissenschaftliche Forschung nicht einordnen, wie viele in unserer Gesellschaft. […] Und es ist Aufgabe des Staates, diese Informationen unzweifelhaft zu vermitteln. Auch in Schulen. Und sorry, da haben die meisten versagt. Wenn wir also Menschen haben, die sich aus Angst und Überforderung nicht impfen lassen, die sich isolieren, die klare Einordnung suchen, aber diese nicht bekommen, ist das ein gesellschaftliches Versagen. Die Medien haben den Schwurbler-Idioten eine breite Bühne geboten. Die Regierung hat sich vor den Massnahmen-Verweigerern eingepisst. Sorry, ihr habt versagt. Und jetzt geht, und räumt euren Scheiss auf.*
>
> *Unvaccinated. An encounter. A friend of mine, who I actually like quite a lot, is over 50 and still not vaccinated. She is not opposed to the measures and is not a militant swearing person. She is simply afraid of the possible consequences of a vaccination. When asked why she is afraid, she cites the information available to her. After all, contradictory statements are repeatedly heard in the media. Studies and counter-studies, even doctors advising against vaccination. My acquaintance is an artist, science and the media are not close to her, have rarely played a role in her life. Now she finds herself at the mercy of a jumble of information that she personally cannot process. She is overwhelmed. She recognizes the danger of Covid, and therefore isolates herself, wears a mask, keeps her distance. She suffers from her fear and also from the consequences of not being vaccinated. Her partner is vaccinated, and they argue about it. The responsibility for her situation lies with media outlets like Nau, 20 Minuten or the #SRFArena, which present unscientific nonsense alongside scientific findings because they are hungry for ratings*

XSL•FO
RenderX

> *or clicks. My acquaintance cannot distinguish between a MIT or ETH study and a survey by a Füdli institute, she cannot classify scientific research, like many in our society. […] And it is the task of the state to communicate this information unequivocally. Also in schools. And sorry, most of them have failed there. So if we have people who, out of fear and overwhelm, don't get vaccinated, isolate themselves, and seek clear categorization but don't get it, that's a social failure. The media has given the swearing idiots a broad stage. The government has pissed itself off the measure objectors. Sorry, you have failed. Now go and clean up your mess* [Row 13 CrowdTangle dataset, September 4, 2021]

Furthermore, public discourse within this theme included discussions on conspiracy theories, which dominated a significant portion of social media discourse, with terms such as "*QAnon"* (a global conspiracy theory movement), showcasing how global narratives infiltrated Swiss social media spaces. These theories often intersected with broader ideological narratives, such as antiestablishment sentiments and critiques of Western political structures (see the presence of the term "USA" in this corpus; Figure 4). Emotional terms, such as "Hetze" (hate speech) and "verkommen" (degraded), suggest the possible adversarial and polarized nature of these discussions, where extreme rhetoric ("extrem" is a term present in this cluster) fuels distrust. Posts associated with conspiracy theories frequently leveraged these narratives to legitimize alternative explanations and sow doubt about the government or the science behind the government's decisions, showing how social media acted during the COVID-19 pandemic in Switzerland as a breeding ground for sensationalist and ideologically driven content.

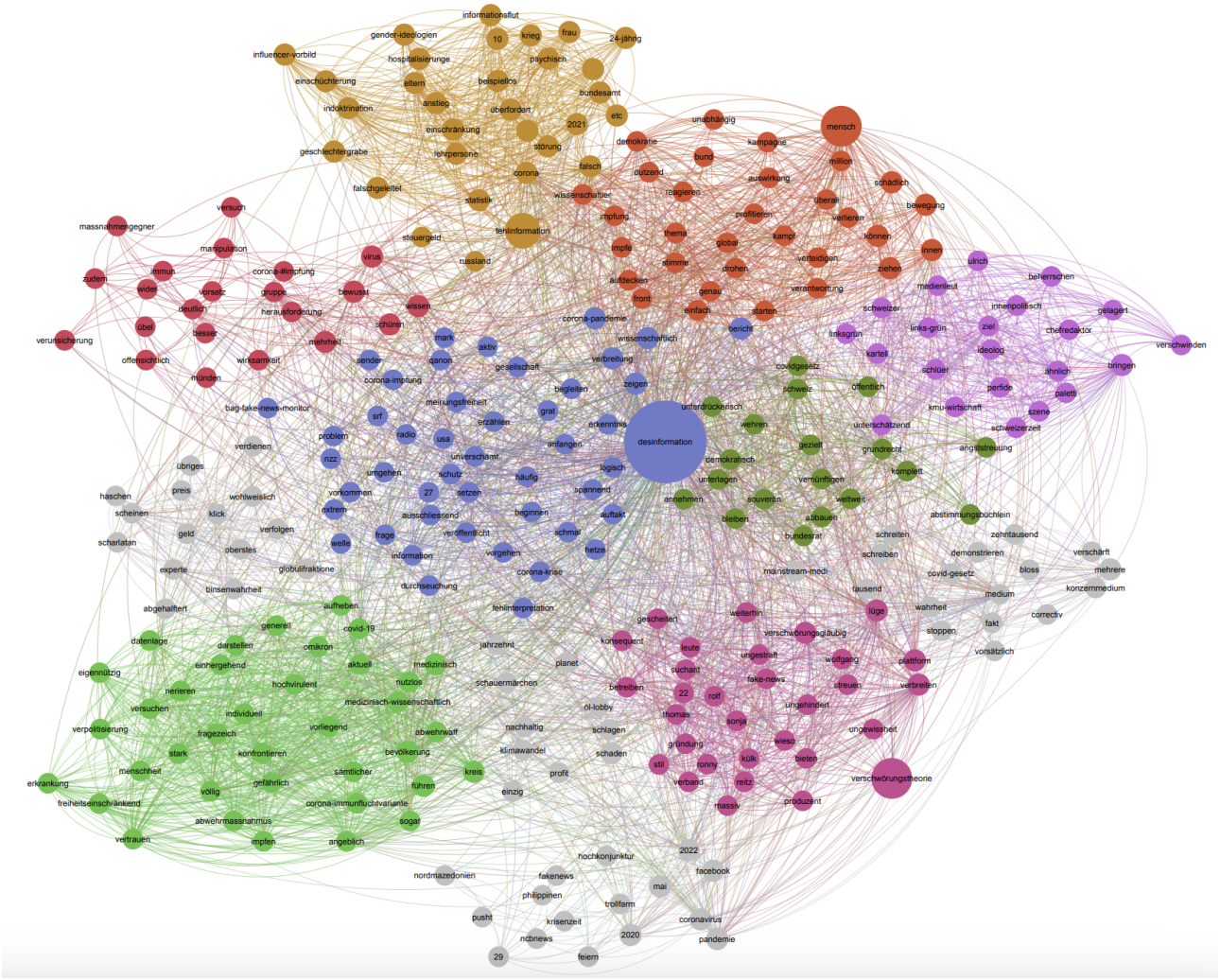
> *Es bleibt nicht mehr viel Zeit, uns komplett dagegen zu wehren und uns für unsere demokratischen Grundrechte zu wehren! Sie wurden und werden gerade weltweit abgebaut! In der Schweiz unterlagen leider die Vernünftigen und das unterdrückerische Covidgesetz wurde wegen gezielter Angststreuung, und Desinformation durch den Bundesrat im Abstimmungsbüchlein, angenommen vom Souverän. Aber es geht nicht um ein Virus! Es geht um unsere Entrechtung und Enteignung! Die WHO soll neue WELTREGIERUNG im Pandemiefall werden! Und die können sie dank dafür nicht geeignetem PCR NAT jederzeit ausrufen! Mehrere Gerichtsurteile bestätigen, dass ein PCR NAT an Gesunden nichts aussagt! Auch sein Erfinder, der dafür einen Nobelpreis bekam, sagte das immer! Dr. Kari Mullis, USA, gestorben 2019.*
>
> *There is not much time left to fight back completely and defend ourselves for our fundamental democratic rights! They have been and are being dismantled worldwide! In Switzerland, unfortunately, the sensible were subject to it and the oppressive Covid law was adopted by the sovereign because of the deliberate spreading of fear and disinformation by the Federal Council in the voting booklet. But it's not about a virus! It's about our disenfranchisement and expropriation! The WHO is supposed to become the new WORLD GOVERNMENT in the event of a pandemic! And thanks to the unsuitable PCR NAT, they can declare it at any time! Several court rulings confirm that a PCR NAT says nothing about the health of a person! Even its inventor, who received a Nobel Prize for it, always said that! Dr. Kari Mullis, USA, died in 2019* [Row 159 Crowdtangle dataset, April 20, 2022]
>
> *Äusserlich stimmt, dass die Summe aller Arbeitsplätze nicht dramatisch gesunken ist. Dies, weil die Staatsbürokratie auch in Corona-Zeiten – teils gerade wegen Corona – ungebremst ausgewuchert ist: Für all die akribische Kontrolle eingesperrter Bürgerinnen und Bürger sowie spitzfindiger Regulierung unterworfener Betriebe. Aber auch, weil weiterhin unverdrossen selbst für Unsinnigstes laufend neue Zweige der Funktionärsbürokratie neu geschaffen werden. In Zürich zum Beispiel eine «Fachkommission» zwecks Beratung des Stadtrats im Hinblick auf die Ausmerzung als rassistisch vermuteter Häuser-Namen. Gezielt wird dabei zum Beispiel auf das seit Jahrzehnten die Zwinglistadt schmückende «Haus zum Mohrenkopf».*
>
> *On the surface, it is true that the total number of jobs has not fallen dramatically. This is because the state bureaucracy has continued to grow unchecked even in the age of the coronavirus – partly precisely because of the coronavirus: for all the meticulous control of locked-up citizens and the meticulous regulation of businesses. But also because new branches of the bureaucratic apparatus continue to be created unabated, even for the most nonsensical of reasons. In Zurich, for example, a "specialist commission" has been set up to advise the city council on the eradication of house names suspected of being racist. One of the targets is the "Haus zum Mohrenkopf," which has graced the Zwinglistadt for decades* [Row 93 Crowdtangle dataset, March 18, 2022]

XSL•FO
RenderX

Figure 3. Co-occurrence plot of the terms most commonly associated with “misinformation” and ”disinformation” (*“Fehlinformation” and “Desinformation”*) in the CrowdTangle dataset on social media. The plot illustrates the strength of co-occurrence between lemmas, with darker blue colors indicating a higher frequency of co-occurrence. The frequency represents the number of times (in percentage, range 0 - 1) in which a lemma has been found in association with “misinformation” and ”disinformation” (either “*Desinformation”* or “*Fehlinformation*”), meaning they were found in the same sentence. Crossed cells indicate no co-occurrence between the listed lemmas and the terms “misinformation” and ”disinformation” (either “*Desinformation”* or “*Fehlinformation*”).

XSL•FO
RenderX

Figure 4. Semantic network map of the lemmas found in the dataset on social media (CrowdTangle) associated with “misinformation” and ”disinformation.” Node size represents lemma frequency, edges represent co-occurrence, and color and topology represent modularity class.



A second theme within the corpus explores the tension between public health measures and individual freedoms, framed through the lens of democracy and civil liberties. Terms such as “demokratisch” (democratic) and “Grundrecht” (fundamental right) reflect discussions about democratic values and protecting fundamental rights during crises. At the same time, terms such as “unterdrückerisch” (oppressive) and “souverän” (sovereign) suggest the presence of polarized narratives within social media discourse, where public health mandates were depicted as overreaching or infringing on personal rights and “sovereignty.” Further focusing on the polarized nature of social media discourse, the semantic network map (Figure 4) suggests that narratives of resistance and compliance clashed. Terms such as “wehren” (to resist) and “gezielt” (targeted), the latter present among the most co-occurring terms with “misinformation” and “disinformation” (Figure 3), suggest organized resistance against perceived institutional overreach, with “abbauen” (dismantle) possibly reflecting arguments aimed at dismantling or modifying public health measures, whereas on the contrary, discussions also framed mandates as reasonable (“vernünftigen”). For example:

> *«Der Bundesrat entschädigt den Schaden, den er selber angerichtet hat nur, wenn er Vollmachten bekommt. Das ist Erpressung.» Nicolas A. Rimoldi, Co-Präsident von MASS-VOLL!, debattiert gegen Manuela Weichelt, Nationalrätin der Grünen, zum Covid-19-Gesetz. Uns wird seit mehr als einem Jahr eine alternativlose Politik präsentiert, die suggeriert, dass nur das massive Eingreifen des Staates in allen Bereichen des Lebens einen Weg aus der selbstverschuldeten Krise darstellt. Dass die vom Bundesrat getroffenen Massnahmen die wirtschaftliche, soziale und politische Krise verursacht haben, wird ignoriert. Die im Covid-19-Gesetz geregelten Finanzhilfen sollen den geschädigten Unternehmen Abhilfe schaffen – doch ein Gesetz, welches die Souveränität des Volkes massiv einschränkt und eine Zweiklassengesellschaft einführt, darf in unserer Demokratie keine ernsthafte Option sein. Sowieso: Die “Entschädigungen” sind nur Almosen, sie helfen nicht. Nur ein NEIN führt dazu, dass alle vollumfassend entschädigt werden! Nutze Deine Stimme für ein NEIN zum Covid-19-Gesetz am 13. Juni - ein NEIN zu Entmachtung, zu Willkür und zu einer alternativlosen Politik. Unterstütze auch Du MASS-VOLL! und sag JA zur Selbstbestimmung. Quelle: Kontrovers – Zentralschweizer Fernsehen*

XSL•FO
RenderX

> *Manuela Weichelt, member of the National Council for the Greens, on the Covid-19 Act. For more than a year, we have been presented with a policy with no alternative, suggesting that the only way out of the self-inflicted crisis is massive state intervention in all areas of life. The fact that the measures taken by the Federal Council have caused the economic, social and political crisis is being ignored. The financial aid regulated by the Covid-19 Act is intended to provide relief for the companies that have suffered damage - but a law that massively restricts the sovereignty of the people and introduces a two-class society must not be a serious option in our democracy. In any case: the "compensation" is just charity, it does not help. Only a NO vote will lead to everyone being fully compensated! Use your vote for a NO to the Covid-19 law on June 13—a NO to disempowerment, to despotism and to a policy without alternatives. You too can support MASS-VOLL! and say YES to self-determination. Source: Kontrovers - Zentralschweizer Fernsehen* [Row 6579 Crowdtangle Dataset, May 29, 2021]

The third theme captured the phenomenon of information overload during the pandemic, where the volume of information—both accurate and false—left the public feeling "überfordert" (overwhelmed), a term strongly associated with "misinformation" and "disinformation" in our analysis (Figure 3). Terms such as "Statistik" (statistics) and "falsch" (false), also among the most commonly co-occurring terms with "misinformation" or "disinformation" (Figure 3), point to discourses questioning the accuracy and reliability of reported data on COVID-19 cases, deaths, and hospitalizations. In addition, terms such as "psychisch" (psychological) and "Störung" (disorder) likely reflect growing concerns about mental health issues linked to prolonged restrictions ("Einschränkung") and the fear-driven narratives spread online. The term "*beispiellos*" (unprecedented) highlights the "unprecedented" nature of these challenges, while "Hospitalisierungen" (hospitalizations) suggests anxiety around the strain on health care systems. All the abovementioned terms, from "psychisch" to "Hospitalisierungen," are included in the list of the most commonly co-occurring terms with "misinformation" and "disinformation" (Figure 3). Furthermore, the presence of the term "Demokratie" (democracy) within this thematic cluster in the semantic network map (Figure 4) reveals that the discourse on fear and anxiety exacerbated by misinformation and disinformation touched upon the tensions and balance between maintaining democratic norms and implementing effective public health measures.

The fourth theme identified suggests that social media discourse in Switzerland was shaped around the relationship between misinformation and disinformation and democratic processes. Terms such as "Demokratie" (democracy) and "*unabhängig*" (independent) point to concerns within public discourse about the erosion of democratic norms due to the spread of misinformation and disinformation, with narratives often calling for accountability ("aufdecken"). The centrality of the term "*mensch*" (human, person) in this cluster—featured as a central node in the semantic network map (Figure 4) and prominently listed among the terms most frequently co-occurring with "misinformation" and "disinformation" (Figure 3)—underlines the human impact of misinformation and disinformation, with discussions focusing on how misinformation and disinformation affects individuals and communities. The term "Verantwortung" (responsibility) likely reflects calls for collective accountability, with an emphasis on the ethical obligation to counteract narratives that are harmful ("schädlich"). Terms such as "Bewegung" (movement) and "verteidigen" (to defend) possibly suggest grassroots efforts within social media discourse to defend what they perceive as human values and societal norms. This theme suggests that, beyond institutional, ideological, and polarized debates, the ultimate stakes of misinformation and disinformation are their tangible impacts on human lives and social cohesion, as recognized by people shaping public discourse on social media.

> *Als ich die Nachricht hörte, dachte ich an einen schlechten Scherz. Doch das Bundesamt für Gesundheit (BAG) bestätigt entsprechende Pressemeldungen: Die Beamten im Innendepartement von SP-Bundesrat Alain Berset haben heimlich einen sogenannten «Fake-News-Monitor» eingerichtet. «Eine spezielle Software wertet für die Behörden soziale Medien, Nachrichtenportale und Zeitungen nach (Falsch-)Informationen aus. So erhalten sie täglich einen Überblick über die Informationen, die in der Schweiz kursieren», schreibt der «Blick». Und fügt treuherzig hinzu, die Software diene dem BAG als «Frühwarnsystem». So könne der Bund mit «Kampagnen» auf Desinformation reagieren, etwa zum Thema Impfen. Ein «Wahrheitsministerium» wie bei Orwell Ich weiss nicht, wie es Ihnen geht – aber mir läuft es dabei kalt den Rücken herunter. Ein Bundesamt für Staatspropaganda? Ein «Wahrheitsministerium»? Irgendwie kommt einem das bekannt vor. George Orwell hat dies in seinem antitotalitären Zukunftsroman «1984», geschrieben 1948, anschaulich vorweggenommen.*
>
> *When I heard the news, I thought it was a bad joke. But the Federal Office of Public Health (FOPH) confirmed the corresponding press reports: the officials in the interior department of SP Federal Councillor Alain Berset have secretly set up a so-called "fake news monitor." "Special software evaluates social media, news portals and newspapers for (false) information for the authorities. This provides them with a daily overview of the information circulating in Switzerland," writes the Blick. And adds, ingenuously, that the software serves as an 'early warning system' for the FOPH. This would enable the federal government to respond to disinformation with "campaigns," for example on the subject of vaccination. A "Ministry of Truth" like in Orwell I don't know about you, but it sends a shiver down my spine. A federal office for state propaganda? A "Ministry of Truth"? Somehow it seems familiar.* [Row 59 Crowdtangle dataset, 22.08.2021]

XSL•FO
RenderX

Discussion

Contested Meanings of Misinformation and Disinformation in Swiss Public Discourse

The analysis shows that the understanding, conceptualization, and use of "misinformation" and "disinformation" operate as a critical point of contention, balancing the priorities of individual liberties and collective responsibilities during times of crisis. This dynamic reflects the inherent tensions in public discourse, as debates clash with the trade-offs between personal freedoms and the broader public good. This aligns with previous findings on Swiss public discourse, where the concept of "freedom" was shaped within the broader debate on balancing individual and collective rights [20]. News media focused on evaluating the impacts of misinformation and disinformation, particularly the tension between journalistic integrity and the amplification of falsehoods. This occurred within a fragmented and polarized information environment, in which social media platforms provided a fertile ground for polarized narratives [34-36]. These narratives frequently intersected with conspiracy theories, emotional appeals, and ideological conflicts, exacerbating societal polarization [37]. If the Factiva dataset captured a narrative of journalistic accountability, highlighting journalism's role in mitigating the spread of falsehoods, the CrowdTangle dataset revealed a markedly different pattern, dominated by a discourse characterized by emotionally charged content and organized resistance to perceived institutional overreach. In mainstream media, misinformation and disinformation are generally considered a challenge, with debates about editorial responsibility and the role of fact-checking in debunking misinformation and disinformation. By contrast, on social media, the same concepts are often discussed and interpreted, as we have seen, as instruments of control, with fact-checking seen as a threat to freedom of expression. We argue that this divergence illustrates a fracture between institutional and grassroots narratives, causing a fragmented and polarized public discourse, where efforts to preserve informational accuracy in one arena can be perceived as censorship in another. We argue that this reflects competing epistemic frameworks that can undermine the existence of points of contact between diverging views and opinions in Swiss society, causing damage and fueling polarization.

In light of these findings and the ongoing societal debates about fact-checking and content moderation [38], we explore potential strategies to reduce fragmentation within polarized public discourse, recognizing that addressing this challenge requires a careful balance between promoting open debate and mitigating the influence of misinformation and disinformation. Rather than assuming that public conversations and differing opinions naturally emerge from well-informed deliberation, we acknowledge that many individuals engage with information in ways that are shaped by cognitive biases, social dynamics, and varying levels of health and media literacy [39]. Moreover, the quality and reliability of the information on which people base their choices are not only essential for individual decision-making but also a prerequisite for democracies that are not merely formal but substantive—where civic participation is grounded in informed judgment rather than manipulation or epistemic asymmetries. Thus, any effort to strengthen public discourse must consider the complexity of how information is processed and acted upon.

For news media, we consider the role of fact-checking mechanisms and editorial accountability as essential tools for ensuring the reliability of information, which need systematic application. While robust verification processes can help reduce the spread of false or misleading content, their effectiveness depends on how deeply they are integrated into journalistic culture and practices, and whether audiences trust and engage with them. Additionally, editorial standards should not only prevent the dissemination of inaccuracies but also encourage nuanced reporting that resists sensationalism and oversimplification, which can contribute to polarization even in the absence of outright falsehoods.

For social media platforms, initiatives aimed at enhancing public engagement and improving information literacy have potential benefits. While some interventions—such as algorithmic adjustments to prioritize credible sources—may help reduce exposure to misinformation, they must be implemented with caution to avoid unintended consequences, such as reinforcing ideological silos, echo chambers, or diminishing pluralism. Likewise, efforts to promote information literacy should not be framed solely as a reactive defense against manipulation, but as a broader proactive means of cultivating critical thinking skills that enable individuals to assess claims.

The extent to which these approaches can mitigate the spread of misinformation and enhance public resilience to its effects likely depends on multiple factors, including institutional cooperation, societal trust in information sources, and the willingness of individuals to engage in reflective, rather than reactive, forms of reasoning. In this sense, this study contributes to understanding societal resilience not merely as an outcome of effective crisis communication, but as a process sustained by epistemic robustness, trust, and deliberative inclusiveness—arguably, resilience is rooted in the capacity of societies to maintain transparent, plural, and reasoned public debate even under conditions of uncertainty. Acknowledging these challenges, we argue, promotes an epistemic culture in which knowledge is critically examined, rather than passively received or reflexively accepted or rejected; this may be a more sustainable long-term goal than attempting to eliminate misinformation outright.

Limitations

This study has certain limitations that should be considered when interpreting the findings. First, the representativeness of the datasets is limited by design: social media data, collected via CrowdTangle, capture publicly shared content, potentially overrepresenting perspectives from certain demographic groups while underrepresenting others, such as individuals who prefer private interactions or alternative communication platforms. Second, although the datasets enable longitudinal analysis, this study does not systematically examine temporal variations in discourse. Third, while the findings highlight significant trends and divergences in framing misinformation and disinformation, the analysis necessarily simplifies complex social constructs, possibly without fully capturing nuances and flattening the

XSL•FO
RenderX

complexity of public discourse. Finally, for future research, there is a need for a deeper, qualitative examination of how misinformation and disinformation have been addressed in public discourse during the COVID-19 pandemic.

Funding

This work was funded by the Swiss National Science Foundation, National Research Programme 80, grant number 408040 - 209947.

Authors' Contributions

Analysis: GS, NBA
Conceptualization: FG, GS, NBA
Data collection: FF, GS
Supervision: SM
Visualization: FG, NBA
Writing (final edit of the manuscript): FF, FG, GS, NBA, SM
Writing (original draft): FG, GS, NBA

Conflicts of Interest

None declared.

References

1. Purnat TD, Nguyen T, Briand S, editors. Managing Infodemics in the 21st Century: Addressing New Public Health Challenges in the Information Ecosystem: Springer International Publishing; 2023. [doi: 10.1007/978-3-031-27789-4]
2. Briand SC, Cinelli M, Nguyen T, et al. Infodemics: a new challenge for public health. Cell 2021 Dec 9;184(25):6010-6014. [doi: 10.1016/j.cell.2021.10.031] [Medline: 34890548]
3. Hart PS, Chinn S, Soroka S. Politicization and polarization in COVID-19 news coverage. Sci Commun 2020 Oct;42(5):679-697. [doi: 10.1177/1075547020950735] [Medline: 38602988]
4. The Lancet Infectious Diseases. The COVID-19 infodemic. Lancet Infect Dis 2020 Aug;20(8):875. [doi: 10.1016/S1473-3099(20)30565-X] [Medline: 32687807]
5. Rocha YM, de Moura GA, Desidério GA, de Oliveira CH, Lourenço FD, de Figueiredo Nicolete LD. The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. J Public Health (Berl) 2023 Jul;31(7):1007-1016. [doi: 10.1007/s10389-021-01658-z]
6. Baines D, Elliott RJR. Defining misinformation, disinformation and malinformation: an urgent need for clarity during the COVID-19 infodemic. : University of Birmingham; 2020 Apr URL: https://repec.cal.bham.ac.uk/pdf/20-06.pdf [accessed 2026-03-06]
7. Larson HJ. Stuck: How Vaccine Rumors Start—and Why They Don't Go Away: Oxford University Press; 2020.
8. Krekó P. Countering conspiracy theories and misinformation. In: Routledge Handbook of Conspiracy Theories: Routledge; 2020:242-255 URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9780429452734-2_8/countering-conspiracy-theories-misinformation-p%C3%A9ter-krek%C3%B3 [accessed 2026-03-06]
9. Farrell J, McConnell K, Brulle R. Evidence-based strategies to combat scientific misinformation. Nat Clim Chang 2019 Mar;9(3):191-195. [doi: 10.1038/s41558-018-0368-6]
10. Wernli D, Clausin M, Antulov-Fantulin N, et al. Building a multisystemic understanding of societal resilience to the COVID-19 pandemic. BMJ Glob Health 2021 Jul;6(7):e006794. [doi: 10.1136/bmjgh-2021-006794] [Medline: 34301677]
11. Rauchfleisch A, Vogler D, Eisenegger M. Public sphere in crisis mode: how the COVID-19 pandemic influenced public discourse and user behaviour in the Swiss Twitter-sphere. Javnost 2021;28(2):129-148. [doi: 10.1080/13183222.2021.1923622] [Medline: 34393592]
12. Scardigno R, Musso P, Cicirelli PG, D'Errico F. Health communication in the time of COVID-19 pandemic: a qualitative analysis of Italian advertisements. Int J Environ Res Public Health 2023 Mar 1;20(5):4424. [doi: 10.3390/ijerph20054424] [Medline: 36901433]
13. Ribeiro MH, Calais PH, Almeida VAF, Meira W. "Everything I disagree with is #fakenews": correlating political polarization and spread of misinformation. arXiv. Preprint posted online on Jan 19, 2017. [doi: 10.48550/arXiv.1706.05924]
14. Tong C, Gill H, Li J, Valenzuela S, Rojas H. "Fake news is anything they say!"—Conceptualization and weaponization of fake news among the American public. Mass Commun Soc 2020 Sep 2;23(5):755-778. [doi: 10.1080/15205436.2020.1789661]
15. Au CH, Ho KKW, Chiu DKW. The role of online misinformation and fake news in ideological polarization: barriers, catalysts, and implications. Inf Syst Front 2022 Aug;24(4):1331-1354. [doi: 10.1007/s10796-021-10133-9]
16. Ferreira Caceres MM, Sosa JP, Lawrence JA, et al. The impact of misinformation on the COVID-19 pandemic. AIMS Public Health 2022;9(2):262-277. [doi: 10.3934/publichealth.2022018] [Medline: 35634019]

XSL•FO
RenderX

17. Cinelli M, Quattrociocchi W, Galeazzi A, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 6;10(1):16598. [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
18. Entman RM. Framing: towards clarification of a fractured paradigm. J Commun 1993 Dec 1;43(4):51-58. [doi: 10.1111/j.1460-2466.1993.tb01304.x]
19. Boosting public discourse: a targeted, evidence-based approach to improving moral reasoning. Swiss National Science Foundation. URL: https://www.nfp80.ch/en/w3TefkOlhfj46IGx/project/boosting-public-discourse-a-targeted-evidence-based-approach-to-improving-moral-reasoning [accessed 2024-12-05]
20. Spitale G, Germani F, Fritschi F, Merten S, Biller-Andorno N. Analyzing freedom: understanding Swiss COVID-19 narratives through NLP analysis. SocArXiv. Preprint posted online on Dec 17, 2024. [doi: 10.31235/osf.io/gehv7]
21. Federal act on research involving human beings (human research act, HRA). Fedlex: The Publication Platform for Federal Law. 2023. URL: https://www.fedlex.admin.ch/eli/cc/2013/617/en [accessed 2026-02-26]
22. Factiva: business intelligence & news monitoring. Dow Jones. URL: https://www.dowjones.com/professional/factiva/ [accessed 2025-01-15]
23. NRP80—boosting public discourse: towards a targeted, evidence-based strategy to improve moral reasoning. OSF. URL: https://osf.io/k4ja5/ [accessed 2025-01-15]
24. CrowdTangle. Transparency Center. URL: https://transparency.meta.com/researchtools/other-datasets/crowdtangle [accessed 2025-01-15]
25. FAQ: local relevance filter. CrowdTangle Help Center. URL: https://web.archive.org/web/20230326191543/https://help.crowdtangle.com/en/articles/4523893-faq-local-relevance-filter [accessed 2025-04-07]
26. Industrial-strength natural language processing in Python. spaCy. URL: https://spacy.io/ [accessed 2025-01-15]
27. About Gephi. Gephi. 2022. URL: https://gephi.org/about/ [accessed 2025-01-15]
28. Lambiotte R, Delvenne JC, Barahona M. Random walks, Markov processes and the multiscale modular organization of complex networks. IEEE Trans Netw Sci Eng 2014 Jul;1(2):76-90. [doi: 10.1109/TNSE.2015.2391998]
29. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech: Theory Exp 2008 Oct 1;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
30. Spitale G, Germani F, Biller-Andorno N. TopicTracker—an advanced software pipeline for text mining on PubMed data: bridging the gap between off-the-shelf tools and code based approaches. Heliyon 2024 Sep 15;10(17):e36351. [doi: 10.1016/j.heliyon.2024.e36351] [Medline: 39281629]
31. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. University of Texas Libraries. 2015 Sep 15. URL: http://hdl.handle.net/2152/31333 [accessed 2025-04-07]
32. Aranda AM, Sele K, Etchanchu H, Guyt JY, Vaara E. From big data to rich theory: integrating critical discourse analysis with structural topic modeling. Eur Manag Rev 2021 Sep;18(3):197-214. [doi: 10.1111/emre.12452]
33. Abram MD, Mancini KT, Parker RD. Methods to integrate natural language processing into qualitative research. Int J Qual Methods 2020 Jan 1;19:1609406920984608. [doi: 10.1177/1609406920984608]
34. Arora SD, Singh GP, Chakraborty A, Maity M. Polarization and social media: a systematic review and research agenda. Technol Forecast Soc Change 2022 Oct;183:121942. [doi: 10.1016/j.techfore.2022.121942]
35. Jiang J, Ren X, Ferrara E. Social media polarization and echo chambers in the context of COVID-19: case study. JMIRx Med 2021;2(3):e29570. [doi: 10.2196/29570] [Medline: 34459833]
36. Kubin E, von Sikorski C. The role of (social) media in political polarization: a systematic review. Ann Int Commun Assoc 2021 Jul 3;45(3):188-206. [doi: 10.1080/23808985.2021.1976070]
37. Tanesini A. Affective polarisation and emotional distortions on social media. Roy Inst Philos Suppl 2022 Oct;92:87-109. [doi: 10.1017/S1358246122000261]
38. Meta backs away from fact checking in the U.S. NPR. 2025. URL: https://www.npr.org/2025/01/12/nx-s1-5252739/meta-backs-away-from-fact-checking-in-the-u-s [accessed 2025-01-15]
39. Redaelli S, Biller-Andorno N, Gloeckler S, Brown J, Spitale G, Germani F. Mastering critical thinking skills is strongly associated with the ability to recognize fakeness and misinformation. Front Educ 2025;10:1577692. [doi: 10.3389/feduc.2025.1577692]

Abbreviations

NLP: natural language processing

XSL•FO
RenderX

Edited by T Purnat; submitted 23.Apr.2025; peer-reviewed by L Hagen, R Scardigno; revised version received 23.Oct.2025; accepted 15.Jan.2026; published 30.Mar.2026.

Please cite as:
Germani F, Spitale G, Fritschi F, Merten S, Biller-Andorno N
Analyzing Misinformation and Disinformation: Understanding Swiss COVID-19 Narratives Through Natural Language Processing Analysis
JMIR Infodemiology 2026;6:e76441
URL: https://infodemiology.jmir.org/2026/1/e76441
doi:10.2196/76441

© Federico Germani, Giovanni Spitale, Franc Fritschi, Sonja Merten, Nikola Biller-Andorno. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 30.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Misinformation Through the Lens of Dutch News Media and Public Health Authorities: Content Analysis of Topics and Narratives Present in Pandemic-Related Misinformation During COVID-19

David Blanco-Herrero, PhD; Bas van den Putte, PhD; Toni G L A van der Meer, PhD

Department of Communication Science, Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam, North Holland, The Netherlands

Corresponding Author:
David Blanco-Herrero, PhD
Department of Communication Science, Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam, North Holland, The Netherlands

Abstract

Background: One of the multiple challenges associated with misinformation is its risk of hindering public health efforts during a health crisis. Misinformation can cover a diversity of thematic topics and is usually spread using a series of nontopic-specific recurring narrative structures. Identifying these 2 misinformation features helps to understand underlying structures and patterns through which misleading information can take shape or be circulated.

Objective: The goal of the study is to obtain a complete overview of the topics and narratives of pandemic-related misinformation cases addressed by 2 key information providers during the COVID-19 pandemic in the Netherlands: news media and public health authorities.

Methods: A quantitative content analysis was conducted on 698 news articles and 1333 tweets published between January 2020 and May 2023.

Results: The study revealed that vaccine-related misinformation, both in terms of topics and narratives, was the most frequently addressed type of misinformation across both sources. The topics and narratives evolved along the pandemic, with the virus origin and danger being more predominant in the misinformation addressed in the early phases and the vaccines gaining presence around 2021.

Conclusions: In total, the study revealed the dominance of vaccine-related narratives as well as the change of misinformation features over time. The deeper understanding of misinformation topics and narratives can inform communication strategies to counter its effects, focusing on the appropriate topics and narrative structures in different moments.

(JMIR Infodemiology 2026;6:e91743) doi:10.2196/91743

KEYWORDS

COVID-19 pandemic; public health misinformation; vaccines; content analysis; agenda-setting

Introduction

Background

In a health crisis, misinformation poses a threat that increases the danger of the health emergency. During the COVID-19 pandemic, misinformation affected individual responses to the crisis [1-3], which hindered the efforts to counter the outbreak. Studies reviewing social media content showed that misinformation about vaccine safety and effectiveness erodes confidence in vaccines and decreases intentions to be vaccinated [4,5]. For example, variation in level of misinformation between US states predicted a decrease of 20% vaccine uptake between states [6]. Beyond vaccination, belief in COVID-19 misinformation is associated with reduced adherence to preventive behaviors such as mask-wearing, physical distancing, and hand hygiene, particularly among individuals with lower trust in institutions [7]. More broadly, widespread misinformation about the virus, vaccines, and public health measures was also observed to undermine trust in authorities and support for evidence-based public health policies [8].

To fully understand the implications of misinformation during this period, it is necessary to go beyond assessing its accuracy or effects and also describe its content. For instance, in their report for the European Parliament, Jacob et al [9] identified key COVID-19 misinformation themes in European countries, such as vaccination and the disease severity. Although identifying misinformation topics is a common approach due to its importance for agenda-setting [10], the study of specific narratives or conspiracy theories that articulate these topics are less frequent (though not completely absent; see, for instance, Kotseva et al [11]).

XSL•FO
RenderX

Misinformation topics are conceptualized as broad and neutral issues, while narratives are specific storyline structures. Topics reflect content areas around which misinformation can circulate or not and can be neutral (eg, vaccines), while narratives follow recurring storyline structures that frame elements of one or more topics in a particular direction (eg, vaccines were intentionally created to harm people). Thus, topics focus on *what* the misinformation is about, whereas narratives are the stories people construct around those topics, showing *how* events are connected and *why* the topic matters to an audience. Using both concepts allows for a more complete understanding of misinformation content, capturing not only the main topics but also the more specific recurring narratives used to spread misinformation, a dimension that has received less attention than the study of topics in previous literature.

To obtain a broad view of misinformation in a given context, it is also important to include more than one type of source of information, unlike most prior studies. Because news media shape public opinion and set the public agenda [12], examining their coverage of misinformation remains relevant, as seen, for instance, in Lwin et al [13] regarding COVID-19 misinformation in Singapore. However, other information providers also play a role. Especially during a health crisis, public health authorities are central in countering the effects of misinformation, as they have direct channels to reach citizens [14,15]. Besides their mediated presence in news media or press conferences, their social media accounts serve as more direct communication channels.

A study on the use of Twitter by public health agencies during the early stages of the pandemic [16] found limited attention to correcting misinformation, perhaps because only the first months were analyzed, when misinformation was insufficiently understood and not enough was known about how to correct it. The authors recommended these organizations "strengthen their efforts to combat misinformation and providing more insights into research." In a similar line, Rasmusen et al [17] found that interventions from public health authorities effectively mitigated the spread of misinformation on social media. Thus, even if addressing misinformation is not always a central target of their communication strategies, public health authorities are well positioned to do so, especially considering that they were the most commonly chosen source by social media users seeking information during the pandemic [18]. However, analyses of their specific approach to misinformation remain scarce.

Therefore, this study introduced 2 main contributions: First, it complemented the analysis of misinformation topics with the analysis of recurrent narratives, and second, it combined 2 types of information sources that addressed pandemic-related misinformation, namely online and print news media and tweets from public health organizations, increasing the diversity of observable strategies and consequently the study's external validity. Although other media types or social media platforms could have been examined, the predominantly textual nature of these 2 sources allows for a coherent analysis strategy. Additionally, the study covered the full duration of the pandemic, offering a broader view of the evolution of misinformation features than previous studies with narrower time frames.

In total, the main goal of our study was to identify the topics and narratives of pandemic-related misinformation cases addressed by news media and public health authorities in the Netherlands during the COVID-19 pandemic. The objective was not to analyze all the misinformation circulating this period but rather the misinformation considered most important by 2 relevant and diverse sources. Because it is unfeasible to accurately capture all false claims and rumors spreading across formats and platforms, analyzing misinformation addressed by relevant sources offers a valid approach, as seen in the studies by Brennen et al [19] and Ceron et al [20], who focused on the content debunked by fact-checking organizations. These approaches not only illustrate the main topic and structure features of misinformation but also reveal the preferences of the studied sources.

This way, we went beyond previous literature that largely sought to map the themes and narratives circulating among the general public, particularly on social media (eg, [11]), with comparatively less attention to how institutional actors responded to misinformation. By examining the topics and narrative structures of misinformation as they were addressed in journalistic coverage and in the communication of health agencies, we shifted the focus from the spread of misinformation to which misinformation is publicly contested and countered. This perspective foregrounds the agenda-setting and framing role of corrective communication and helps identify which misinformation themes receive attention and when.

In our study, we did not attempt to establish what qualifies as misinformation but rather understand what news media and public health authorities perceive as such. Thus, misinformation was conceptualized as any claim that the studied news articles or tweets addressed as such, without further exploring whether the alleged misinformation cases addressed by the studied actors actually entailed misinformation as defined by scientific definitions (ie, a verifiably inaccurate or misleading assertion about COVID-19, its prevention, or vaccination that conflicted with the best available scientific evidence or official guidance at the time [21]) or whether that also includes normative statements (eg, whether measures were "too strict" or "unfair") and disagreements over policy preferences without a falsifiable factual component.

Before delving more specifically into the literature that underpinned this research, it is relevant to briefly contextualize the Dutch case, where the study was conducted. In the Netherlands, the government response to the COVID-19 pandemic combined voluntary compliance, personal responsibility, and fluctuating proportional restrictions [22,23]. Throughout the pandemic, Prime Minister Mark Rutte and successive health ministers held regular televised press conferences to explain measures and justify policy choices, with the National Institute for Public Health and the Environment (RIVM) providing weekly epidemiological updates and guidance. The Dutch media environment during COVID-19 was characterized by high consumption by Dutch citizens of traditional news outlets and generally high, although fluctuating, levels of trust in journalism and governmental information [24,25]. There was, however, space for contestation and skepticism and the spread of conspiracy narratives, especially

XSL•FO
RenderX

on social media [26], something that even fostered social movements [27]. Thus, the Netherlands combined a relatively trusted, expert-oriented news system, in which the institutional response to the crisis coexisted with contestation that was partly fueled by misinformation and conspiracy theories, making it a relevant case to examine how mainstream news media and public health authorities addressed pandemic-related misinformation.

Misinformation Topics and Recurring Narratives During the COVID-19 Pandemic

Following the longstanding line of work on agenda-setting [12] and considering as sources both news media and public health authorities, it was relevant to identify the topics that were more predominantly present in misinformation instances, as false information also has the capacity to set the agenda [28]. Previous examples are a report by Bruns et al [29], whose review of empirical evidence from social and behavioral sciences showed that COVID-19 misinformation did not focus only on health aspects but also included misinformation about political responses to the pandemic or the origins of the virus and its severity. Focusing exclusively on the first months of the crisis and using the misinformation corrected by fact-checkers as a reference, Brennen et al [19] and Salaverría et al [30] also studied the main topics present in misinformation cases. Building upon these studies, our study combined a timeframe covering the whole duration of the declaration of pandemic by the World Health Organization (WHO) while also including 2 diverse sources. We asked 3 research questions (RQs).

RQ1 asked "What were the main topics present in the pandemic-related misinformation addressed by printed and online news media as well as Twitter accounts of public health authorities in the Netherlands during the COVID-19 pandemic?"

As an important novelty, this study went beyond topics, trying to also identify misinformation narratives. Topics are the *subject matter* or domain around which misinformation clusters, although they could also be used to structure types of content other than misinformation (eg, vaccination can be a topic around which misinformation spreads but also a broader topic of health information). Narratives, by contrast, go beyond topical labels and relate to recurring storylines and argumentative logics that give meaning and structure to misinformation, with ideas being linked causally and temporally, not just thematically (eg, vaccines are created by an elite to cause harm). Although narratives can differ in their level of complexity, they often include a "cause-and-effect relationships between events that take place over a particular period that impact particular characters" [31].

This narrative dimension plays a distinctive role in the dynamics of misinformation, as it allows rhetorical strategies or constructions that go beyond merely misstated facts. In fact, misinformation has been frequently observed to repeat itself or evolve around specific stories and tropes [32]. Part of the successful spread of misinformation is explained by its capacity to exploit recurring narratives, making use of familiarity biases and the illusory truth effect, which is the tendency to believe that (false) information is correct after repeated exposure [33,34].

The study of misinformation narratives during the COVID-19 pandemic has been tackled in a few studies. For example, Kotseva et al [11] identified 12 super-narratives that were repeatedly observed: Vaccine-related narratives, Criticism restrictions, Conspiracy theories, Criticism of EU and inter/national actors, Claims of authoritarianism and dystopia, Geopolitics, Health-related narratives, Fearmongering, Downplaying COVID-19, Anti-minority narratives, Distrust toward media, and Other. Although with a broader operationalization than the one adopted in our study and collected exclusively on news articles, these super-narratives offer one of the most comprehensive perspectives to date to understand misinformation beyond its topics; they were therefore used to inform our research design and instruments.

However, the investigations addressing misinformation narratives frequently used shorter timeframes that did not cover the whole duration of the pandemic and did not include a diversity of sources for the analysis, especially by looking at the content of public health authorities. To provide a broader overview, we posed the following question (RQ2): "What were the main recurrent narratives present in the pandemic-related misinformation addressed by printed and online news media as well as Twitter accounts of public health authorities in the Netherlands during the COVID-19 pandemic?"

Misinformation topics and narratives were not static. Both Jacob et al [9] and Kotseva et al [11] observed that the presence of misinformation topics and specific narratives changed over time during the pandemic. Further comprehending the time evolution of misinformation features is relevant to understand how they might have been affected not only by the volume of misinformation circulating at a given time but also by additional factors, such as the roles of the sources or the shifting societal goals during the pandemic [35]. Once again, previous research is limited in its partial covering of the pandemic's duration, and analyses often do not combine the study of diverse misinformation features (ie, topics and narratives) with diverse sources (ie, news media and public health authorities). Therefore, with the goal of studying how the topics and narratives present in the misinformation addressed by news media and public health authorities changed over the duration of the pandemic, we posed the last research question (RQ3): "How did the main (1) topics and (2) recurrent narratives present in pandemic-related misinformation addressed by printed and online news media as well as Twitter accounts of public health authorities in the Netherlands change along the COVID-19 pandemic?"

Methods

Procedure

This research was based on a quantitative content analysis of national news media and Twitter posts of public health authorities in the Netherlands. Although the COVID-19 pandemic and the spread of misinformation were global phenomena, the analysis of misinformation features from a national perspective is justified by the possibility of using comparable sources for the analysis as well as by the response

XSL•FO
RenderX

to the health crisis, which was mostly determined by national governments and differed from country to country.

The period of analysis comprised from January 1, 2020 (because on January 24, 2020, the first news items about COVID-19 were published in the Netherlands), to May 5, 2023 (day on which the WHO officially ended the pandemic declaration). Given the different phases within the pandemic and in order to answer RQ3, we followed the process by Solovei et al [24] to determine the phases of the COVID-19 pandemic in the Netherlands. These phases were defined based on the waves of infections and hospitalizations [23] and the stringency index developed by Hale et al [36] to reflect the response measures enacted by governments during the pandemic. The 6 phases used in this study were (1) the dawn of the pandemic and the first wave (January 2020-May 2020), (2) the first interwave period of relief (June 2020-August 2020), (3) the second and third waves (September 2020-December 2020 and March 2021-April 2021) with no relaxation of measures in between, (4) the second period of relief (May 2021-September 2021), (5) the fourth wave (October 2021-December 2021), and (6) the final relief period until the end of the pandemic declaration by the WHO (January 2022-May 2023).

Content was collected from national newspapers (both printed and online) and from the Twitter accounts of 5 regional and national public health organizations. Although Twitter was not the most used social media platform in the Netherlands during the period of study—around 20% of the Dutch population were Twitter users according to data from the National Social Media Study [37,38]—the choice of this platform for the analysis is explained by its open and accessible nature as well as by its role as a central stage for public debate and the provision of information by public institutions [39], particularly during the COVID-19 health crisis [40]. This approach was also used by previous studies addressing Twitter communication of public health authorities during the pandemic [41]. It should be noted that the Twitter platform changed its name to X in July 2023; the original name has been consistently used across the manuscript, since that was the name of the platform during the entirety of the studied period.

Although data collection and analysis of news media and Twitter occurred in parallel, the process, materials, and instruments were channel-specific and therefore slightly different for both types of data. In both cases, the strategy was driven by the communication patterns of the sources: Given the greater amount of information, search strings were used for news media, something that was not necessary for the Twitter posts, which could be filtered manually. Regarding the news media, Nexis Uni was used to identify articles dealing with pandemic-related misinformation in the 8 main Dutch national newspapers. Given the limited presence of online-only news media in this database, a complementary search was conducted on Google News for the 2 main digital news platforms in the Netherlands: NU.nl and NOS.nl.

The search string used in these cases was [(nepnieuws* OR misinformatie* OR desinformatie* OR fake OR “valse informatie”) AND (covid* OR corona*)]. This translates into English as [(“fake news” OR misinformation OR disinformation OR fake OR “false information”) AND (covid OR corona)]. The use of wildcards is explained by the possibility of creating words in Dutch by combining terms, such as “desinformatiecampagne,” meaning “disinformation campaign.“ This reflects the study’s focus on news articles in which misinformation was explicitly constructed as such and not necessarily those in which false claims or COVID-19 were discussed implicitly. This search led to 2151 cases, which were subject to a preliminary analysis to identify cases in which pandemic-related misinformation was sufficiently present. This additional screening ensured that only articles where pandemic-related misinformation was a relevant part of the story were retained, removing cases where it was mentioned only as background information (eg, political misinformation during the 2020 US presidential elections where the pandemic was merely contextual). The final sample comprised 698 cases in which pandemic-related misinformation was central to the story. All cases were in Dutch and included news and opinion articles, excluding other formats such as quizzes and weather forecasts.

Regarding the Twitter posts of public health authorities, 5 institutions were selected: GGD (Gemeentelijke Gezondheidsdienst) Amsterdam, GGD Rotterdam-Rijnmond, GGD Utrecht Regio, the RIVM, and the Ministry of Health. The first 3 are regional or local institutions in charge of the implementation of emergency measures during the pandemic: Those were the organizations that led the testing and vaccination process in their region and produced much of the communication that reached their local communities. The 3 selected accounts are those with the largest number of followers and publications. The last 2 are the national institutions in charge of the design of health measures and strategies (eg, vaccination, testing) that the different GGDs would implement. Our selection combines the communication of larger national organizations with the communication of more local and hands-on ones.

For a Twitter post to be selected, it was not enough to provide information on a pandemic-related topic (eg, where the vaccination centers are); it had to convey information that would counter false rumors or help clarify doubts or fears among the population. An example can be found in the post from the Health Ministry on January 19, 2022 (Original Dutch text: “Vragg jij je ook af of je niet beter kunt wachten op een vaccin tegen de omikronvariant? Immunoloog Marjolein van Egmond van het @amsterdamumc, raad dat niet aan. Ze legt uit waarom.”): “Are you also wondering if it wouldn’t be better to wait for a vaccine against the omicron variant? Immunologist Marjolein van Egmond of the @amsterdamumc advises against it. She explains why.” This was followed by a video of this immunologist, in which those concerns were disputed. Although the post did not directly mention misinformation, it alludes to and tries to counter fears and doubts that can be supported by misinformation.

Two trained coders independently analyzed the timeline of each of the 5 accounts during the studied period, searching for posts tackling pandemic-related misinformation or providing information that could be used to counter misinformation. An exploratory observation of the activity of the public health authorities on their social media showed that they do not usually correct misinformation explicitly, but they rather provide

XSL•FO
RenderX

information that can be used to counter the effects of misinformation. This led us to adopt a wider conceptualization of the ways of dealing with misinformation, rather than limiting the search to explicit mentions to misinformation. To safeguard reliability, relevant cases were selected only when both coders agreed on their adequacy for the goals of the study, thus ensuring that the selected cases were indeed referring to misinformation. This led to a total of 1333 cases.

Content from both news media and Twitter accounts was analyzed using the same codebook, which was slightly adapted to fit the particulars of both types of content (see Multimedia Appendices 1 and 2). The analysis was preceded by a training process as well as a series of intercoder reliability (ICR) tests to ensure the reliability of the coding. For the news articles and Twitter posts, 95 of the 698 analyzed cases (13.6%) and 135 of the 1333 analyzed cases (10.1%), respectively, were double-coded. The results of these tests are shown in Table 1. Because all the variables were dichotomous and many were highly skewed, we used prevalence-adjusted and bias-adjusted kappa (PABAK) as our primary reliability coefficient, which is less sensitive to prevalence and bias in such situations [42]. Although Krippendorff alpha is widely regarded as a standard reliability index in content analysis, particularly for multicategory nominal data and scale construction [43], given our data structure, the use of Krippendorff alpha was expected to underestimate agreement for rare categories. Therefore, PABAK provided a more informative summary of coder agreement. On average, the ICR values reached a satisfactory 0.89, with only 2 values offering less than substantial (<0.60) agreement; in these 2 cases, caution in the interpretation of the results is recommended.

XSL•FO
RenderX

Table . Intercoder reliability measures for each variable.

Variable	News outlets, PABAK[a]	Public health authorities, PABAK
Topic (Presence)		
Misinformation on the origin of the virus	0.90	1.00
Misinformation on the danger and spread of the virus	0.80	0.66
Misinformation on vaccines	0.85	0.88
Misinformation on pharmaceutical measures, other than vaccines	0.90	0.69
Misinformation on nonpharmaceutical measures	0.77	0.87
Misinformation on political issues related to the pandemic	0.87	1.00
Misinformation on economic issues related to the pandemic	0.98	1.00
Misinformation on social issues related to the pandemic	1.00	0.90
Misinformation on technological issues related to the pandemic	0.90	0.99
Misinformation on other topics related to the pandemic	0.98	0.97
Narrative (Presence)		
The virus has been intentionally created	0.71	1.00
The virus does not exist or is less harmful than pretended	0.75	0.67
Lockdowns or other measures are used to control the population or to limit their freedom	0.90	0.97
Useless or dangerous remedies are effective	0.96	0.99
Vaccines do not work or are not sufficiently tested	0.94	0.56
Vaccines are intentionally harmful or evil	0.92	0.97
The pandemic is part of larger scheme or conspiracy theory	0.83	1.00
Bill Gates is involved in some way with the virus, the measures, or the vaccines	1.00	1.00
George Soros is involved in some way with the virus, the measures, or the vaccines	1.00	1.00
Other recurring narratives	1.00	0.32

[a]PABAK: prevalence-adjusted and bias-adjusted kappa.

Variables

The codebook started with a series of descriptives, followed by the main variables related to the features of misinformation. The descriptive variables included the date of publication, the source, and, for the Twitter posts, the type (original or a retweet). For the study of misinformation topics and narratives, a hybrid deductive-inductive coding strategy was adopted. Following previous studies [9,11,19,29] as well as a set of interviews with information intermediaries who had tackled misinformation during the COVID-19 pandemic in the Netherlands [44], 9 topics and 9 narratives were coded. As a complement, open categories were included to capture additional topics and narratives. The final variables were topics and narratives.

For topics, the presence (1) or absence (0) of the following pandemic-related misinformation topics were coded (more than one option could be selected): Origin of the virus; Danger and spread of the virus; Vaccines; Pharmaceutical measures, other than vaccines; Nonpharmaceutical measures; Political issues related to the pandemic; Economic issues related to the pandemic; Social issues related to the pandemic; Technological issues related to the pandemic; and Other topics (including a follow-up open question).

XSL•FO
RenderX

For narratives, the presence (1) or absence (0) of the following pandemic-related narratives were coded (more than one option could be selected): The virus has been intentionally created; The virus does not exist or is less harmful than pretended; Lockdowns or other measures are used to control the population or to limit their freedom; Useless or dangerous remedies are effective; Vaccines do not work or are not sufficiently tested; Vaccines are intentionally harmful or evil; The pandemic is part of a larger scheme or conspiracy; Bill Gates is involved in some way with the virus, the measures, or the vaccines; George Soros is involved in some way with the virus, the measures or the vaccines; and Others (including a follow-up open question).

Ethical Considerations

This research was conducted after preregistration [45] and ethical approval from the Ethics Review Board of the Faculty of Social and Behavioral Sciences of the University of Amsterdam (number FMG-9528).

Results

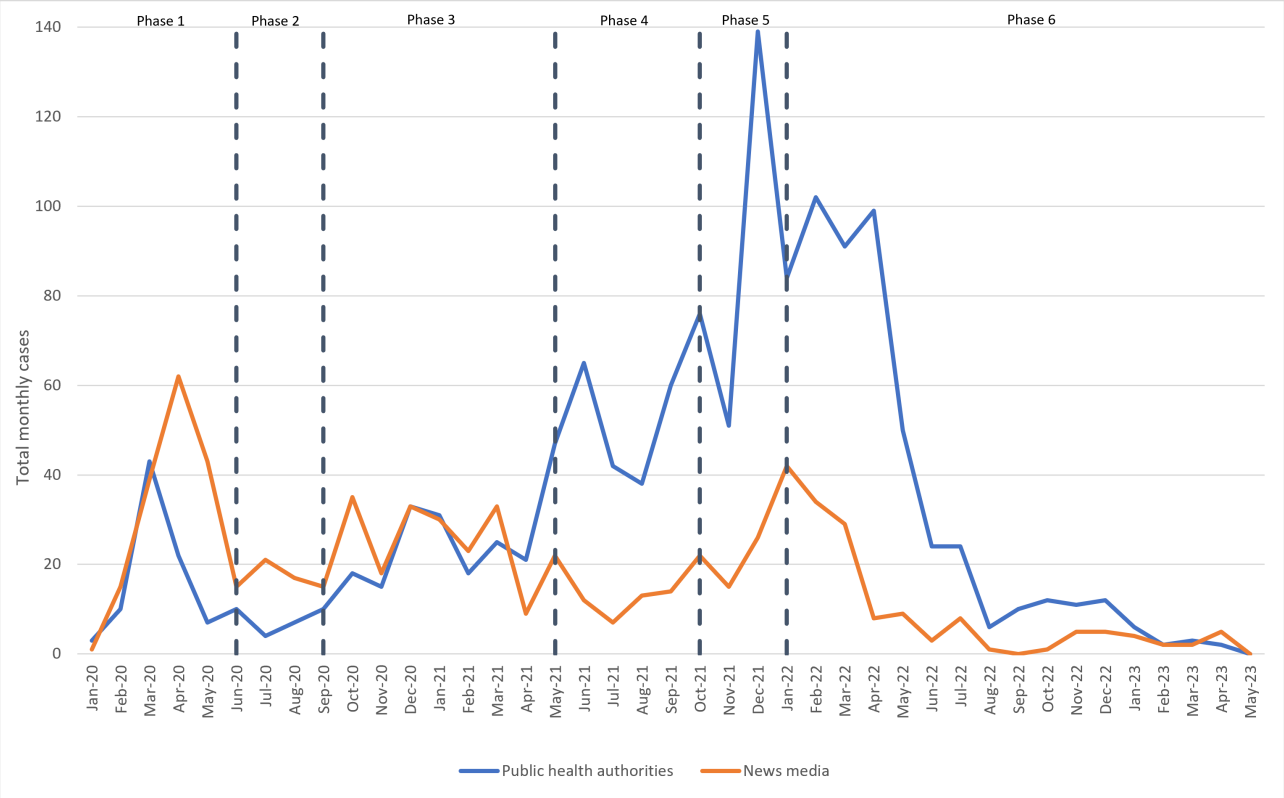
Sample Description

The final sample included 698 news articles and 1333 Twitter posts from public health authorities that had dealt with misinformation during the COVID-19 pandemic. The subsample of news articles was dominated by printed newspaper articles (650/698, 93.1%), with only 48 articles from online only newspapers (48/698, 6.9%). The printed newspaper *De Volkskrant* had the most articles in the sample (144/698, 20.6%), and the online only NU.nl had the smallest presence (18/698, 2.6%). Articles were on average 892.15 (SD 765.43) words long, with the shortest article containing 34 words and the longest containing 7318 words.

For the Twitter posts, the 1333 cases were distributed more unevenly, with a clear predominance of GGD in the Utrecht region (793/1333, 59.5%). Given its predominance in the sample, we tested whether GGD Utrecht's posts showed specific patterns that could affect the overall interpretations, but no relevant differences were observed. In contrast, GGD Amsterdam had only 43 (43/1333, 3.2%) posts in the sample. Most of the posts were original publications (1226/1333, 92%), while the others (107/1333, 8%) were retweets or mentions of existing posts, frequently from the same account or another public health institution.

Figure 1 shows the monthly distribution of the number of cases per dataset during the period analyzed. Although some shared patterns can be identified, the peaks in each of the 2 subsamples took place in different moments. The highest number of news articles was published during the first phase, when the first wave of the pandemic surprised the Netherlands and other Western countries. For the public health authorities, the most active period included the months at the end of year 2021 and the first months of 2022, during which the Netherlands experienced an important increase in the number of infections while the vaccination campaign was also ongoing.

Figure 1. Time distribution of the analyzed cases.



Topics and Narratives

To reply to RQ1, vaccination was the most frequent misinformation topic among both the news articles and the Twitter posts (Table 2). The topics were more diverse in the news articles; besides vaccination, the most frequent topics related to the spread, danger, and origin of the virus and to pharmaceutical measures other than vaccines. In the Twitter

XSL•FO
RenderX

accounts of public health authorities, more than one-half (817/1333, 61.3%) of the posts concentrated on vaccination, followed by pharmaceutical and nonpharmaceutical measures to counter the pandemic.

Table . Presence of misinformation topics in the sample.

Topics	News articles (n=698), n (%)	Twitter posts (n=1333), n (%)
Misinformation on vaccines (eg, vaccines include microchips)	222 (31.8)	817 (61.3)
Misinformation on the danger and spread of the virus (eg, the virus is not real)	146 (20.9)	111 (8.3)
Misinformation on the origin of the virus (eg, the virus was intentionally created)	141 (20.2)	1 (0.1)
Misinformation on pharmaceutical measures, other than vaccines (eg, hydroxychloroquine kills the virus)	131 (18.8)	342 (25.7)
Misinformation on technological issues related to the pandemic (eg, the COVID-19 app is used to track citizens)	47 (6.7)	25 (1.9)
Misinformation on nonpharmaceutical measures (eg, the lockdown is used to control the population)	40 (5.7)	130 (9.8)
Misinformation on social issues related to the pandemic (eg, immigrants are not respecting the restrictions)	17 (2.4)	13 (1)
Misinformation on political issues related to the pandemic (eg, alleged declarations of politicians)	16 (2.3)	1 (0.1)
Misinformation on economic issues related to the pandemic (eg, the pandemic is destroying a particular economic field)	3 (0.4)	0 (0)
Other topics	0 (0)	0 (0)
No topic specified	205 (29.4)	126 (9.5)

In a significant number of cases, no specific topic was identified, especially in the news articles: This was usually the case with messages that addressed false information in the context of the COVID-19 pandemic but did so as a general threat or trying to warn or educate people on the phenomenon of misinformation without connecting it to a specific topic.

Regarding the more specific narratives explored in RQ2, different patterns were observed in both types of sources, as shown in Table 3. Among news articles, the intentional creation of the virus or its lack of danger were the most repeated narratives, together with useless or dangerous remedies being effective to fight the virus. Interestingly, a more detailed analysis of the cases coded as “Other” narratives showed that a relevant and recurring narrative was the alleged connection between 5G mobile technology and COVID-19.

XSL•FO
RenderX

Table . Presence of misinformation narratives in the sample.

Recurrent narratives	News articles (n=698), n (%)	Twitter posts (n=1333), n (%)
The virus has been intentionally created	116 (16.6)	0 (0)
The virus does not exist or is less harmful that pretended	115 (16.5)	40 (3)
Useless or dangerous remedies are effective	108 (15.5)	9 (0.7)
Bill Gates is involved in some way with the virus, the measures, or the vaccines	62 (8.9)	0 (0)
Vaccines are intentionally harmful or evil	58 (8.3)	5 (0.4)
The pandemic is part of larger scheme or conspiracy	47 (6.7)	0 (0)
Vaccines do not work or are not sufficiently tested	37 (5.3)	208 (15.6)
Lockdowns or other measures are used to control the population or to limit their freedom	29 (4.2)	12 (0.9)
George Soros is involved in some way with the virus, the measures, or the vaccines	11 (1.6)	0 (0)
Other narratives		
All other narratives	15 (2.1)	133 (10)
Related to the connection between 5G and the pandemic	15 (2.1)	0 (0)
Related to the vaccination process (locations, the relationship of vaccination to other measures, or the need for booster vaccines)	0 (0)	42 (3.1)
Related to testing (compulsoriness, locations, relationship to other measures)	0 (0)	40 (3)
Related to the implementation of measures (nonpharmaceutical interventions being in place or not at a given time)	0 (0)	35 (2.6)
Related to the activity of the public health authorities themselves	0 (0)	10 (0.8)
Related to warnings about phishing and other cybercrimes	0 (0)	6 (0.5)
No recurring narrative specified	330 (47.3)	966 (72.5)

In the case of the Twitter posts, recurring narratives were clearly focused on vaccines. The category of “Other” narratives was quite frequently present, usually due to more specific narratives within the predesigned recurring stories. Some of those narratives related to testing (mostly, the lack of effectiveness of the tests) or to the vaccination process (eg, claims about not needing booster doses of the vaccine; this vaccination process category could be considered a particular narrative connected to those of vaccines not working or being harmful). Other recurring themes identified during the study related to the activity of the studied public health authorities (eg, the workers appearing unexpectedly in houses) or to warnings about scams or phishing in the context of the pandemic. Given the relevance and abundance of these categories, they were added to the predesigned codes for a more detailed observation, as shown in Table 3.

In 47.3% (330/698) of the news articles and 72.5% (966/1333) of the Twitter posts, no recurring narrative was explicitly mentioned. This is significantly higher than the one observed in relation to the topics. This is reasonable, given the higher specificity of narratives and because misinformation could circulate around a diversity of topics that were not always connected to recurring narratives.

XSL•FO
RenderX

Given the connection between topics and narratives, it is no coincidence that vaccine-related narratives were predominant while vaccines were also dominant as a topic. To test that relation between topics and narratives, complementary analyses were conducted, observing the correlation between the presence of topics and narratives. The significant correlations (P<.001) are available in Multimedia Appendix 3.

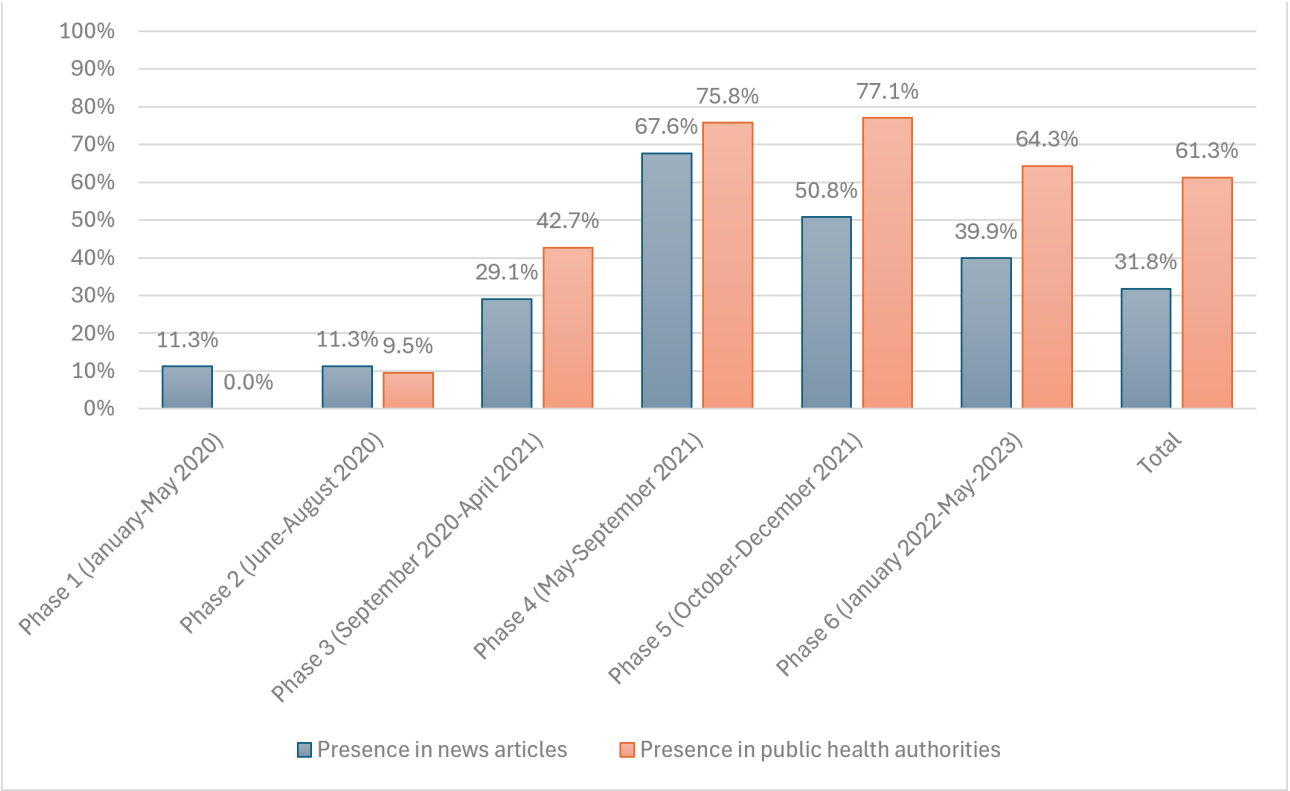
For the news articles, the biggest effect sizes were found in pairs of variables that would be expected to be related: The topic on the origin of the virus had a strong positive correlation with the narrative on the intentional creation of the virus (r_{696}=0.849, P<.001); the topic on the danger and spread of the virus was similarly correlated with the narrative of the virus not existing or being less harmful than pretended by the official discourse (r_{696}=0.826, P<.001); and the topic on pharmaceutical measures other than vaccines had a strong positive correlation with the narrative on the effectiveness of useless or dangerous remedies (r_{696}=0.839, P<.001). It was also interesting to observe some cross-cutting narratives, like the involvement of Bill Gates on the pandemic or the pandemic being part of a larger scheme, that were positively correlated with a diversity of other topics and recurring narratives.

The strongest correlations within the Twitter posts were between the presence of misinformation on technological issues related to the pandemic and warning about phishing and other cybercrimes (r_{1331}=0.486, P<.001) and between the presence of misinformation on social issues related to the pandemic and the recurring narratives related to the activity of the public health authorities themselves (r_{1331}=0.434, P<.001). Contrary to what was observed in the news articles, in which all significant correlations were positive, among the Twitters posts, there were several negative correlations. The strongest was between the presence of misinformation on vaccines and the presence of misinformation on pharmaceutical measures other than vaccines (r_{1331}=−0.418, P<.001).

Temporary Evolution of Topics and Narratives Present in Pandemic-Related Misinformation

The third RQ asked about the potential differences in the presence that the studied topics and narratives had during the different pandemic phases. To study this, χ^2 tests were performed, comparing the mean presence of each topic and narrative over the 6 phases. In general terms, patterns in both types of sources were similar, and most variables showed statistically significant differences across phases. This means that the misinformation content that news media and Twitter accounts of public health authorities decided to deal with varied significantly over the pandemic. Some relevant patterns were the higher presence of misinformation related to the danger and spread of the virus, to its origin, and to the different measures to counter it (including the specific narratives related to these topics) at proportionally higher rates in the first phases, when there was less certainty and more rumors about these issues. In contrast, misinformation on the vaccines increased with time. As an example, Figures 2-4 show the distribution across the topics and narratives with the highest effect size (ie, Cramer V) and therefore a strongest change across different phases: the topic of misinformation on vaccines, the topic of misinformation on the danger and spread of the virus, and the recurring narrative related to the vaccination process that emerged during the analysis (only for the Twitter posts). The figures showing the evolution of the presence of the rest of variables as well as the detailed results of the χ^2 tests are available in Multimedia Appendix 4.

Figure 2. Percentage of studied cases addressing misinformation on vaccines in the 2 studied types of sources across the different phases of the pandemic.



XSL•FO
RenderX

Figure 3. Percentage of studied cases addressing misinformation on the danger and spread of the virus in the 2 studied types of sources across the different phases of the pandemic.

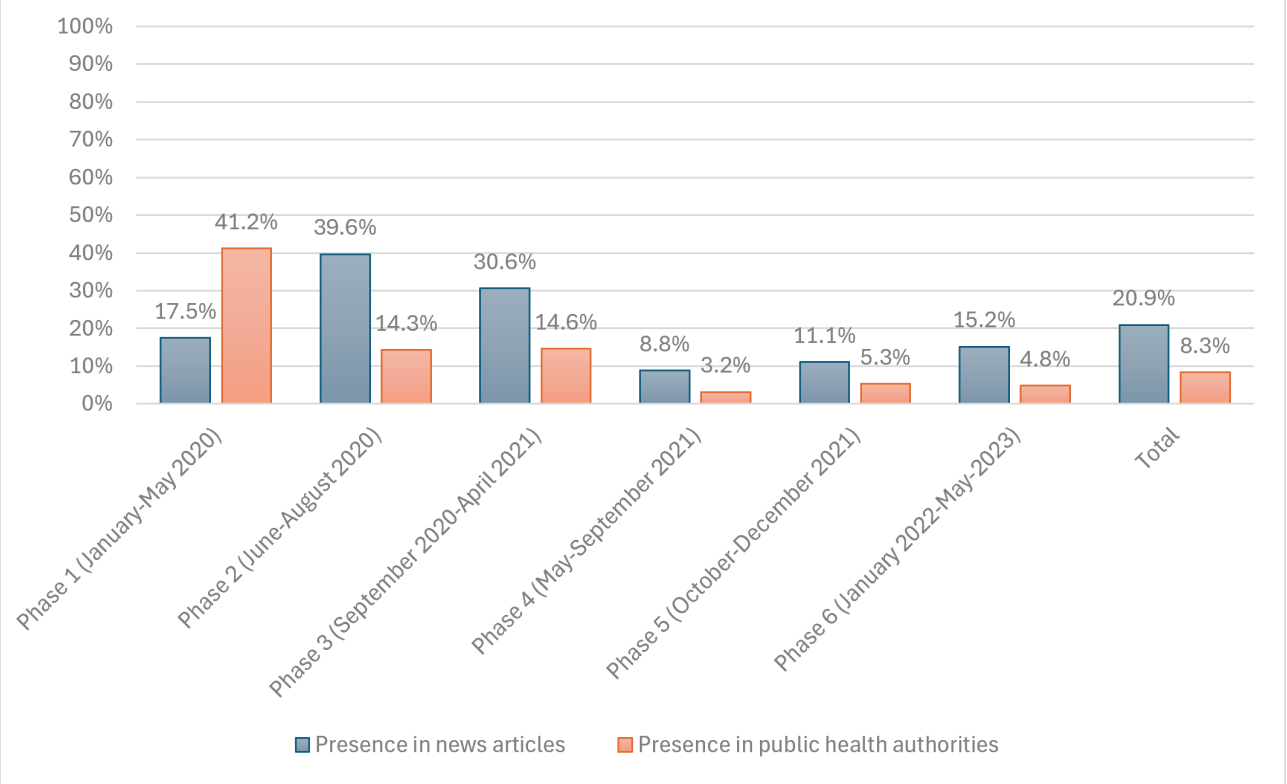
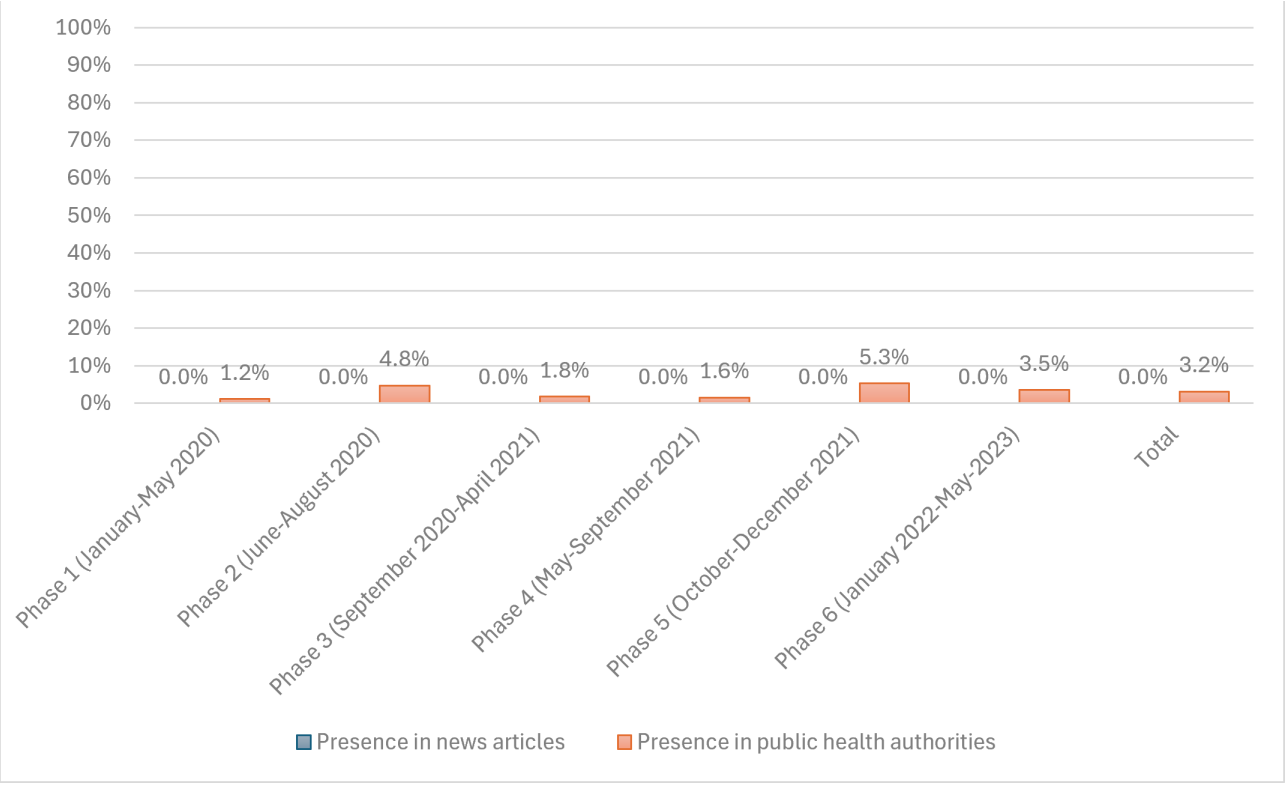


Figure 4. Percentage of studied cases addressing misinformation on the recurring narrative related to the vaccination process in the 2 studied types of sources across the different phases of the pandemic.



The 2 misinformation topics presented in Figures 2 and 3 had strong changes across the pandemic, albeit in different directions. Vaccine misinformation gained presence as the pandemic advanced, peaking in phases 4 and 5 during the second half of 2021 when the vaccination campaign was in full swing in the Netherlands. This is also the case for the recurring narratives on the vaccines not working or being insufficiently tested and the vaccines being intentionally harmful (Figure 2).

XSL•FO
RenderX

In contrast, misinformation on the danger and spread of the virus was more present in the beginning of the pandemic, and it became less relevant in the last phases (Figure 3).

Figure 4, which focuses on misinformation narratives related to the vaccination process (locations, the relationship of vaccination to other measures, or the need for booster vaccines), shows strong fluctuations across phases but without a clear trend; this could be explained by the comparably lower number of cases, which allows for a few instances to have larger impacts on the comparisons, as well as by the spread of specific rumors, such as the discovery of a vaccine when it was not the case (during Phase 2) or the lack of importance of booster vaccines (common in phase 5). Notably, this narrative was only present among public health authorities because it generally focused on the practical aspects of the process, something that specifically concerned the public health authorities involved in the vaccination campaign.

Discussion

Our study explored the content of misinformation addressed by news media and public health authorities in the Netherlands. By identifying the presence of topics and recurring narratives, it was easier to obtain a clearer picture of the type of false information that the 2 types of studied sources considered more relevant.

The first research question explored the topics of the misinformation cases with which the studied sources dealt. The main topic was the vaccines in both types of sources. Other than vaccines, news media also paid attention to misinformation on the origin, spread, and dangerousness of the virus as well as pharmaceutical measures other than vaccines. In the case of public health authorities, there was less diversity in the addressed misinformation topics, with a particularly strong focus on vaccines and on other measures to control the pandemic, either pharmaceutical, such as medicine and remedies, or nonpharmaceutical, such as lockdowns and tests.

The second research question sought to go beyond topics, digging into more specific recurring narratives. In the news articles, the intentional creation of the virus and the claims about the virus not existing or not being as dangerous were the most abundant. Misinformation narratives regarding specific remedies against the virus (eg, the use of hydroxychloroquine or some natural remedies without a scientific basis) or about the evil behind the vaccines were also frequent. In the case of the public health authorities, the most common recurring narrative focused on the potential harms of the vaccines but without focusing on the intentional evil or harm behind them.

Even if it was not the goal of our study to comprehend the mechanism through which the studied actors decided to address a pandemic-related misinformation case or not, the findings needed to be interpreted in light of the broader, bidirectional dynamics between institutional actors and the public. Prominent misinformation narratives shared on social media among citizens can shape which topics journalists choose to cover and which claims public health authorities decide to address. Thus, the focus on vaccine-related topics and narratives by both sources aligns with existing research that found vaccines were a main misinformation topic [9,11]. Because there is evidence that false information can have agenda-setting capacities [28], it is reasonable that the topics and narratives in Dutch news media and public health communication do not emerge in isolation but are partly responsive to misinformation circulating among the public. However, this dialogue with existing misinformation is not only explained by the abundance of a case of misinformation, as in the case of vaccine misinformation [46], but also by its implications for public health [29]. For instance, a May 2021 Eurobarometer found that misinformation could make people delay or even reject vaccination against COVID-19 [47], and other studies have confirmed the role of misinformation in reduced vaccination uptake [6]. Thus, the potential harm of misinformation to society and to the pandemic response, together with the prominence of a misinformation case in the public and social media discourse, is likely to be a key driver of the choice of institutional actors to address said misinformation case. Future studies are encouraged to further interpret these choices.

More specific observations can be made for each of the studied sources. Regarding news media, their attention to misinformation on the origin and spread of the virus, its remedies, and measures to counter it was also identified in a previous report on misinformation in 2020 [48]. Both the topics and narratives tended to be consistent with some of the most frequent conspiracy theories identified in previous studies. This includes storylines like the alleged involvement of Bill Gates or the connection between the virus and 5G mobile technology [49,50]. On the other hand, public health authorities paid more attention to stories that affected the process of vaccination, testing, and application of measures. One of the reasons behind this is that the studied public health organizations were in charge of the implementation of these measures. As shown in a previous interview study with information intermediaries that included local public health authorities [44], the choices regarding what cases of misinformation to deal with was influenced by the activity and goals of the organization.

It should be noted that, although we differentiated between topics and narratives, there were strong correlations between some of them. For instance, for cases in which a vaccine-related narrative (eg, vaccines are intentionally harmful or evil or vaccines do not work or are not sufficiently tested) was present, the vaccine topic was present as well. The same applied to the narratives on the intentional creation of the virus, which were strongly correlated with the topic on the origin of the virus. Although some topics and narratives were connected by definition, this does not mean that narratives exclusively aligned within one topic; for instance, some of the narratives connecting the pandemic with larger conspiracy schemes or accusing public persons like Bill Gates of their role in the pandemic touched upon the topics of the origin of the virus, the vaccines, and political aspects. In total, our analyses also showed which narratives were most often used in relationship to some topics; this combined observation of topics and recurring narratives offers a more complete picture of the misinformation that the studied sources addressed during the COVID-19 pandemic.

XSL•FO
RenderX

To obtain a comprehensive picture, the third research question of the study sought to go beyond the identification of misinformation topics and narratives and explored the time differences in the distribution of misinformation cases present in news outlets and Twitter posts of public health authorities. In general terms, news media showed a peak of attention to pandemic-related misinformation during the first months of the pandemic when the stories about the origin of the virus, its spread, and the ways to combat the disease were particularly relevant. On the other hand, public health authorities were most active during the year 2021 and beginning of 2022, a period in which the vaccination process was in full swing. More specifically, this is the time in which the vaccine boosters were distributed, which received specific attention as certain narratives argued that they were not needed.

Both the news outlets and the public health authorities paid attention to different topics over different periods of time. The clearest examples are the the higher rates of topics and narratives related to the origin and danger of the virus as well as the measures and strategies to counter it in the early stages, while vaccine-related misinformation was proportionally more present after 2021 in the third and fourth phases of the pandemic. These observations generally concur with those of past research [11], although we managed to combine and correlate topics and narratives, going beyond previous, more descriptive studies. There is, however, room for a deeper understanding of the temporary evolution of the pandemic-related misinformation cases addressed by the studied actors. Our study identified that attention to different topics and narratives among news articles and Twitter posts of public health authorities varied during the COVID-19 pandemic; future research could link this time-varying attention to indicators of public information-seeking and online engagement (eg, search queries, social media activity) for the same narratives, following the approach by Southwell et al [51] of relating virus outbreak news coverage to patterns in online behavior.

To adequately interpret the findings of the study, it is important to consider the different approach to misinformation by news media articles and the Twitter accounts of public health authorities. As summarized in previous paragraphs, some points of convergence emerged, the clearest being the notable attention devoted to vaccine-related misinformation and the similar trends across time. These patterns indicate a connection between the 2 types of sources, not only because they both based their activity on existing circulating misinformation but also because risk communication strategies used by health agencies can influence news media coverage [52] and because public health authorities are common sources of information for journalists covering health topics, especially during pandemics [53,54].

Despite these similarities, there were important divergences in how each type of source addressed misinformation based on their distinct social roles. Journalists tended to examine the impact of misinformation on an issue, covering a wider range of themes with greater depth. This enabled news media to have more diversity in terms of misinformation topics and narratives. Such coverage aligns with the findings of Blanco-Herrero et al [55], who interviewed journalists and identified the level of widespread and newsworthiness of a misinformation case as the main motivators for journalists to deal with them. Thus, they would focus more often on larger news stories, usually with a political and international dimension (eg, the vaccine is evil or the virus is created to control us), as they seek to inform about newsworthy events and developments in society.

In contrast, public health authorities focused on more practical dimensions and decisions that could be affected by misinformation (eg, this measure is now in place, how and when the second dose of the vaccine was needed or distributed), as their communication strategy is connected to their function, which is the promotion of public health and the implementation of measures to counter the effects of the health crisis. Unlike news media, who could cover a diversity of topics based on what was considered newsworthy or relevant for the audience, these authorities limited the use of their Twitter accounts to face rumors, doubts, or questions that could be motivated by misinformation and that affected their usual activity and objectives, without tackling elements that went beyond these areas. Misinformation was rarely explicitly mentioned. Rather than confronting or dissecting it, the studied authorities tended to provide additional information to counter the effects of rumors and misbeliefs.

Two additional factors also shaped these differences. First, platform characteristics determined the style, format, and size of the messages: News articles were longer textual pieces produced by professional journalists; meanwhile Twitter posts were shorter, could include multimedia elements, and were produced by the communication teams of the institutions. Second, our sampling criteria differed: News articles were selected when pandemic-related misinformation was a central element of the text, irrespective of how this misinformation was addressed, which differs from the strategy used in sampling the Twitter posts, where a case was considered when it could be used to counter misinformation even if that was not the sole purpose and there was no direct mention of misinformation. As explained in the methodology, this broader inclusion criterion was applied to public health authorities' Twitter posts to capture a representative and analytically viable sample of their communication. Applying the more restrictive criteria used for news media would have drastically reduced the dataset, limiting the ability to observe patterns and draw meaningful conclusions.

Although these different sampling strategies were justified by the need to obtain a relevant perspective of the studied materials, they did not allow for valid comparisons between the 2 types of studied sources. Thus, we did not conduct these comparisons in our study. The differences between both types of sources were addressed only for a more complete interpretation of the results.

Another limitation is the study's inability to determine the precise reasons why a topic or narrative was mentioned. More in-depth qualitative interview studies that use this content analysis as a basis are needed to understand these decision-making processes. Finally, the study's observations are valid for the Dutch setting, and the findings might not hold true for other national settings. Although replication studies in other countries are recommended, our research did not significantly differ from previous investigations with more

XSL•FO
RenderX

limited time frames or different sources [9,11] or less focused on misinformation [40]. Arguably, the Netherlands does not constitute an outlier in the way the COVID-19 pandemic was addressed [36].

Nonetheless, these limitations do not undermine the core aim of the research. By incorporating 2 distinct types of sources, this study offers a broader perspective on the topics and narratives present in misinformation than previous studies focusing on a single source type, often fact checks [20,30]. Our approach yielded a more comprehensive picture of the misinformation addressed by key information sources in the Netherlands during the pandemic, demonstrating the predominance of vaccine misinformation, the fluctuating nature of misinformation during a crisis, and the different approaches to it that different information sources adopt.

Acknowledgments

The authors thank the students assistants who participated in different coding tasks (in alphabetical order): Anastasia Bogoslavets, Mary Evans, Eva Fritz, John Gatev, Yingqi Huang, Merve Kaya, Lieve Labrujère, Phuong Bao Lien Nguyen, Carlo Partyan, and Oliwia Wyciszkiewicz.

Funding

This project was funded by ZonMw Netherlands, as part of the *BePrepared consortium*, under project number 10710022210002.

Data Availability

Data are openly available in the next OSF archive [56] or upon request to the corresponding author.

Authors' Contributions

Conceptualization: DBH, BvdP, TvdM
Data curation: DBH
Formal analysis: DBH
Funding acquisition: BvdP
Investigation: DBH
Methodology: DBH, BvdP, TvdM
Project administration; DBH, BvdP
Resources: DBH, BvdP, TvdM
Software: DBH
Supervision: BvdP, TvdM
Validation: DBH, BvdP, TvdM
Visualization: DBH
Writing – original draft: DBH
Writing – review & editing: DBH, BvdP, TvdM

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook for news articles.
[PDF File, 229 KB - infodemiology_v6i1e91743_app1.pdf]

Multimedia Appendix 2
Codebook for Twitter posts.
[PDF File, 222 KB - infodemiology_v6i1e91743_app2.pdf]

Multimedia Appendix 3
Correlations between the presence of different topics and narratives.
[DOCX File, 23 KB - infodemiology_v6i1e91743_app3.docx]

Multimedia Appendix 4
Chi-square tests and figures.
[DOCX File, 585 KB - infodemiology_v6i1e91743_app4.docx]

XSL•FO
RenderX

References

1. Barua Z, Barua S, Aktar S, Kabir N, Li M. Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. Prog Disaster Sci 2020 Dec;8:100119. [doi: 10.1016/j.pdisas.2020.100119] [Medline: 34173443]
2. Greene CM, Murphy G. Quantifying the effects of fake news on behavior: evidence from a study of COVID-19 misinformation. J Exp Psychol Appl 2021 Dec;27(4):773-784. [doi: 10.1037/xap0000371] [Medline: 34110860]
3. Bhattacharya S, Singh A. Unravelling the infodemic: a systematic review of misinformation dynamics during the COVID-19 pandemic. Front Commun 2025;10:1560936. [doi: 10.3389/fcomm.2025.1560936]
4. Skafle I, Nordahl-Hansen A, Quintana DS, Wynn R, Gabarron E. Misinformation about COVID-19 vaccines on social media: rapid review. J Med Internet Res 2022 Aug 4;24(8):e37367. [doi: 10.2196/37367] [Medline: 35816685]
5. Ruggeri K, Vanderslott S, Yamada Y, et al. Behavioural interventions to reduce vaccine hesitancy driven by misinformation on social media. BMJ 2024 Jan 16;384:e076542. [doi: 10.1136/bmj-2023-076542] [Medline: 38228339]
6. Pierri F, Perry BL, DeVerna MR, et al. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. Sci Rep 2022 Apr 26;12(1):5966. [doi: 10.1038/s41598-022-10070-w] [Medline: 35474313]
7. Lee SJ, Lee CJ, Hwang H. The impact of COVID-19 misinformation and trust in institutions on preventive behaviors. Health Educ Res 2023 Jan 20;38(1):95-105. [doi: 10.1093/her/cyac038] [Medline: 36564938]
8. Kisa S, Kisa A. A comprehensive analysis of COVID-19 misinformation, public health impacts, and communication strategies: scoping review. J Med Internet Res 2024 Aug 21;26:e56931. [doi: 10.2196/56931] [Medline: 39167790]
9. Jacob C, Hausemer P, Zagoni-Bogsch A, Diers-Lawson A. The effect of communication and disinformation during the COVID-19 pandemic. : European Parliament; 2023 URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2023/740063/IPOL_STU(2023)740063_EN.pdf
10. Unlu A, Truong S, Sawhney N, Tammi T. Setting the misinformation agenda: modeling COVID-19 narratives in Twitter communities. New Media & Society 2025 Jul;27(7):3973-3997. [doi: 10.1177/14614448241232079]
11. Kotseva B, Vianini I, Nikolaidis N, et al. Trend analysis of COVID-19 mis/disinformation narratives-a 3-year study. PLoS ONE 2023;18(11):e0291423. [doi: 10.1371/journal.pone.0291423] [Medline: 37976242]
12. McCombs ME, Shaw DL. The agenda-setting function of mass media. Public Opin Q 1972;36(2):176. [doi: 10.1086/267990]
13. Lwin MO, Lee SY, Panchapakesan C, Tandoc E. Mainstream news media's role in public health communication during crises: assessment of coverage and correction of COVID-19 misinformation. Health Commun 2023 Jan;38(1):160-168. [doi: 10.1080/10410236.2021.1937842] [Medline: 34157919]
14. Drescher LS, Roosen J, Aue K, Dressel K, Schär W, Götz A. The spread of COVID-19 crisis communication by German public authorities and experts on Twitter: quantitative content analysis. JMIR Public Health Surveill 2021 Dec 22;7(12):e31834. [doi: 10.2196/31834] [Medline: 34710054]
15. Balogun BA, Hogden A, Kemp N, Yang L, Agaliotis M. Public health agencies' use of social media for communication during pandemics: a scoping review of the literature. Osong Public Health Res Perspect 2023 Aug;14(4):235-251. [doi: 10.24171/j.phrp.2023.0095] [Medline: 37652679]
16. Maslowska E, Guo Z, Oh SH, Smit E. Exploring public health agencies' communication on social media in the early stages of the COVID-19 pandemic. EJHC 2025;6(1):1-26. [doi: 10.47368/ejhc.2025.101]
17. Rasmussen J, Lindekilde L, Petersen MB. Public health communication reduces COVID-19 misinformation sharing and boosts self-efficacy. J Exp Polit Sci 2024;11(3):327-342. [doi: 10.1017/XPS.2024.2]
18. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. J Med Internet Res 2021 Jun 11;23(6):e29802. [doi: 10.2196/29802] [Medline: 34043526]
19. Brennen JS, Simon FM, Howard PN, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. Reuters Institute for the Study of Journalism. 2020. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2026-08-28]
20. Ceron W, de-Lima-Santos MF, Quiles MG. Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content. Online Social Networks and Media 2021 Jan;21:100116. [doi: 10.1016/j.osnem.2020.100116]
21. Southwell BG, Brennen JSB, Paquin R, Boudewyns V, Zeng J. Defining and measuring scientific misinformation. Ann Am Acad Pol Soc Sci 2022 Mar;700(1):98-111. [doi: 10.1177/00027162221084709]
22. Coronavirus tijdlijn. Rijksoverheid. URL: https://www.rijksoverheid.nl/themas/overheid-en-democratie/coronavirus-tijdlijn [accessed 2026-07-17]
23. Weekly update COVID-19. National Institute for Public Health and the Environment (RIVM). 2026. URL: https://www.rivm.nl/en/coronavirus-covid-19/current/weekly-update [accessed 2026-08-28]
24. Solovei A, van Weert JCM, van den Putte B, et al. Patterns of news consumption during the COVID-19 pandemic crisis: a 2.5 year longitudinal study in the Netherlands. Journal Stud 2024;25(16):1968-1989. [doi: 10.1080/1461670X.2024.2407944] [Medline: 39611048]
25. Yousuf H, van der Linden S, van Essen T, et al. Dutch perspectives toward governmental trust, vaccination, myths, and knowledge about vaccines and COVID-19. JAMA Netw Open 2021 Dec 1;4(12):e2140529. [doi: 10.1001/jamanetworkopen.2021.40529] [Medline: 34967887]

XSL•FO
RenderX

26. van Dijck J, Alinejad D. Social media and trust in scientific expertise: debating the Covid-19 pandemic in The Netherlands. Social Media + Society 2020 Oct;6(4). [doi: 10.1177/2056305120981057]
27. Harambam J. The proliferation of alternative media: how corona conspiracy theories in the Netherlands fostered new social movements. In: Butter M, Knight P, editors. Covid Conspiracy Theories in Global Perspective: Routledge; 2023:252-267. [doi: 10.4324/9781003330769-24]
28. Vargo CJ, Guo L, Amazeen MA. The agenda-setting power of fake news: a big data analysis of the online media landscape from 2014 to 2016. New Media & Society 2018 May;20(5):2028-2049. [doi: 10.1177/1461444817712086]
29. Bruns H, Dessart FJ, Pantazi M. COVID-19 Misinformation: Preparing for Future Crises: Publications Office of the European Union; 2022. [doi: 10.2760/41905]
30. Salaverría R, Buslón N, López-Pan F, León B, López-Goñi I, Erviti MC. Desinformación en tiempos de pandemia: tipología de los bulos sobre la Covid-19. EPI ;29(3). [doi: 10.3145/epi.2020.may.15]
31. Dahlstrom MF. Using narratives and storytelling to communicate science with nonexpert audiences. Proc Natl Acad Sci U S A 2014 Sep 16;111 Suppl 4(Suppl 4):13614-13620. [doi: 10.1073/pnas.1320645111] [Medline: 25225368]
32. Lewandowsky S, Cook J, Ecker U, et al. The Debunking Handbook 2020. [doi: 10.17910/b7.1182]
33. Hassan A, Barber SJ. The effects of repetition frequency on the illusory truth effect. Cogn Res Princ Implic 2021 May 13;6(1):38. [doi: 10.1186/s41235-021-00301-5] [Medline: 33983553]
34. Pennycook G, Cannon TD, Rand DG. Prior exposure increases perceived accuracy of fake news. J Exp Psychol Gen 2018 Dec;147(12):1865-1880. [doi: 10.1037/xge0000465] [Medline: 30247057]
35. Shen C, Song Z, He P, Liu L, Xiong Z. Online rumors during the COVID-19 pandemic: co-evolution of themes and emotions. Front Public Health 2024;12:1375731. [doi: 10.3389/fpubh.2024.1375731] [Medline: 38919926]
36. Hale T, Angrist N, Goldszmidt R, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 2021 Apr;5(4):529-538. [doi: 10.1038/s41562-021-01079-8] [Medline: 33686204]
37. van der Vee N, Boekee S, Hoekstra H, Peters O. Nationale social media onderzoek 2020. In: Newcom Research & Consultancy 2020. URL: https://www.studiohey.nl/file/2020-nsmo-rapportage-2020-publicatieversie-25012020.pdf [accessed 2026-01-20]
38. Hoekstra H, Jonker T, van der Veer N. Nationale Social Media Onderzoek 2023: Newcom Research & Consultancy; 2023. URL: https://www.newcom.nl/wp-content/uploads/2023/01/Newcom-Nationale-Social-Media-Onderzoek-2023.pdf [accessed 2026-01-20]
39. Tuñón Navarro J, Oporto Santofimia E. EU Member States' institutional Twitter campaigns on COVID-19 vaccination: analyses of Germany, Spain, France and Italy. Vaccines (Basel) 2023 Mar 9;11(3):619. [doi: 10.3390/vaccines11030619] [Medline: 36992204]
40. Kompani K, Deml MJ, Mahdavian F, Koval O, Arora S, Broqvist H. Who said what: a multi-country content analysis of European health organisations' COVID-19 social media communication. Int J Public Health 2022;67:1604973. [doi: 10.3389/ijph.2022.1604973] [Medline: 36213137]
41. Guidry JPD, O'Donnell NH, Meganck SL, et al. Tweeting a pandemic: communicating #COVID19 across the globe. Health Commun 2023 Oct;38(11):2377-2386. [doi: 10.1080/10410236.2022.2071192] [Medline: 35510413]
42. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993 May;46(5):423-429. [doi: 10.1016/0895-4356(93)90018-v] [Medline: 8501467]
43. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Methods Meas 2007 Apr;1(1):77-89. [doi: 10.1080/19312450709336664]
44. Blanco-Herrero D, Van Den Putte B, Van Der Meer T. Tailored and trust-based: strategies of information intermediaries countering health misinformation during the COVID-19 pandemic. Information, Communication & Society 2026:1-19. [doi: 10.1080/1369118X.2026.2648687]
45. Study 2 (content analysis). the role of intermediaries during public health crisis. WP1.2 of the beprepared consortium. OSF. URL: https://osf.io/cqvh4/overview?view_only=00f872fc6f0045d798f364d5b26c3176
46. Gabarron E, Oyeyemi SO, Wynn R. COVID-19-related misinformation on social media: a systematic review. Bull World Health Organ 2021 Jun 1;99(6):455-463A. [doi: 10.2471/BLT.20.276782] [Medline: 34108756]
47. Flash Eurobarometer 494: attitudes on vaccination against Covid-19. European Union. URL: https://data.europa.eu/data/datasets/s2512_494_eng?locale=en [accessed 2026-08-29]
48. Siwakoti S, Yadav K, Thange I, et al. Localized misinformation in a global pandemic: report on COVID-19 narratives around the world. Princeton University.: Empirical Studies of Conflict Project; 2021. URL: https://esoc.princeton.edu/publications/localized-misinformation-global-pandemic-report-covid-19-narratives-around-world [accessed 2026-08-29]
49. Bruns A, Harrington S, Hurcombe E. 'Corona? 5G? or both?': the dynamics of COVID-19/5G conspiracy theories on Facebook. Media International Australia 2020 Nov;177(1):12-29. [doi: 10.1177/1329878X20946113]
50. Evanega S, Lynas M, Adams J, Smolenyak K. Coronavirus misinformation: quantifying sources and themes in the COVID-19 'infodemic' (preprint). Journal of Medical Internet Research. Preprint posted online on 2020. [doi: 10.2196/preprints.25143]
51. Southwell BG, Dolina S, Jimenez-Magdaleno K, Squiers LB, Kelly BJ. Zika virus-related news coverage and online behavior, United States, Guatemala, and Brazil. Emerg Infect Dis 2016 Jul;22(7):1320-1321. [doi: 10.3201/eid2207.160415] [Medline: 27100826]

XSL•FO
RenderX

52. Lee ST, Basnyat I. From press release to news: mapping the framing of the 2009 H1N1 A influenza pandemic. Health Commun 2013;28(2):119-132. [doi: 10.1080/10410236.2012.658550] [Medline: 22439616]
53. Husemann S, Fischer F. Content analysis of press coverage during the H1N1 influenza pandemic in Germany 2009-2010. BMC Public Health 2015 Apr 15;15:386. [doi: 10.1186/s12889-015-1742-1] [Medline: 25887542]
54. Morani M, Cushion S, Kyriakidou M, Soo N. Expert voices in the news reporting of the coronavirus pandemic: a study of UK television news bulletins and their audiences. Journalism (Lond) 2022 Dec;23(12):2513-2532. [doi: 10.1177/14648849221127629] [Medline: 38603191]
55. Blanco-Herrero D, van der Meer T, van den Putte B. Role adoption and decision making: how Dutch journalists dealt with misinformation during the COVID-19 pandemic. Journalism Practice 2026 Sep 2;20(9):2732-2750. [doi: 10.1080/17512786.2025.2464925]
56. 2na3s / files / osfstorage. OSF. URL: https://osf.io/2na3s/files/osfstorage [accessed 2026-08-28]

Abbreviations

GGD: Gemeentelijke Gezondheidsdienst
ICR: intercoder reliability
PABAK: prevalence-adjusted and bias-adjusted kappa
RIVM: National Institute for Public Health and the Environment
RQ: research question
WHO: World Health Organization

Edited by M Haupt; submitted 21.Jan.2026; peer-reviewed by B Balogun, B Southwell; revised version received 27.Jul.2026; accepted 07.Aug.2026; published 04.Sep.2026.

Please cite as:
Blanco-Herrero D, van den Putte B, van der Meer TGLA
Misinformation Through the Lens of Dutch News Media and Public Health Authorities: Content Analysis of Topics and Narratives Present in Pandemic-Related Misinformation During COVID-19
JMIR Infodemiology 2026;6:e91743
URL: https://infodemiology.jmir.org/2026/1/e91743
doi:10.2196/91743

© David Blanco-Herrero, Bas van den Putte, Toni G L A van der Meer. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 4.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

YouTube Videos as Educational Resources for Seizure Clusters and Treatment Options: Cross-Sectional Observational Study

Taylor S Elliott[1,2*], BS, PVCert; Andrew J Gienapp[3*], MS; James W Wheless[2,4*], BScPharm, MD

[1]College of Medicine, University of Tennessee Health Science Center, 910 Madison Ave, Memphis, TN, United States

[2]Comprehensive Epilepsy Program, Neuroscience Institute, Methodist Le Bonheur Healthcare, Memphis, TN, United States

[3]Children's Foundation Research Institute, Le Bonheur Children's Hospital, Memphis, TN, United States

[4]Department of Pediatrics, Division of Pediatric Neurology, University of Tennessee Health Science Center, Memphis, TN, United States

[*]all authors contributed equally

Corresponding Author:
Taylor S Elliott, BS, PVCert
College of Medicine, University of Tennessee Health Science Center, 910 Madison Ave, Memphis, TN, United States

Abstract

Background: Patients with seizure clusters and their families often use YouTube videos as resources for information about the illness and how to use concomitant rescue medications; however, many videos include misinformation or incorrect information that could potentially cause harm.

Objective: This study aims to examine YouTube videos intended as educational and instructional resources for seizure clusters and treatment options posted since the approval of intranasal benzodiazepine rescue medications in 2019 and to assess their accuracy and quality using a 3-item rating scale.

Methods: We searched YouTube for 10 specific terms related to seizure clusters and treatment options and analyzed videos from the first 20 results for each to mimic internet user behavior when choosing results with search engines. Videos were rated on 3 criteria: accurate seizure cluster definition, appropriate US Food and Drug Administration (FDA)–approved treatment recommendation, and correct visual demonstration of rescue medication use. A single author rated the videos based on the presence of specific subcriteria of all 3 primary criteria in each video. Score assignment was based on the number of present subcriteria.

Results: Seizure clusters and trade names generated the highest number of videos meeting criteria; other search terms, including nonproprietary names, garnered <8 videos each. Only 19.5% (n=39) of the 200 collected videos provided accurate descriptions of seizure clusters, while 61.0% (n=122) offered no definition, 37.0% (n=74) recommended at least 1 FDA-approved rescue medication, and 37.5% (n=75) offered no rescue therapy at all; 26.5% (n=53) provided a full demonstration of rescue medicine, while 62.5% (n=125) did not. Only 4.5% (n=9) of videos included all subcriteria of the 3 primary criteria. Search results for specific trade names yielded more results, while less specific terminology returned unrelated or unhelpful videos.

Conclusions: Accurate YouTube videos of seizure cluster management are limited and represent a knowledge gap that should be addressed to avoid patient harm. To facilitate better navigation of educational and instructional resources, clinicians should have a ready list of videos available to give to patients or provide a list of search terms that are known to produce accurate results to facilitate better searches. Reputable health care sources could also produce brief, helpful videos of each rescue therapy with search terms that direct patients to their content.

(JMIR Infodemiology 2026;6:e98768) doi:10.2196/98768

KEYWORDS

caregivers; education; rescue medication; seizure clusters; YouTube; intranasal; diazepam; midazolam

Introduction

Seizure clusters are often characterized by seizures that deviate from a patient's normal seizure pattern, appearing intermittently in groups over hours or days [1]. Seizure clusters have been referred to by various names in the past, including seizure flurry, acute repetitive seizures, serial seizures, and crescendo seizures [1-5]. Criteria for seizure clusters range broadly. For example, acute repetitive seizure criteria include experiencing multiple complex partial (now called focal with impaired consciousness) or generalized seizures in 24 hours for adults or 12 hours for children. Additionally, seizure cluster definitions can include experiencing at least 3 seizures in 24 hours [6-11], 3 seizures within 4 hours [8,10,12], or an atypical pattern of seizures that differs from a patient's usual pattern [12,13]. Despite being recognized as dangerous and potentially fatal [14,15], seizure clusters do not have an established definition [1,16-18]. The International League Against Epilepsy, for example, has yet to

XSL•FO
RenderX

establish an official classification. The only available guidance derives from a broad consensus issued by the US Food and Drug Administration (FDA) [1,13] that provides a vague definition of "intermittent, stereotypic episodes of frequent seizure activity (ie, seizure clusters, acute repetitive seizures) that are distinct from a patient's usual seizure pattern" [19,20].

These different and sometimes conflicting definitions for seizure clusters have led to inconsistent treatment advice with studies providing different recommendations [16-18]. Some recommend only FDA-approved rescue medications—fast-acting medicines given at home to stop seizure clusters—for treating seizure clusters, including rectal diazepam, intranasal diazepam, and intranasal midazolam [21,22]. Nonindicated treatments can include extra doses of maintenance antiseizure medications or oral benzodiazepines [15]. For patients, seizure clusters are challenging to manage without the perplexities of definition and treatment options. The lack of an established definition contributes to the confusion due to the variability of potential triggers. Additionally, patients may experience their own unique set of triggers, or they may not be able to define precise physiologic changes or activities that enhance the likelihood of seizure clusters [13]. Auras may or may not be present to signal the onset of cluster seizures. Being able to recognize seizure clusters and understanding appropriate rescue treatments to prevent and control this acute condition is paramount.

Prior studies have shown that patients and caregivers seek health information from websites, journals, and educational videos, with YouTube being a prominent platform [23,24]. Searches conducted by YouTube users for seizure information suggest a demand for resources; however, the accuracy of the information in these videos is inconsistent. Videos can contain varying levels of accuracy and lack useful information. St-Pierre et al [25] found that 45.3% of the videos assessed in their study focused on personal experience, while 63.5% were considered educational. They also found that 18.1% contained wrong or misleading information. Two raters in Fat et al's [26] study found 60% and 64%, respectively, of the videos they assessed were accurate, but only 10 videos were given the study's highest possible rating. Gonulal et al [27] rated 53.3% of the videos in their study as low quality and rated only 1 video as high quality. Wong et al [28] found that most videos (44%) were created by users relating personal experiences or anecdotes, while only 38% were informative or educational. Only 51% of the videos were considered accurate, 40% were neutral, and 9% were inaccurate. Finally, a review by Osman et al [29] concluded overall that YouTube was not a reliable source and that only 32% of produced content was considered unbiased.

Given the lack of classification for seizure clusters and inconsistencies with available treatments, it is no surprise that patients and caregivers may become confused, resort to searching online for information as a result, and heed advice given on social media or found in videos and on websites [14]. While some videos by health care professionals and pharmaceutical companies do include legitimate educational and instructional created content, there is a vast number of sources featuring individuals sharing personal experiences as caregivers and offering what they have learned as sound advice. Most of these videos contain content created by nonmedical professionals. Without expert clinical input to verify information, the advice may be partially or completely incorrect, missing key information, or specific to certain patients and not generalizable. These sources introduce high-risk misinformation that may cause harm.

There are few studies on YouTube videos as a resource for seizure clusters and rescue treatments, many of which do not assess the accuracy of video content. Most studies focus on non–seizure cluster topics. The purpose of this study therefore was to identify and assess the accuracy of YouTube videos intended as educational and instructional resources for seizure clusters and rescue treatments [24,30]. The research question we posed for the study is the following: are the instructional/educational videos on YouTube for seizure clusters and rescue treatments accurate? We hypothesized that most videos analyzed for this study were inaccurate, reflecting a need for improved instruction and education.

Methods

Study Design

This was a cross-sectional observational study that evaluated and rated the accuracy of YouTube videos about seizure clusters and treatments. Videos were selected from YouTube based on 3 criteria: seizure clusters definition, rescue therapy indication, and therapy demonstration. Videos were found using 10 specific terms with YouTube's search feature: (1) *seizure clusters*, (2) *acute repetitive seizures*, (3) *status epilepticus*, (4) *seizure rescue medications*, (5) *rectal benzodiazepines*, (6) *nasal benzodiazepines*, (7) *Nayzilam* (the trade name for intranasal midazolam), (8) *Valtoco* (the trade name for intranasal diazepam), (9) *midazolam*, and (10) *diazepam* [26,28,31]. Searches were performed on YouTube without an account (ie, logged out of YouTube) and in a private browser to avoid potential account-related influences, such as algorithmically determined search results based on an account's prior search and viewing history.

Searches were conducted by the first author (TSE) between June 2023 and March 2024. Inclusion criteria were videos containing at least 1 search term in the video's title or description, publication between 2019 and 2024, and inclusion of educational or instructional content. The year 2019 was set as the starting point for search results to include videos created after FDA approval of new intranasal benzodiazepine rescue medications (ie, intranasal midazolam in 2019 and intranasal diazepam in 2020). Exclusion criteria were videos published before 2019, content not specific to seizure clusters or rescue treatments, absence of all 10 search terms in the title or description, or absence of an educational or instructional component, and sponsored videos (ie, videos that appear at the top of search results via paid placement). No filters for language were applied.

Videos from the top 20 search results for each search term were selected for analysis for a total of 200 videos. The cutoff of 20 was intended to mimic how search engine users focus mostly on the first page of results [32]. Specifically, marketing research about search engine user behavior has found that 75% of users

XSL•FO
RenderX

never go to the second page of results [33] and that only 21% click on more than 1 result [34], meaning that most users click only 1 link on the first page. Since search terms were closely related and YouTube's video algorithm recommends videos with similar titles and content, we permitted and included videos that appeared in the search results of multiple search terms. If this occurred, the video was counted once in the overall dataset, but its performance was tabulated under each search term where it appeared.

Video Analysis

Videos from the top 20 search results for each search term were selected for analysis based on 3 primary criteria mentioned above (seizure cluster definition, treatment recommendation, and demonstration of treatment). Subcriteria for each primary criterion were established by the senior author (an epileptologist). The definition used for seizure clusters was based on the FDA's broad consensus, as it currently was the only definition available. An accurate definition had to include mention of 2 or more seizures within a 24-hour period. An accurate recommendation for treatment had to include a suggestion to use an FDA-approved medication and a description of the official label indication for appropriate age and route of administration. An accurate video demonstration of how to use the rescue medication had to include images of the device, a verbal discussion of the steps to administer the medication, and an example administration performed with a live person, a dummy, or computer-generated images (Table 1).

Table . Rating system for YouTube videos.

Primary criteria	Number of subcriteria present	Subcriteria score
Did the video (1) mention and (2) accurately define seizure clusters?	Both	1.0
	One	0.5
	None	0.0
Did the video (1) suggest and (2) discuss indication of Food and Drug Administration–approved rescue medicine?	Both	1.0
	One	0.5
	None	0.0
Did the video (1) include a demonstration of administering the rescue medication and (2) was it accurate?	Both	1.0
	One	0.5
	None	0.0

Videos were assessed for quality with a video rating system developed and validated originally by Fat et al [26] in their study evaluating YouTube videos as educational tools for infantile spasms. The present study adapted this system by excluding technical quality criteria originally set by Fat et al [26] (ie, quality of lighting, sound, etc) and focusing solely on topical quality based on the primary criteria for seizure clusters in place of infantile spasm criteria.

For the present study, the 3 seizure cluster criteria were divided into 2 subcriteria (Table 1). Scoring was based on the presence of both, one, or none of the subcriteria for each primary criterion. Scores for each video were then tallied, reflecting the number of elements present and correctly conveyed. Overall video scores ranged between a minimum score of 0 to a maximum score of 3, which was an author-developed pragmatic tool rather than a validated instrument. Cutoffs were chosen a priori by the authors to distinguish videos that omitted the most critical elements from those that provided comprehensive instruction. Levels of quality were rated with the following ranges of overall scores: 0.0 - 0.5=very poor quality, 1.0 - 1.5=poor quality, 2.0 - 2.5=good quality, and 3.0=great quality. The first author (TSE) conducted all video ratings.

Ethical Considerations

This study did not include any protected health information from electronic medical records, so an institutional review board approval was not required. Any personal information related to the videos or YouTube channels discussed in the manuscript was deidentified for the study.

Results

Search Results for Total Sample

A total of 200 videos were collected for study with 20 videos per search term. Video posting dates ranged from 2019 to 2024 (Table 2). Mean overall length of all 200 videos was 13.5 (SD 19.0, range 0.3 - 120) minutes long. The longest length by search term was *acute repetitive seizures* (mean 37.4, SD 38.7, range 1 - 120), followed by *status epilepticus* (mean 22.1, SD 24.8, range 4 - 100) and *seizure clusters* (mean 12.7, SD 14.8, range 0.5 - 52). Videos retrieved from the search term *nayzilam* had the shortest video length (mean 4.2, SD 3.0, range 1 - 14). Most videos were presented/narrated by or based their authority on the credentials of a medical doctor (n=34, 17.0%), hospital (n=37, 13.5%), or company (n=23, 11.5%). Some videos included combined authority, such as a medical doctor in combination with a hospital (n=12, 6.0%), organization (n=6, 3.0%), company (n=3, 1.5%), or medical school (n=3, 1.5%). The YouTube channels with the greatest numbers of posted videos were Neurelis (n=20, 10.0%), Valley Children's (n=19, 9.5%), Epilepsy Foundation (n=16, 8.0%), and UCB (n=9, 4.5%). Eighty-two channels posted only a single video.

XSL•FO
RenderX

Table . Characteristics of sampled videos (N=200).

Characteristic	Value
Earliest year of posting	2019
Latest year of posting	2024
Video duration for total sample (minutes), mean (SD, range)	12.5 (19.0, 0.3 - 120)
Video duration by search term (minutes), mean (SD, range)	
Acute repetitive seizures (n=20)	37.4 (38.7, 1-120)
Status epilepticus (n=20)	22.1 (24.8, 4-100)
Seizure clusters (n=20)	12.7 (14.8, 0.5 - 52)
Valtoco (n=20)	11.6 (11.9, 3-45)
Seizure rescue medications (n=20)	12.3 (13.2, 2-52)
Diazepam (n=20)	7.7 (7.0, 2-29)
Rectal benzodiazepines (n=20)	7.5 (5.2, 1-20)
Nasal benzodiazepines (n=20)	6.3 (2.2, 0.3 - 20)
Midazolam (n=20)	3.7 (2.2, 2-29)
Nayzilam (n=20)	4.2 (3.0, 1-14)
Video by presenter and credential, n (%)	
Medical doctor	34 (17.0)
Hospital	27 (13.5)
Company	23 (11.5)
Organization	18 (9.0)
Registered nurse	15 (7.5)
None	14 (7.0)
Medical doctor, hospital	12 (6.0)
Doctor of pharmacy	6 (3.0)
Medical doctor, organization	6 (3.0)
Registered nurse, company	5 (2.5)
Patient	4 (2.0)
Patient, company	4 (2.0)
Registered pharmacist	4 (2.0)
Medical doctor, company	3 (1.5)
Medical doctor, medical school	3 (1.5)
Patient, parent, company	3 (1.5)
Emergency medical technician	2 (1.0)
Medical doctor (psychiatrist)	2 (1.0)
News source	2 (1.0)
Patient, parent	2 (1.0)
Unconfirmed/possible medical doctor	2 (1.0)
Doctor of pharmacy, company	1 (0.5)
Medical doctor, hospital, organization	1 (0.5)
Medical doctor, news source	1 (0.5)
Medical doctor (general practitioner)	1 (0.5)
Parent, organization	1 (0.5)
Physician assistant, certified	1 (0.5)

XSL•FO
RenderX

Characteristic	Value
Registered nurse, organization	1 (0.5)
Veterinary medical doctor	1 (0.5)
Veterinary medical doctor, veterinary school	1 (0.5)
Videos posted by channel name, n (%)	
Neurelis	20 (10.0)
Valley Children's	19 (9.5)
Epilepsy Foundation	16 (8.0)
UCB	9 (4.5)
Neurology Live	6 (3.0)
Drug Talk	4 (2.0)
Cleveland Clinic	3 (1.5)
CURE Epilepsy	3 (1.5)
Orlando Health	3 (1.5)
VJNeurology	3 (1.5)
Andrew Kim MD	2 (1.0)
Child Neurology Foundation	2 (1.0)
drbillakotamd	2 (1.0)
ILAE - International League Against Epilepsy	2 (1.0)
Johns Hopkins Medicine	2 (1.0)
Just a Different Life	2 (1.0)
Medical Videos [Animated]	2 (1.0)
Medicating Normal	2 (1.0)
Medzcool	2 (1.0)
Mini Med Lessons	2 (1.0)
Nurse Ryan	2 (1.0)
Paediatric Emergencies	2 (1.0)
Pixorize	2 (1.0)
RegisteredNurseRN	2 (1.0)
Seattle Children's	2 (1.0)
Strong Medicine	2 (1.0)
Channels with only 1 (0.5%) video each	82 (41.0)

Videos from the top 20 search results for each search term included 103 (51.5%) total videos scored as very poor, 24 (12.0%) scored as poor, 51 (25.5%) scored as good, and 22 (11.0%) scored as great (Table 3). The search term that retrieved the highest scoring videos was *valtoco* (n=20, 10.0%), which garnered 3 great, 15 good, 2 poor, and 0 very poor ratings. Conversely, the search term that retrieved the lowest scoring videos was *status epilepticus* with 20 (10.0%) videos rated very poor. Only 9 (4.5%) distinct videos in the total sample met all 3 inclusion criteria. (Links to these videos are provided in Multimedia Appendix 1.)

XSL•FO
RenderX

Table . Videos by search term ratings (N=200).

Total sample and search term	Very poor, n (%)	Poor, n (%)	Good, n (%)	Great, n (%)
All videos (N=200)	103 (51.5)	24 (12.0)	51 (25.5)	22 (11.0)
Acute repetitive seizures (n=20)	11 (5.5)	3 (1.5)	6 (3.0)	0 (0.0)
Diazepam (n=20)	17 (8.5)	0 (0.0)	0 (0.0)	3 (1.5)
Midazolam (n=20)	15 (7.5)	1 (0.5)	3 (1.5)	1 (0.5)
Nasal benzodiazepines (n=20)	12 (6.0)	2 (1.0)	4 (2.0)	2 (1.0)
Nayzilam (n=20)	3 (1.5)	5 (2.5)	9 (4.5)	3 (1.5)
Rectal benzodiazepines (n=20)	16 (8.0)	0 (0.0)	2 (1.0)	2 (1.0)
Seizure clusters (n=20)	4 (2.0)	5 (2.5)	6 (3.0)	5 (2.5)
Seizure rescue medications (n=20)	5 (2.5)	6 (3.0)	6 (3.0)	3 (1.5)
Status epilepticus (n=20)	20 (1.0)	0 (0.0)	0 (0.0)	0 (0.0)
Valtoco (n=20)	0 (0.0)	2 (1.0)	15 (7.5)	3 (1.5)

Analysis of primary criteria and subcriteria found that only 39 (19.5%) videos provided accurate descriptions of seizure clusters, 39 (19.5%) defined seizure clusters partially, and 122 (61.0%) offered no definition at all (Table 4). Of the total sample, 74 (37.0%) videos recommended at least 1 FDA-approved rescue medication, 32 (16.0%) suggested an unapproved medication for seizure emergencies, 19 (9.5%) discussed both approved and unapproved medications, and 75 (37.5%) offered no rescue therapy at all. Of the 74 videos that did recommend an FDA-approved medication, 69 (93.2%) fully and accurately described its indication, while 3 (4.05%) only offered a partial discussion of rescue indication and 2 (2.7%) included none. Of the total sample, 53 videos (26.5%) provided a full demonstration of rescue medicine, 22 (11.0%) provided an incomplete demonstration, and 125 (62.5%) did not provide any demonstration. Of the combined 75 videos that provided either a full or partial demonstration of the medication, 16 (21.3%) showcased the device alone, 5 (6.67%) included a verbal description of its administration, 6 (8.0%) included both, and 49 (65.3%) provided a guided demonstration with a human being, a dummy, or computer-generated images.

XSL•FO
RenderX

Table . Analysis of video quality criteria based on search results of each search term (N=200).

Total sample and search term	Seizure clusters mentioned, n (%)		Definition given in videos that mentioned seizure clusters, n (%)[a]		Rescue medications discussed, n (%)				Medicine indication(s) discussed/defined, n (%)			Medicine administration demonstrated, n (%)		
	Yes	No	Full and accurate	Incomplete or inaccurate	FDA[b]-approved	Non-FDA-approved	Both	None	Full	Partial	None	Full and accurate	Incomplete or inaccurate	None
All videos (N=200)	78 (39.0)	122 (61.0)	39/78 (50.0)	39/78 (50.0)	74 (37.0)	17 (8.5)	16 (8.0)	93 (46.5)	69 (34.5)	3 (1.5)	128 (64.0)	52 (26.0)	23 (11.5)	125 (62.5)
Seizure clusters (n=20)	15 (75.0)	5 (25.0)	10/15 (66.7)	5/15 (33.3)	9 (45.0)	2 (10.0)	4 (20.0)	5 (25.0)	12 (60.0)	1 (5.0)	7 (35.0)	6 (30.0)	1 (5.0)	13 (65.0)
Acute repetitive seizures (n=20)	8 (40.0)	12 (60.0)	4/8 (50.0)	4/8 (50.0)	3 (15.0)	5 (25.0)	4 (20.0)	8 (40.0)	4 (20.0)	1 (5.0)	15 (75.0)	1 (5.0)	6 (30.0)	13 (65.0)
Status epilepticus (n=20)	0 (0.0)	20 (100.0)	0/0 (0.0)	0/0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	20 (100.0)	0 (0.0)	0 (0.0)	20 (100)	0 (0.0)	0 (0.0)	20 (100.0)
Seizure rescue medications (n=20)	12 (60.0)	8 (40.0)	5/12 (41.7)	7/12 (58.3)	8 (40.0)	4 (20.0)	5 (25.0)	3 (15.0)	12 (60.0)	1 (5.0)	7 (35.0)	9 (45.0)	5 (25.0)	6 (30.0)
Rectal benzodiazepines (n=20)	3 (15.0)	17 (85.0)	3/3 (100.0)	0/3 (0.0)	4 (20.0)	5 (25.0)	0 (0.0)	11 (55.0)	3 (15.0)	0 (0.0)	17 (85.0)	3 (15.0)	0 (0.0)	17 (85.0)
Nasal benzodiazepines (n=20)	6 (30.0)	14 (70.0)	2/6 (33.3)	4/6 (66.7)	8 (40.0)	0 (0.0)	0 (0.0)	12 (60.0)	5 (25.0)	0 (0.0)	15 (75.0)	6 (30.0)	0 (0.0)	14 (70.0)
Nayzilam (n=20)	9 (45.0)	11 (55.0)	3/9 (33.3)	6/9 (66.7)	15 (75.0)	0 (0.0)	2 (10.0)	3 (15.0)	11 (55.0)	0 (0.0)	9 (45.0)	12 (60.0)	1 (5.0)	7 (35.0)
Valtoco (n=20)	18 (90.0)	2 (10.0)	8/18 (44.4)	10/18 (55.6)	19 (95.0)	0 (0.0)	1 (5.0)	0 (0.0)	18 (90.0)	0 (0.0)	0 (0.0)	9 (45.0)	8 (40.0)	3 (15.0)
Midazolam (n=20)	3 (15.0)	17 (85.0)	1/3 (33.3)	2/3 (66.7)	5 (25.0)	1 (5.0)	0 (0.0)	14 (70.0)	2 (10.0)	0 (0.0)	18 (90.0)	4 (20.0)	1 (5.0)	15 (75.0)
Diazepam (n=20)	4 (20.0)	16 (80.0)	3/4 (75.0)	1/4 (25.0)	3 (15.0)	0 (0.0)	0 (0.0)	17 (85.0)	2 (10.0)	0 (0.0)	18 (90.0)	2 (10.0)	1 (5.0)	1 (85.0)

[a]Denominator varies by search term based on the number of videos that mentioned seizure clusters and is indicated in each cell.

[b]FDA: US Food and Drug Administration.

Search Results by Term

Across search terms, there was variability in terms of content, discussion, and demonstration (Table 4 and Table 5). Direct and clinically relevant searches with terms such as *seizure clusters*, *seizure rescue medications*, *Valtoco*, and *Nayzilam* were more likely to return videos with higher accuracy and better demonstrations (ie, most or all subcriteria of the primary criteria were met). For example, results for the term *seizure clusters* included 5 (25.0%) great and 6 (20.0%) good videos. Results for the term *seizure rescue medications* included 3 (15.0%) great and 6 (30.0%) good videos. Results for the term

XSL•FO
RenderX

Valtoco included 3 (15.0%) great and 15 (75.0%) good videos. Results for the term *Nayzilam* included 3 great (15.0%) and 9 good (45.0%) videos.

Table . Analysis of the type of administration demonstration by search term.

Search term (n=20 for each)	Images only, n (%)	Verbal description only, n (%)	Images and verbal description, n (%)	Full demonstration presented[a], n (%)
Seizure clusters	1 (5.0)	—[b]	4 (20.0)	2 (10.0)
Acute repetitive seizures	5 (25.0)	—	1 (5.0)	—
Status epilepticus	—	—	—	—
Seizure rescue medications	1 (5.0)	3 (15.0)	—	10 (50.0)
Rectal benzodiazepines	—	—	—	3 (15.0)
Nasal benzodiazepines	—	—	—	6 (30.0)
Nayzilam	1 (5.0)	—	—	12 (60.0)
Valtoco	8 (40.0)	—	—	9 (45.0)
Midazolam	—	—	—	4 (20.0)
Diazepam	—	—	—	3 (15.0)

[a]Demonstration performed with an individual, a dummy, or computer-generated images.

[b]Not applicable.

Conversely, searching for generic medication names (*midazolam* and *diazepam*) or their drug class (*benzodiazepines*) resulted in videos with lower quality grades. Specifically, results for the term *midazolam* returned only 3 (15.0%) videos that discussed seizure clusters, 2 (10.0%) provided information about indications for treatment, and 4 (20.0%) included a demonstration of medication use, while 14 (70.0%) discussed incorrect therapies or no therapies at all. Results for the term *diazepam* generated similar results, with only 4 (20.0%) videos providing a definition for seizure clusters, 17 (85.0%) videos lacking discussions of rescue medications, 18 (90.0%) lacking indications to treat seizure clusters, and 17 (85.0%) lacking demonstrations. Results for the terms *rectal benzodiazepines* and *nasal benzodiazepines* were similar.

Terms associated with seizure clusters—*acute repetitive seizures* and *status epilepticus*—yielded few videos for patient and health care provider education. Searching for *acute repetitive seizures* resulted in 8 (40.0%) videos that mentioned seizure clusters but only 3 (15.0%) described their management with rescue medication, 4 (20.0%) discussed why they should be treated, and only 1 (5.0%) offered visualizations of how to use appropriate therapies. Interestingly, across all metrics, *status epilepticus* failed to provide any videos that included a seizure cluster description, suggested treatment option, and demonstration of medication.

Discussion

Principal Findings

Given the unpredictability of seizure clusters, their difference from normal seizure patterns, and their higher risk for injury and sudden unexpected death in epilepsy, effective instruction for administering rescue medications is vital. This analysis determined that even while using highly specific terminology related to seizure emergencies and rescue therapies, most YouTube videos collected were rated as poor or very poor (combined n=127, 63.5%). By comparison, only 25.5% (n=51) of videos were rated as good, and 11.0% (n=22) were rated as great. *Seizure clusters*, *Nayzilam*, and *Valtoco* were the only terms in which the number of effective videos found exceeded the number of ineffective videos and were more likely to be seen by medical professionals, possibly because of their link to medical jargon. The results from this study showed that specificity is key to navigating YouTube, and that search terms directly linked to seizure clusters and approved treatments were more effective for gathering accurate videos.

One of the principal findings of this study was that a large proportion of videos contained missing information. While 19.5% (n=39) of the videos provided accurate descriptions of seizure clusters and another 19.5% (n=39) provided a partial or inaccurate description, most videos (n=122, 61.0%) included no definition at all. This pattern also repeated with the other primary criteria. Only 37.0% (n=74) of the videos provided information about FDA-approved treatments, and 37.5% (n=75) did not provide any recommended treatments. While 34.5% (n=69) of the videos included the full indication for the rescue treatment and 1.5% (n=3) included a partial indication, 64.0% (n=128) included none. Only 26.0% (n=52) provided a complete demonstration of the therapy and 11.5% (n=23) provided an incomplete demonstration, and 62.5% (n=125) provided none. These findings highlight a wide gap in knowledge about seizure clusters in video resources on YouTube. Patients with seizure clusters do not have many video resources to choose from. If they found videos with missing information, it is likely that they would not be able to recognize that this information was absent.

Most of the videos with high ratings provided either sufficient details about rescue indication, purpose, and pharmacology, or they provided a well-organized and thorough procedural guide for administration of the medication—few achieved both. As search terms deviated from seizure clusters toward other seizure

XSL•FO
RenderX

emergencies and their treatment, including *status epilepticus* and *rectal benzodiazepines*, the quality of videos decreased. Similarly, as expected, using more general search terminology retrieved videos with less applicability to seizure clusters and rescue medications. These types of search results could be problematic for anyone looking for accurate information. Patients and their caregivers would need to sift through videos with missing information, and roughly half of the videos that did include information could be wrong.

Literature

There are few papers in the literature that have analyzed YouTube video content and seizure clusters. A 2026 study of 563 total videos across multiple platforms, including YouTube, by St-Pierre et al [25] found that 45.3% of the total videos focused on personal experience and 18.1% contained wrong or misleading information. The majority of YouTube videos were labeled as educational (63.5%), but only 8% were considered to be inaccurate. Fat et al [26] assessed 28 videos intended as a resource for patients with infantile spasms, and 2 raters found 60% and 64% of videos to be accurate, respectively, with only 10 videos rated as excellent. Unlike Osman et al [29], the authors concluded that YouTube could be a beneficial patient resource with guided searches; however, Fat et al [26] used a methodology that included technical quality in their rating system, which composed 5 out of 7 criteria and comprised 45% of the total score. Only 1 criterion worth 1 point was used for diagnostic accuracy, and 1 criterion worth 5 points was used to rate the overall quality of the video. As a result, videos could have high scores due to excellent video production and presentation. It is possible that most of the scores were thus based on superficial video and style qualities, and that a more careful assessment of health information content would yield results closer in line with Osman et al [29]. Gonulal et al [27] published a 2026 systematic analysis of YouTube videos for epilepsy-specific exercises and found that 53.3% of the 45 total videos were rated as low quality and only 1 video rated as high quality. Wong et al [28] studied the presentation of seizures and epilepsy in 100 YouTube videos and found that most videos (44%) were created by users relating personal experiences or anecdotes, while only 38% were informative or educational. Only 51% of the videos were considered accurate, 40% were neutral, and 9% were inaccurate.

A recent 2026 cross-sectional analysis by Alpua and Yoldaş [35] assessed the top 50 search results for keywords related to functional neurological disorder. Using several quality scales (Global Quality Scale, modified DISCERN, and *Journal of American Medical Association* benchmark criteria), the study assessed the sample as having “moderate” quality overall but “suboptimal adherence to health information standards.” Analysis of user engagement found a strong correlation between view and like count but a limited association with information quality. A 2020 systematic literature review of 202 studies by Osman et al [29] examined medical and health information in general presented in YouTube videos and concluded that YouTube was not a reliable source. Only 32% of the 22,300 videos analyzed produced content the authors considered unbiased. The authors also reported that videos retrieved in search results were likely selected based on popularity (the number of views and likes) as opposed to any metrics about content. Using this type of methodology suggests that videos with high production values but incorrect information could be driven to the top of search results based solely on engagement. In their study of YouTube videos for functional neurological disorder education, Alpua and Yoldaş [35] emphasized this view by pointedly noting that “engagement metrics do not reliably reflect informational accuracy.”

While YouTube and video demonstrations have their advantages, misinformation is a concern. Content posted by nonmedical professionals was highly variable and often rated as poor, while videos of recorded virtual conferences or of organization-based rescue therapy guides were more likely to be rated as effective due to the inclusion of medical expertise [30,31]. When seizure clusters are the primary concern, understanding their nuances and personalizing therapy is paramount to achieving optimal outcomes; therefore, general videos about epilepsy or seizure therapies may not be ideal for guiding nonmedical professionals, underpinning the need for highly specific searches or direct links to high-quality videos [13,15,22].

Despite nearly half (n=98, 49.0%) of the videos in our study being narrated and published by health care professionals and organizations, there was still a high degree of variability in seizure cluster definition and treatment recommendation. Videos made by actual patients and caregivers appeared more focused on sharing their experience than educating others about seizure management. During instances of seizure emergencies, caregivers in need of a concise, step-by-step tutorial for rescue medications may find incomplete or incorrect information. Additionally, inefficient search terms could further hinder this process. Thus, recommended videos, search terms, and other informational resources for patients and caregivers should be integrated into patient-physician conversations to assist with education and answer lingering questions. Providing this information could prevent patients from accessing inaccurate information, as well as promoting trust in the physician-patient relationship. It could also reduce potential confusion regarding signs and treatment of seizure clusters and empower caregivers to research and administer rescue therapy themselves.

Because of the variability of accuracy and quality of videos found for this study, clinicians should provide specific search terms and direct links to vetted videos for families. Epilepsy centers and professional organizations could also produce their own short, standardized demonstration videos aligned with product labels. There is an opportunity for hospital systems to create curated playlists by audience (eg, families, school nurses, health care providers), as already done by some groups.

Limitations

There are several limitations associated with this study. A major methodological limitation includes the time window in which searches were conducted. Searches were not done at the same time but rather spread out over months, meaning that there is a high degree of variability in the results. Other limitations include restricting the number of videos to the first 20 results, including repeated videos from different search terms, and providing no analysis or survey of actual patient feedback about video quality

XSL•FO
RenderX

to determine video usefulness. The study also used a single rater, thus causing potential selection bias, which is likely to be present in the results, and a potential issue with rater reliability. Additionally, this study relied entirely on YouTube's search function, which is itself subject to various and unknown proprietary internal algorithms for gathering results. This concern is especially amplified considering the increasing use of AI among tech companies. Despite this study's efforts to limit algorithmic influence from YouTube, there is no guarantee that results were completely unbiased. Finally, the fluidity of online content, which is in a constant state of change, means that future studies are unlikely to replicate our results.

Conclusions

Seizure rescue medications are vital to halting seizure clusters, but seizure rescue medications can be challenging when their use and purpose are not properly explained or demonstrated. This study found that most videos posted between 2019 and 2024 within the top 20 search results for 10 seizure-related terms included missing information or lacked key elements necessary for safe and effective rescue use. Many videos were inconsistent in the information conveyed, and some included incorrect or wrongly interpreted information. The stark contrast found in the search results among different search terms indicates a gap in accessibility of educational tools that should be highlighted and remedied by health care professionals.

Being able to differentiate seizure clusters from status epilepticus and recognizing appropriate rescue treatments to prevent and control these acute conditions is paramount and further underscores the gap in accessible information for the public. In the short term, clinicians should have a ready list of effective YouTube videos to give patients during outpatient and inpatient discussions, provided in a card, brochure, or document that includes links to the most appropriate videos. Alternatively, clinicians could provide specific search terms—such as *seizure rescue medications* or proprietary product names—that are known to return useful and accurate videos. Finally, health care professionals and organizations could create their own videos with easily identifiable information, such as hospital or clinic names, and use targeted search engine optimization strategies that could increase the likelihood that patients are directed to their content and away from inaccurate content.

Future studies on this topic should include the development and validation of an instrument for standardizing video quality (one with the potential to be generalizable to other studies of social media content beyond seizure clusters). Other potential avenues for research include examinations of search results from the years following this study's cutoff, quantitative analysis of viewer data (eg, counts of views, likes, and comments), and qualitative analysis of comment text to determine most-used videos and their accuracy. Finally, longitudinal monitoring of content quality over time could also enable health care providers to locate gaps and assist in creating media content with current up-to-date seizure cluster treatment information.

Funding

No financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The videos from which all data were collected for the purposes of this study are readily accessible and publicly available on the YouTube media platform.

Authors' Contributions

Conceptualization: TSE, JWW
Formal analysis: TSE, AJG, JWW
Investigation: TSE
Methodology: TSE, JWW
Project administration: JWW
Supervision: AJG, JWW
Writing – original draft: TSE, AJG, JWW
Writing – review & editing: TSE, AJG, JWW
Validation: AJG
Visualization: TSE, AJG

Conflicts of Interest

TSE and AJG: none declared. JWW: advisor/consultant for Biomarin Pharmaceutical Inc, CombiMatrix, Eisai Inc, GW Pharmaceuticals, Lundbeck Inc, Neurelis Inc, Neuro Event Labs, NeuroPace Inc, Supernus Pharmaceuticals Inc, and Upsher - Smith Laboratories Inc; speakers bureau member for Aquestive Therapeutics Inc, Biomarin Pharmaceutical Inc, Cyberonics Inc, Eisai Inc, Greenwich LifeSciences Inc, LivaNova PLC, Lundbeck Inc, Mallinckrodt, Neurelis Inc, SK Life Science Inc, Supernus Pharmaceuticals Inc, UCB, Upsher - Smith Laboratories Inc, Zogenix Inc; clinical research grants from Acorda Therapeutics, Aucta Pharmaceuticals Inc, Envision Pharma Inc, EpiWatch, GW Pharmaceuticals, and Insys.

XSL•FO
RenderX

Multimedia Appendix 1
Video links to highest quality videos.
[DOCX File, 16 KB - infodemiology_v6i1e98768_app1.docx]

References

1. Gidal B, Klein P, Hirsch LJ. Seizure clusters, rescue treatments, seizure action plans: unmet needs and emerging formulations. Epilepsy Behav 2020 Nov;112:107391. [doi: 10.1016/j.yebeh.2020.107391] [Medline: 32898744]
2. Dreifuss FE, Rosman NP, Cloyd JC, et al. A comparison of rectal diazepam gel and placebo for acute repetitive seizures. N Engl J Med 1998 Jun 25;338(26):1869-1875. [doi: 10.1056/NEJM199806253382602] [Medline: 9637805]
3. Cereghino JJ, Mitchell WG, Murphy J, Kriel RL, Rosenfeld WE, Trevathan E. Treating repetitive seizures with a rectal diazepam formulation: a randomized study. Neurology 1998 Nov;51(5):1274-1282. [doi: 10.1212/wnl.51.5.1274] [Medline: 9818845]
4. Mitchell WG. Status epilepticus and acute serial seizures in children. J Child Neurol 2002 Jan;17(1 Suppl):S36-S43. [doi: 10.1177/08830738020170010501] [Medline: 11918461]
5. Haut SR. Seizure clustering. Epilepsy Behav 2006 Feb;8(1):50-55. [doi: 10.1016/j.yebeh.2005.08.018] [Medline: 16246629]
6. Yen DJ, Chen C, Shih YH, et al. Antiepileptic drug withdrawal in patients with temporal lobe epilepsy undergoing presurgical video-EEG monitoring. Epilepsia 2001 Feb;42(2):251-255. [doi: 10.1046/j.1528-1157.2001.15100.x] [Medline: 11240598]
7. Haut SR, Swick C, Freeman K, Spencer S. Seizure clustering during epilepsy monitoring. Epilepsia 2002 Jul;43(7):711-715. [doi: 10.1046/j.1528-1157.2002.26401.x] [Medline: 12102673]
8. Rose AB, McCabe PH, Gilliam FG, et al. Occurrence of seizure clusters and status epilepticus during inpatient video-EEG monitoring. Neurology 2003 Mar 25;60(6):975-978. [doi: 10.1212/01.wnl.0000053748.83309.28] [Medline: 12654962]
9. Haut SR, Lipton RB, LeValley AJ, Hall CB, Shinnar S. Identifying seizure clusters in patients with epilepsy. Neurology 2005 Oct 25;65(8):1313-1315. [doi: 10.1212/01.wnl.0000180685.84547.7f] [Medline: 16247068]
10. Noe KH, Drazkowski JF. Safety of long-term video-electroencephalographic monitoring for evaluation of epilepsy. Mayo Clin Proc 2009 Jun;84(6):495-500. [doi: 10.4065/84.6.495] [Medline: 19483165]
11. Di Gennaro G, Picardi A, Sparano A, et al. Seizure clusters and adverse events during pre-surgical video-EEG monitoring with a slow anti-epileptic drug (AED) taper. Clin Neurophysiol 2012 Mar;123(3):486-488. [doi: 10.1016/j.clinph.2011.08.011] [Medline: 21920813]
12. Martinez C, Sullivan T, Hauser WA. Prevalence of acute repetitive seizures (ARS) in the United Kingdom. Epilepsy Res 2009 Dec;87(2-3):137-143. [doi: 10.1016/j.eplepsyres.2009.08.006] [Medline: 19748226]
13. Haut SR. Seizure clusters: characteristics and treatment. Curr Opin Neurol 2015 Apr;28(2):143-150. [doi: 10.1097/WCO.0000000000000177] [Medline: 25695133]
14. Penovich PE, Buelow J, Steinberg K, Sirven J, Wheless J. Burden of seizure clusters on patients with epilepsy and caregivers: survey of patient, caregiver, and clinician perspectives. Neurologist 2017 Nov;22(6):207-214. [doi: 10.1097/NRL.0000000000000140] [Medline: 29095321]
15. Jafarpour S, Hirsch LJ, Gaínza-Lein M, Kellinghaus C, Detyniecki K. Seizure cluster: definition, prevalence, consequences, and management. Seizure 2019 May;68:9-15. [doi: 10.1016/j.seizure.2018.05.013] [Medline: 29871784]
16. Buchhalter J, Shafer PO, Buelow JM, et al. Preferred practices for rescue treatment of seizure clusters: a consensus-driven, multi-stakeholder approach. Epilepsy Behav 2021 Apr;117:107836. [doi: 10.1016/j.yebeh.2021.107836] [Medline: 33640567]
17. Pina-Garza JE, Chez M, Cloyd J, et al. Outpatient management of prolonged seizures and seizure clusters to prevent progression to a higher-level emergency: consensus recommendations of an expert working group. Epileptic Disord 2024 Aug;26(4):484-497. [doi: 10.1002/epd2.20243] [Medline: 38813941]
18. Bassani G, Pasini E, Mostacci B, et al. Seizure clusters: current concepts in definition and treatment. J Clin Med 2026 Feb 28;15(5):1847. [doi: 10.3390/jcm15051847] [Medline: 41827262]
19. Specchio N, Wirrell EC, Scheffer IE, et al. International League Against Epilepsy classification and definition of epilepsy syndromes with onset in childhood: position paper by the ILAE Task Force on Nosology and Definitions. Epilepsia 2022 Jun;63(6):1398-1442. [doi: 10.1111/epi.17241] [Medline: 35503717]
20. Wirrell EC, Nabbout R, Scheffer IE, et al. Methodology for classification and definition of epilepsy syndromes with list of syndromes: report of the ILAE Task Force on Nosology and definitions. Epilepsia 2022 Jun;63(6):1333-1348. [doi: 10.1111/epi.17237] [Medline: 35503715]
21. Becker DA, Wheless JW, Sirven J, Tatum WO, Rabinowicz AL, Carrazana E. Treatment of seizure clusters in epilepsy: a narrative review on rescue therapies. Neurol Ther 2023 Oct;12(5):1439-1455. [doi: 10.1007/s40120-023-00515-3] [Medline: 37341903]
22. Wheless JW, Gidal B, Rabinowicz AL, Carrazana E. Practical questions about rescue medications for acute treatment of seizure clusters in children and adolescents with epilepsy in the USA: expanding treatment options to address unmet needs. Paediatr Drugs 2024 Jan;26(1):49-57. [doi: 10.1007/s40272-023-00601-x] [Medline: 37902940]
23. Brigo F, Igwe SC, Ausserer H, et al. Why do people Google epilepsy? An infodemiological study of online behavior for epilepsy-related search terms. Epilepsy Behav 2014 Feb;31:67-70. [doi: 10.1016/j.yebeh.2013.11.020] [Medline: 24361764]

XSL•FO
RenderX

24. Meng Y, Elkaim L, Wang J, et al. Social media in epilepsy: a quantitative and qualitative analysis. Epilepsy Behav 2017 Jun;71(Pt A):79-84. [doi: 10.1016/j.yebeh.2017.04.033] [Medline: 28554148]
25. St-Pierre M, Li J, Toffa DH, Robin A, Nguyen DK. Disinformation and misinformation in epilepsy: an analysis of multiplatform short-form social media video content. Epileptic Disord 2026 Jun 15. [doi: 10.1002/epd2.70309] [Medline: 42295743]
26. Fat MJL, Doja A, Barrowman N, Sell E. YouTube videos as a teaching tool and patient resource for infantile spasms. J Child Neurol 2011 Jul;26(7):804-809. [doi: 10.1177/0883073811402345] [Medline: 21551373]
27. Gonulal Y, Denizci T, Yesilkaya M, Agascioglu E, Aykac S, Singin RHO. Quality and safety of YouTube videos on epilepsy-specific exercises: a systematic analysis using a novel evaluation tool. Seizure 2026 Mar;136:51-59. [doi: 10.1016/j.seizure.2026.02.002] [Medline: 41679102]
28. Wong VSS, Stevenson M, Selwa L. The presentation of seizures and epilepsy in YouTube videos. Epilepsy Behav 2013 Apr;27(1):247-250. [doi: 10.1016/j.yebeh.2013.01.017] [Medline: 23453635]
29. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? a systematic review. BMC Med Educ 2022 May 19;22(1):382. [doi: 10.1186/s12909-022-03446-z] [Medline: 35590410]
30. Bhoot N, Gohil AV, Usgaokar K, Ranpariya K, Yadav R, Nanda A. Epilepsy videos on YouTube: a cross-sectional observational study. Cureus 2023 Aug 22;15(8):e43916. [doi: 10.7759/cureus.43916] [Medline: 37746477]
31. Borusiak P, Langer T, Tibussek D, et al. YouTube as a source of information for children with paroxysmal episodes. Klin Padiatr 2013 Dec;225(7):394-397. [doi: 10.1055/s-0033-1353142] [Medline: 24158883]
32. Dean B. How people use Google search (new user behavior study). Backlinko. 2025 Apr 14. URL: https://backlinko.com/google-user-behavior [accessed 2026-09-14]
33. Dean B. 74 important SEO statistics. Backlinko. 2026 Jan 16. URL: https://backlinko.com/seo-stats [accessed 2026-09-14]
34. Fishkin R. The state of searcher behavior revealed through 23 remarkable statistics. Moz. 2017 Mar 14. URL: https://moz.com/blog/state-of-searcher-behavior-revealed [accessed 2026-09-14]
35. Alpua M, Yoldaş TK. Functional neurological disorder on YouTube: how reliable is the information? Front Public Health 2026;14:1802156. [doi: 10.3389/fpubh.2026.1802156] [Medline: 42022815]

Abbreviations

FDA: Food and Drug Administration

Edited by T Mackey; submitted 18.Apr.2026; peer-reviewed by E Carrazana, R Singin; revised version received 24.Jul.2026; accepted 31.Aug.2026; published 30.Sep.2026.

Please cite as:
Elliott TS, Gienapp AJ, Wheless JW
YouTube Videos as Educational Resources for Seizure Clusters and Treatment Options: Cross-Sectional Observational Study
JMIR Infodemiology 2026;6:e98768
URL: https://infodemiology.jmir.org/2026/1/e98768
doi:10.2196/98768

© Taylor S Elliott, Andrew J Gienapp, James W Wheless. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 30.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Repeated Exposure to Polarized Health-Related News on Explicit and Implicit Attitudes Toward Dietary Supplements: Online Experimental Study

Eugen-Călin Secară, MSc; Nicolae-Adrian Opre, Prof Dr

Department of Psychology, Faculty of Psychology and Educational Sciences, Universitatea Babe -Bolyai, Mihail Kogălniceanu 1, Cluj-Napoca, Cluj, Romania

Corresponding Author:
Nicolae-Adrian Opre, Prof Dr
Department of Psychology, Faculty of Psychology and Educational Sciences, Universitatea Babe -Bolyai, Mihail Kogălniceanu 1, Cluj-Napoca, Cluj, Romania

Abstract

Background: Repetition is a central feature of digital news consumption, where engagement-driven algorithms often expose users to similar health-related content. Prior research suggests that repeated exposure can influence perceived truth and evaluation, but most studies use brief or decontextualized stimuli and have rarely distinguished between explicit and implicit attitudes. Little is known about how repeated exposure to full-length, polarized health news shapes explicit and implicit attitudes, particularly toward familiar products such as dietary supplements.

Objective: This study aims to examine whether 2 weeks of exposure to positively, negatively, or mixed-valence health news articles would alter explicit and implicit attitudes toward dietary supplements, and whether engagement mediated these effects or initial attitudes moderated them.

Methods: In a preregistered 4 (group: PRO, CON, MIX, and control) × 3 (time: T0, T1, and T2) mixed experimental study, 228 participants (174 women, PRO: n=68, CON: n=51, MIX: n=52, control: n=57; mean age 2.81, SD 4.88 years) were randomly assigned to receive one full-length article per day for 2 weeks. Articles presented positive (PRO), negative (CON), mixed (MIX), or neutral space-related information (control). Explicit attitudes toward dietary supplements (perceived efficiency, harmfulness, and willingness to recommend) were assessed with visual analog scales, and implicit attitudes with an Implicit Association Test at baseline, 1 week, and 2 weeks. Mixed analysis of covariances (ANCOVAs) controlled for self-reported exposure to health-related news. Mediation, moderation, and moderated mediation analyses examined whether time spent reading and baseline attitudes influenced change.

Results: The change in implicit attitudes was not significant ($F_{6,\ 428}$=1.90; P=.08). In contrast, explicit attitudes showed a significant group × time interaction ($F_{5.04,\ 428}$=14.34; P<.001). Explicit attitudes became more favorable in the PRO group (M_{Δ}=14.86, SE=5.93, d=0.30; P=.02) and less in the CON (M_{Δ}=–53.42, SE=7.35, d=1.02; P<.001) and MIX groups (M_{Δ}=–24.56, SE=6.77, d=0.50; P=.001). At T2, the CON group reported lower explicit attitudes than the control group (M_{Δ}=–32.84, SE=1.75, d=0.32; P=.02), and the PRO group scored higher than the CON and MIX groups (M_{Δ}=52.08, d=0.52; P<.001 and M_{Δ}=27.07, d=0.29; P=.03). Exploratory item analyses showed large effects of negative exposure on perceived harmfulness and recommendability. No mediation by reading time and no moderation by baseline attitudes were supported.

Conclusions: Repeated exposure to polarized health-related news shifted explicit but not implicit attitudes toward dietary supplements. Negative content exerted the strongest influence. Engagement and prior attitudes did not meaningfully shape these outcomes. These findings suggest that even brief, routine exposure to polarized health information may accumulate into explicit evaluative change, underscoring the importance of balanced digital news environments.

(JMIR Infodemiology 2026;6:e88632) doi:10.2196/88632

KEYWORDS

illusory truth effect; mere exposure effect; implicit attitudes; explicit attitudes; health-related content; polarization; repeated exposure; misinformation; public health; health information; echo chamber

XSL•FO
RenderX

Introduction

Background

Exposure to online information is increasingly shaped by algorithms designed to optimize user engagement. Combined with the decline of traditional news sources [1] and the growing prevalence of paid promotional content within news feeds [2], this shift has raised concerns about exposure to information that aligns with users' existing beliefs while limiting contact with opposing viewpoints, a phenomenon commonly described as echo chambers [3,4]. Although recent evidence suggests that the polarizing effects of political echo chambers are limited for most users [5-8], health-related content (HRC) has exhibited a marked informational divide, particularly during the COVID-19 pandemic [9], where automated distribution mechanisms such as bots contributed to the spread of unreliable information [10].

Unlike political content, HRC is frequently linked to direct commercial interests, often promoting specific products or brands. Because sensationalist content generates greater engagement than balanced reporting [11], commercial incentives may encourage polarized framing and selective presentation of evidence. Beyond restricting informational diversity, health-related echo chambers repeatedly expose users to similar information, engaging cognitive processes through which repetition can alter evaluations regardless of accuracy [12,13].

Repetition and Evaluation

The mere exposure effect describes the tendency for repeated encounters with a stimulus to increase positive affect and perceived familiarity even in the absence of conscious recognition or evaluative intent [12], an effect observed across a wide range of stimuli including words, symbols, and faces [14]. In online health communication, repeated exposure to a claim, product, or figure may therefore enhance likability, irrespective of its accuracy.

Closely related is the illusory truth effect, whereby repeated statements are judged as more truthful than novel ones [13,15]. This effect has been demonstrated across variations in plausibility [16], cognitive ability, and cognitive style [17], and persists even when information is explicitly qualified or labeled as uncertain, improbable, or advertising [18,19]. Initial repetitions produce the greatest increases in truth judgments [20] and spaced repetitions exert stronger effects [21,22]. Studies examining HRC have demonstrated increased belief in both true and false repeated claims about tobacco [23] and COVID-19 [24]. Repetition increases processing fluency, the ease with which information is perceived and retrieved, which is then misattributed to truthfulness rather than familiarity itself [25,26].

In algorithmically mediated environments, where users are repeatedly exposed to similar health-related information, such fluency-driven processes may systematically influence evaluations and beliefs.

Explicit and Implicit Attitudes

Human behavior reflects the joint influence of explicit and implicit attitudes, 2 evaluative systems that differ in awareness, control, and their impact on action. Explicit attitudes are consciously accessible evaluations that can be reported and reflected upon, whereas implicit attitudes are automatic associations that may operate outside conscious awareness and diverge from self-reported beliefs, influencing spontaneous or habitual behavior [27-29].

Explicit attitudes predict deliberate and controlled actions, such as making health decisions after reflection or expressing stated opinions [30]. Implicit attitudes predict spontaneous or nonverbal behaviors, such as avoidance, facial expression, or tone, particularly under conditions of limited time or cognitive resources [31]. The interaction between the 2 systems can produce inconsistency between what people *say* and what they *do*, a discrepancy relevant for understanding reactions to polarized health information, where explicit belief change may not be reflected in behavioral tendencies.

Within polarized information environments, repeated exposure can be expected to affect explicit attitudes through fluency-driven increases in perceived truth (ie, the illusory truth effect) and to influence implicit attitudes through consistent evaluative pairings that strengthen affective associations [32] or through the presentation of crucial information which leads to the reinterpretation of previously held attitudes [33-35].

The Current Research

The aim of this study was to investigate how repeated exposure to polarized health information influences explicit and implicit attitudes toward dietary supplements. Prior studies on the illusory truth effect and evaluative conditioning have typically relied on isolated statements or short stimuli, limiting ecological validity. To better approximate real-world information exposure, this study examined the effects of repeated exposure to full-length online articles containing positive, negative, or neutral information.

Repeated exposure to congruent information was expected to shift attitudes in the direction of the presented valence. Specifically, positive exposure was expected to increase favorable implicit (H1a) and explicit (H1b) attitudes, whereas negative exposure was expected to decrease favorable implicit (H2a) and explicit (H2b) attitudes. Exposure to both positive and negative information was expected to produce changes in attitudes over time (H3a, H3b). Comparisons with a neutral control condition were expected to clarify whether any observed changes reflected informational valence rather than repeated measurement or the passage of time (H4-H6).

Beyond these primary effects, the study examined whether baseline attitudes moderated attitude change (H7), anticipating that participants holding more positive initial views of dietary supplements would show smaller changes when exposed to positive information but larger negative shifts in the negative exposure condition, based on how diagnostic the nature of the presented information is with regard to previously held beliefs. Moreover, the relationship between informational valence and attitude change was expected to be mediated by engagement, such that greater time spent reading articles would predict stronger attitude change (H8). Finally, this mediation was hypothesized to be moderated by initial attitudes, reflecting the

XSL•FO
RenderX

interaction between prior beliefs and the effects of repeated exposure (H9).

Methods

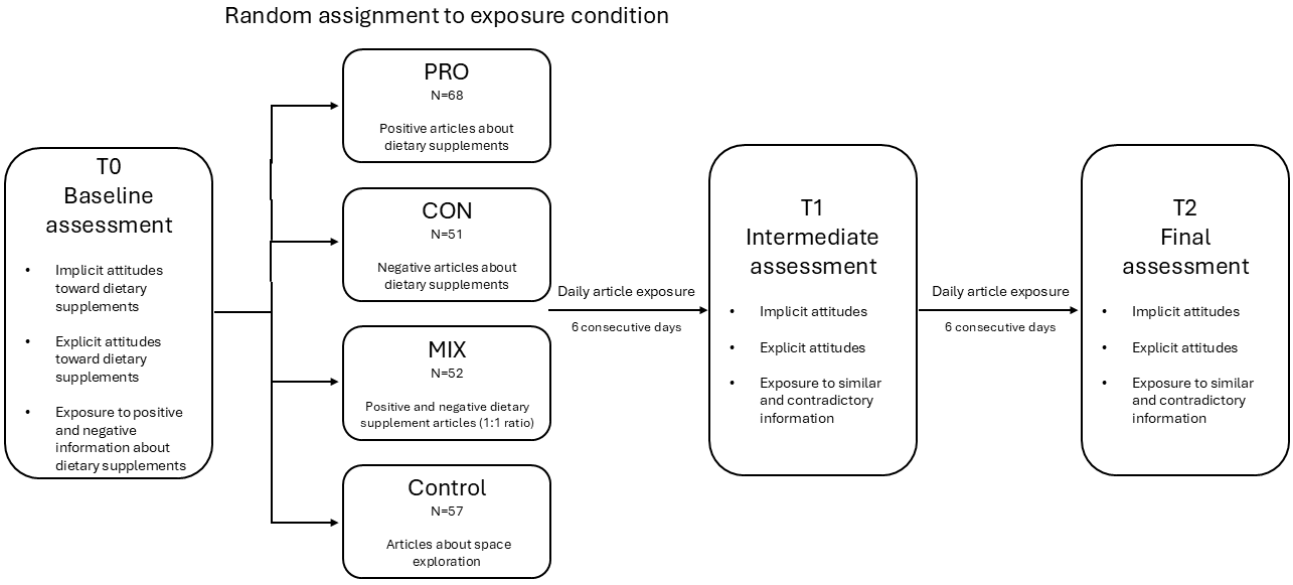
Design and Procedure

The study used a 4×3 mixed experimental design. The between-subjects factor was group, defined by the type of articles participants read (PRO: positive information about dietary supplements; CON: negative information about dietary supplements; MIX: both positive and negative information regarding dietary supplements presented across different articles; or Control: neutral content about space exploration, unrelated to health), and the within-subjects factor was Time, with assessments at baseline (T0), 1 week (T1), and 2 weeks (T2). Participants were randomized with replacement by the Gorilla online experimental platform [36] into 1 of the 4 groups before exposure began. An overview of the study design is presented in Figure 1.

Participants accessed the study via a link to the Gorilla platform. They were presented with a cover story stating that the purpose of the project was to assess personal characteristics of readers who enjoy different types of online articles. After providing informed consent, participants completed the initial assessment (T0) and were randomized into 1 of the 4 groups, being presented with the first article of their assigned condition. After reading, they rated the article on a 4-star scale and were informed that their next article would be available the following day.

Each participant then received 1 article per day for 6 consecutive days, followed on the seventh day by the second assessment (T1). The procedure was repeated for a second cycle of 6 daily articles, after which participants completed the final assessment (T2). The platform recorded time spent reading each article (in seconds). Each assessment included measures of explicit attitudes, implicit attitudes, and frequency of online exposure to health information that either aligned with or contradicted the study content.

Figure 1. Study design.



The article stimuli were developed in November 2022. To approximate the type of information users may encounter when searching for dietary supplement-related content online, the first 20 Google results pages were screened for a range of commonly searched dietary supplement topics spanning general supplement use and several health-related domains. The complete list of search terms is provided in Table S1 of Multimedia Appendix 1.

Candidate articles were selected to represent recurring perspectives commonly encountered in online discussions of dietary supplements. Attention was paid to recurring themes related to nutrition, natural and traditional remedies, mechanisms and effectiveness, and safety considerations. Within each perspective, preference was given to articles that presented more developed arguments, referred to supporting evidence or expert recommendations where available, and remained understandable to a general audience. To maintain consistency across stimuli, selected articles included a short introduction and favored repeated references to supplement-related terms such as “vitamins,” “minerals,” “nutrients,” and “amino acids.”

Information from selected articles was reorganized and edited to create stimuli of comparable length (approximately 450 - 500 words). When source articles contained both positive and negative information, content was separated by valence to maintain experimental control while preserving the original thematic content. For example, one negative stimulus (C1) was constructed from a Buiced article [37], retaining only information concerning potential side effects, limited premarket evidence requirements, and the distinction between naturalness and safety. This material was sufficient to reach the target length without supplementation from additional sources.

In contrast, one positive stimulus (P1) drew its introduction from a Nova Southeastern University article [38] and retained positively framed information concerning supplement effectiveness, safety, and marketplace oversight. Because this content did not reach the target length, additional material

XSL•FO
RenderX

regarding rigorous state regulation of the dietary supplement marketplace from the Buiced article [37] was incorporated.

Using this procedure, 28 stimuli were initially developed and pretested (Table S1 in Multimedia Appendix 1). Of these, 24 stimuli were ultimately included in the study (12 PRO and 12 CON articles). Table 1 summarizes the topics and recurring themes represented in the final stimulus set, together with illustrative examples of how these themes were instantiated within each stimulus. Full stimuli are available in Multimedia Appendix 2.

Table . Recurring themes represented across dietary supplement topics and their instantiation within the article stimuli.

Theme	General supplement use	Sleep	Weight loss	Influenza	Mental health	Diabetes
Nutrition	P1[a] (nutrient gaps), P2 (deficiencies), C1[b] (diet first)	P4 (vitamin B6; magnesium)	P5 (diet support)	P8 (vitamin C, D), C8 (diet)	P10 (mood enhancer)	P11 (micronutrients), C12 (diet first)
Natural or traditional remedies	P2 (plant based), C2 (natural ≠ safe)	P3 (valerian; 5-HTP), P4 (lemon balm; GABA), C3 (sleep aids), C4 (valerian; chamomile)	P6 (green tea; chitosan), C6 (weight-loss botanicals)	P7 (ginger; turmeric), P8 (garlic; green tea), C7 (herbal supplements)	P9 (passionflower; ashwagandha), P10 (L-theanine; probiotics), C9 (kava; valerian), C10 (passionflower; St. John's wort)	P12 (bitter melon; chromium), C11 (Ayurvedic products)
Mechanisms, evidence, safety	P1 (evidence), P2 (expert opinion), C1 (toxicity; interactions), C2 (adverse effects)	P3 (sleep quality), P4 (sleep regulation), C3 (side effects; interactions), C4 (efficacy concerns; interactions)	P5 (weight management), P6 (fat metabolism), C5 (limited efficacy; adverse effects), C6 (stimulant risks)	P7 (immune enhancement), P8 (viral inhibition), C7 (limited evidence; interactions), C8 (antioxidant risks)	P9 (anxiety reduction), P10 (mood regulation), C9 (side effects; interactions), C10 (limited evidence; risks)	P11 (glucose control), P12 (AMPK; glucose uptake), C11 (hypoglycemia risk), C12 (limited evidence; interactions)
Regulation, product quality	P1 (GMP standards), C1 (premarket exemption), C2 (quality variability)	C3 (unregulated products), C4 (ingredient variability)	C5 (fraudulent products), C6 (FDA[c] bans)	C7 (quality variability)	C9 (quality concerns)	C11 (illegal products), C12 (marketing loopholes)

[a]P: positive article (PRO exposure group).

[b]C: negative article (CON exposure group).

[c]FDA: US Food and Drug Administration.

The MIX condition was intended to approximate exposure to competing health narratives encountered across different information sources, rather than balanced reporting within a single article. Due to a technical error in the platform, participants in the MIX group did not receive articles in the intended randomized order. Instead, they were systematically presented with a positive article followed by a negative one, such that the day immediately preceding each assessment (T1 and T2) always featured a negative (CON-type) article. This deviation was consistent across all MIX participants and was documented prior to data analysis.

During the initial visit to the platform, participants provided contact information that was used to remind them to engage daily with the study materials. They subsequently received reminders via email, SMS text message, and, when no response was recorded by the evening, by telephone. Because the platform could not advance without completing the daily task, participants who missed a day were automatically removed from the study and received no further notifications.

Participants

Participants were recruited from the Babe -Bolyai University student population, who were offered course credit in exchange for participation, and from the general Romanian population through social media posts and targeted online advertisements. Inclusion criteria required participants to be at least 18 years old and to have access to a keyboard-compatible device for completing the online tasks. A power analysis based on the procedures described by Preacher et al [39] indicated that a total sample of 200 participants (50 per group) would provide sufficient power to detect a medium-sized conditional indirect effect in moderated mediation models. Recruitment was planned to reach this target, and data collection concluded once it was achieved, with participants who had already begun the study allowed to complete their participation.

A total of 345 participants initially enrolled in the study. Of these, 117 dropped out during the study, resulting in a final sample of 228 participants. Demographic characteristics of the final sample are presented in Table 2. Participants were distributed across conditions as follows: PRO (n=68), CON (n=51), MIX (n=52), and Control (n=57).

XSL•FO
RenderX

Table . Demographic characteristics of the final sample.

Characteristics	Values
Age (years), mean (SD)	20.81 (4.88)
Gender, n (%)	
Women	174 (76.3)
Men	52 (22.8)
Other or prefer not to say	2 (0.9)
Qualifications, n (%)	
Primary education	1 (0.4)
Secondary education	169 (74.1)
Undergraduate degree	42 (18.4)
Graduate degree	16 (7)

Measures

Implicit Attitudes

Implicit attitudes toward dietary supplements were assessed using a computerized Implicit Association Test (IAT, [40]) administered via the Gorilla online platform. The IAT measured the relative strength of automatic associations between dietary supplements and positive versus negative evaluative attributes.

Participants categorized stimuli into one of two target categories—dietary supplements (eg, “vitamins,” “minerals,” “nutrients”) and objects (eg, “trousers,” “chair,” “poster”)—and two attribute categories—positive (eg, “healthy,” “good,” “safe”) and negative (eg, “risky,” “bad,” “dangerous”). During the task, single words appeared sequentially in the center of the screen, and participants were instructed to classify each as quickly and accurately as possible using 2 response keys corresponding to the category pairings displayed on-screen. The task alternated between congruent blocks (eg, dietary supplements + positive and objects + negative) and incongruent blocks (eg, dietary supplements + negative and objects + positive). Reaction times and accuracy were recorded. Incorrect responses triggered an error message, and the trial continued only after the correct key was pressed, penalizing errors through longer response times [41].

An IAT D-score was computed for each participant at each time point (T0, T1, and T2) following standard scoring procedures [41]. The D-score was calculated by subtracting the mean reaction time for congruent blocks from that for incongruent blocks and dividing by the pooled standard deviation of all response latencies. Higher D-scores indicated more favorable implicit attitudes toward dietary supplements.

Explicit Attitudes

Explicit attitudes were measured using three visual analog scale (VAS) questions: “To what degree do you consider dietary supplements to be efficient?”; “To what degree do you consider dietary supplements to be harmful?” (reverse-scored); and “To what degree would you recommend dietary supplements to a loved one?” Scores were aggregated across items at each time point to yield an explicit attitude index, with higher values reflecting more favorable evaluations. Internal consistency was acceptable at every evaluation (Cronbach α=0.79, 0.75, and 0.76, respectively).

The frequency of exposure to health-related news and similar news was assessed using the following questions: “How often do you read health-related news?” (only at T0); “During the last week, how often have you encountered positive news articles related to dietary supplements?” (only at T0), “During the last week, how often have you encountered negative news articles related to dietary supplements?” (only at T0) “During the last week, how often have you encountered similar news articles to those presented on this platform?” (T1 and T2), “During the last week, how often have you encountered news articles that contradicted those presented on this platform?” (T1 and T2); and “Approximately how much time did you spend last week reading health-related news?.”

Statistical Analysis

Data were analyzed using mixed analysis of covariance (ANCOVA) models with group as the between-subjects factor and time as the within-subjects factor. Dependent variables were explicit and implicit attitude scores. Covariates included time spent reading health-related news and self-reported exposure to health-related news on the topic of dietary supplements, assessing both similar and contradictory news.

Planned contrasts compared T2-T0 changes within the PRO, CON, and MIX groups and between-group differences at T2. One-tailed significance tests were used in accordance with our directional hypotheses (PRO and CON), and 2-tailed tests for the nondirectional hypothesis about the MIX group.

Moderation, mediation, and moderated mediation analyses were conducted using PROCESS [42]. These analyses tested whether baseline attitudes moderated the effect of group assignment on attitude change, whether time spent reading mediated this effect, and whether this mediation was further moderated by initial attitudes or exposure to similar news. Indirect effects were evaluated using bias-corrected bootstrapped confidence intervals (BootCI).

All data were analyzed using IBM SPSS 25.

XSL•FO
RenderX

Ethical Considerations

The research was conducted within the framework of the author's doctoral project at Babe -Bolyai University, which received a doctoral scholarship from the Romanian Ministry of Education. At the time of doctoral enrollment (2019), ethical approval for the doctoral research project was granted through the institutional approval process rather than through separate approvals for individual studies. Consequently, the project is registered under reference number 14459/01.10.2019. The responsible ethics committee was the Research Ethics Subcommittee of Babe -Bolyai University, Cluj-Napoca.

All participants provided informed consent prior to enrollment. Participants were compensated either with course credit (students) or entry into a lottery for one of ten €20 (US $21.65) prizes. Participants were debriefed about the true purpose of the study upon completion of data collection. Participant data were securely collected and stored on the Gorilla.sc research platform. Access to the research data was restricted to the research team, and all analyses and reporting were conducted using anonymized, aggregated data. Contact information was collected solely for the purpose of sending reminders during the 2-week study period and, where applicable, administering course credit or participation in the optional prize lottery. Research team members responsible for contacting participants had access only to participants' participation status (ie, whether they had completed the current day's session) and not to their study responses. Following completion of the study, final debriefing, allocation of course credit, and administration of the optional lottery, all contact information was permanently deleted.

Results

Descriptive statistics for the primary study variables are presented in Table 3. These include explicit and implicit attitudes toward dietary supplements, assessed at baseline (T0), 1 week (T1), and 2 weeks (T2), as well as a measure of participant engagement with the study materials, operationalized as the average time spent reading the articles.

Table . Descriptive statistics.

Variables	PRO (n=68), mean (SD)	CON (n=51), mean (SD)	MIX (n=52), mean (SD)	Control (n=57), mean (SD)
T0				
Implicit attitudes	−0.08 (0.42)	0 (0.48)	−0.12 (0.46)	−0.17 (0.43)
Explicit attitudes	190.21 (62.74)	202.25 (56.37)	200.17 (44.44)	203.60 (48.83)
Efficiency rating	61.99 (24.68)	66.86 (20.82)	67.79 (16.83)	67.18 (20.17)
Harmfulness rating	29 (19.87)	24.25 (18.10)	29.06 (20.62)	26.16 (17.44)
Recommendation willingness	56.22 (27.11)	58.65 (25.43)	60.44 (19.42)	61.58 (22.79)
Time spent reading study articles (s)	1164.82 (751.06)	1205.21 (1117.17)	1556.27 (106.60)	1399.18 (856.19)
T1				
Implicit attitudes	−0.05 (0.43)	−0.13 (0.47)	−0.30 (0.47)	−0.18 (0.45)
Explicit attitudes	201.31 (51.03)	150.80 (57.67)	184.92 (41.35)	188.12 (49.91)
Efficiency rating	67.93 (21.33)	49.82 (21.43)	64.29 (18.69)	62.72 (20.07)
Harmfulness rating	27.69 (20.65)	45.35 (22.19)	35.02 (18.06)	28.77 (18.48)
Recommendation willingness	60.07 (22.32)	45.33 (23.12)	54.65 (21.34)	53.18 (23.06)
T2				
Implicit attitudes	−0.15 (0.41)	−0.09 (0.39)	−0.23 (0.41)	−0.13 (0.40)
Explicit attitudes	206.15 (49.22)	148.06 (55.75)	176.40 (45.61)	187.54 (52.23)
Efficiency rating	68.94 (18.67)	47.37 (22.13)	61.58 (20.45)	60.65 (22.06)
Harmfulness rating	26.01 (18.36)	44.08 (20.93)	37.62 (21.13)	25.98 (16.56)
Recommendation willingness	62.22 (20.84)	43.76 (23.65)	51.44 (23.20)	51.88 (24.04)

Self-reported frequency of exposure to health-related news, time spent reading such news, and exposure to positive and negative articles about dietary supplements outside the experimental context were also assessed at each time point and included in the analyses as control variables. Full descriptive data for these control measures are available in Table S2 of Multimedia Appendix 1.

To examine changes in implicit attitudes, a 4×3 mixed-design ANCOVA was conducted on IAT *D*-scores across the 3 assessments, with group (PRO, CON, MIX, and Control) as the

XSL•FO
RenderX

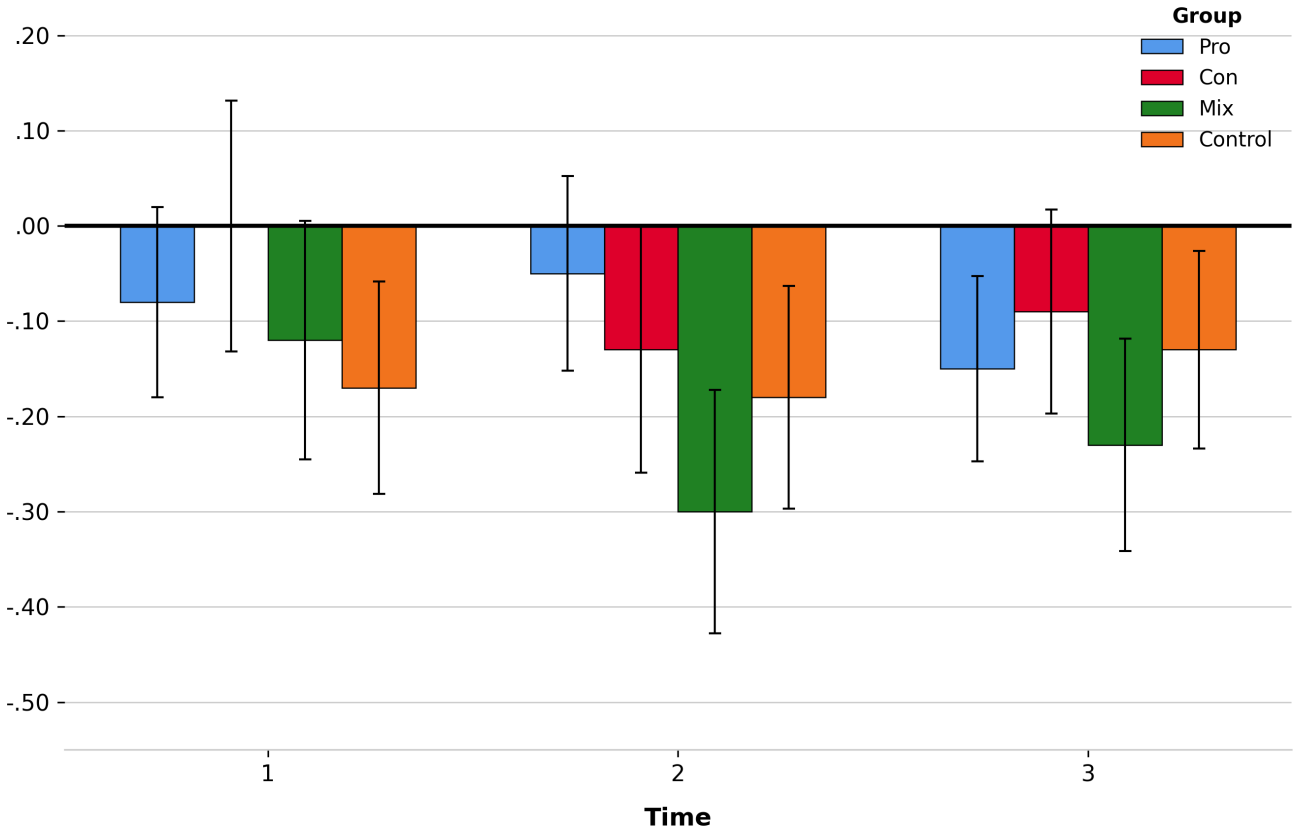
between-subjects factor and time (T0, T1, and T2) as the within-subjects factor. Time spent reading health-related news, frequency of exposure to health-related information (T0, T1, and T2), and time spent reading pro- and antisupplement articles (T0, T1, and T2) were included as covariates. The analysis revealed that the Group × Time interaction was not statistically significant ($F_{6, 428}$=1.90, η^2 =0.03; P=.08). Given the absence of a significant interaction, no follow-up contrasts were performed. Hence, the collected data were insufficient to support H1a, H2a, H3a, H4a, H5a, and H6a.

As illustrated in Figure 2, implicit attitudes remained relatively stable across time and exposure conditions.

To analyze changes in explicit attitudes toward dietary supplements, a 4×3 mixed-design ANCOVA was conducted on the composite explicit attitude scores, with group (PRO, CON, MIX, and control) as the between-subjects factor and time (T0, T1, and T2) as the within-subjects factor. Time spent reading health-related news, frequency of exposure to health-related information (T0, T1, and T2), and time spent reading pro- and anti-supplement articles (T0, T1, and T2) were included as covariates.

The analysis revealed a significant Group × Time interaction ($F_{5.04, 428}$=14.34, η^2 =0.167; P<.001), showing that changes in explicit attitudes differed across exposure conditions.

Figure 2. Changes in implicit attitudes toward dietary supplements across time as a function of exposure condition.



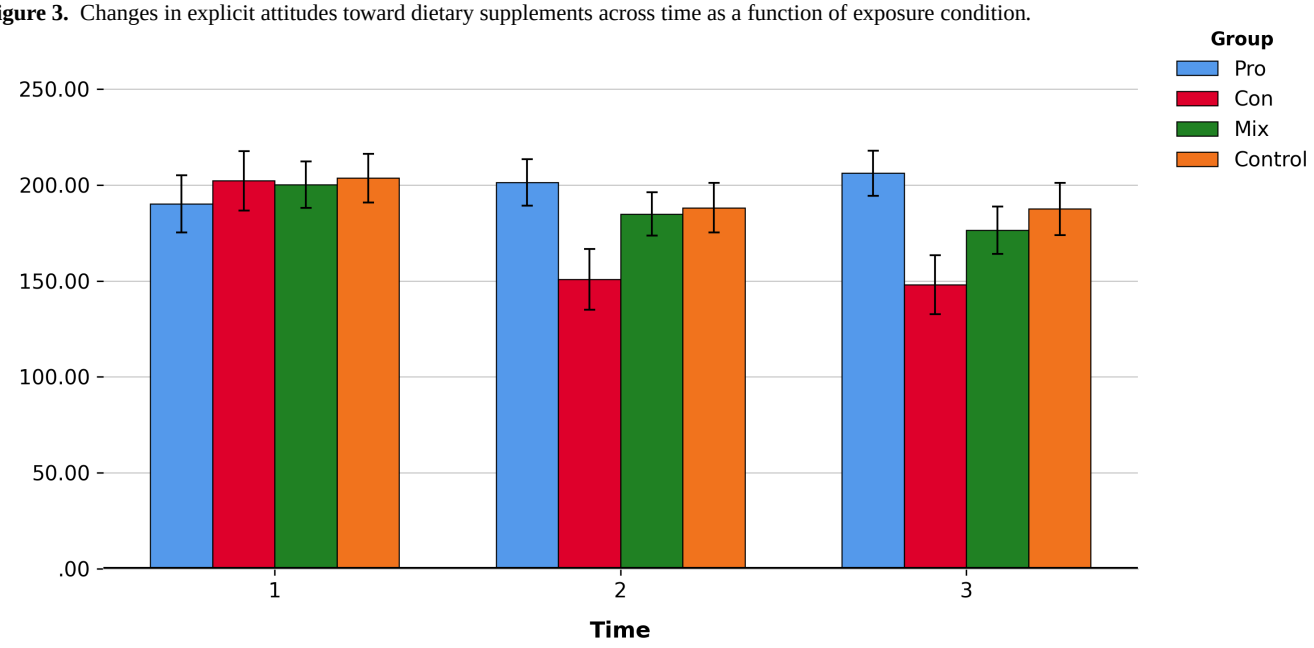
Preregistered within-group comparisons, illustrated in Figure 3, showed that participants in the PRO condition developed significantly more favorable explicit attitudes from baseline to the final assessment (MΔ=−14.86, SE=5.93, d=0.30; $P_{1\text{-tailed}}$=.02). Participants in the CON condition showed the opposite pattern, reporting less favorable attitudes over time (M_Δ=−53.42, SE=7.35, d=1.02; $P_{1\text{-tailed}}$<.001). The MIX group exhibited a significant change consistent with the most recent exposure (ie, negative article, M_Δ=−24.56, SE=6.77, d=0.50; P=.001), whereas the control group showed no significant change (M_Δ=−14.74, SE=6.63, d=0.29; P=.08). H1b, H2b, and H3b are therefore supported by the collected data.

Between-group comparisons at the final assessment (T2) confirmed this pattern. The CON group reported significantly lower explicit attitudes than the control group (M_Δ=−32.84, SE=10.75, d=0.32; P=.02). Explicit attitudes in the PRO group were not significantly different from the control group (M_Δ=19.24, SE=9.29, d=0.20; P=.22) but were significantly higher than in the CON group (M_Δ=52.08, SE=10.03, d=0.52; P<.001), and than in the MIX group (M_Δ=27.07, SE=9.38, d=0.29; P=.03). No other between-group differences were significant. Data provided clear support for H5b and partial support for H4b.

Together, these results indicate that repeated exposure to polarized information altered explicit, but not implicit, attitudes toward dietary supplements, and that these changes occurred in the direction of the informational valence of the exposure.

XSL•FO
RenderX

Figure 3. Changes in explicit attitudes toward dietary supplements across time as a function of exposure condition.



In order to test whether baseline attitudes moderated the relationship between group assignment and attitude change, 2 multiple regression models were conducted. The model having implicit attitude change as criterion was statistically significant ($F_{11,\ 216}$=16.80, R^2=0.46; *P*<.001), but the only significant predictor was baseline implicit attitudes (*b*=–.84, *SE*=.11; *P*<.001). For the model having explicit attitude change as criterion, neither implicit ($F_{3,\ 216}$=0.04, ΔR^2=0.00; *P*=.99) nor explicit attitudes ($F_{3,\ 216}$=1.94, ΔR^2=0.02; *P*=.12) moderated the effect. Thus, the collected data were insufficient to support H7a and H7b.

Given that the mixed-design ANCOVA revealed no significant interaction between time and group for implicit attitudes and no consistent within-group changes, mediation analyses involving implicit attitudes as the outcome were not pursued. Therefore, H8a and H9a are not supported.

For explicit attitudes, the simple mediation model testing whether time spent reading mediated the relationship between exposure condition and explicit attitudes at T2 failed to meet the assumptions for mediation, as the path from group to reading time was nonsignificant ($F_{3,\ 224}$=0.26, R^2=0.03; *P*=.12). The collected data were not sufficient to support H8b.

To further examine whether engagement effects depended on participants' initial evaluations, 2 preregistered moderated mediation models (PROCESS Model 7 [42]) were tested. When baseline implicit attitudes were included as a moderator of the Group to Reading (a) path, the model was still not significant ($F_{7,\ 220}$=0.98, R^2=0.03; *P*=.45). The analysis with baseline explicit attitudes as the moderator also yielded a nonsignificant model ($F_{7,\ 220}$=1.06, R^2=0.03; *P*=.39). Therefore, H9 is not supported by the collected data. A graphical representation of the tested moderated mediation models is provided in Figure S1 of Multimedia Appendix 1.

To further examine the changes in explicit attitudes, the 3 items composing the composite score (ie, perceived efficiency, perceived harmfulness, and recommendability of dietary supplements) were analyzed separately using 4×3 mixed-design ANCOVAs with the same covariates as the main model. These analyses were treated as exploratory as they have not been preregistered. Figure 4 presents the changes in perceived efficiency, harmfulness, and recommendability across time and exposure conditions.

The multivariate test revealed a significant time × group interaction ($F_{18,\ 1344}$=6.27, η^2_p=0.078; *P*<.001). At the univariate level, all 3 attitude items showed significant time × group interactions (all *Ps*<.001, *partial* η^2s=0.09-0.14).

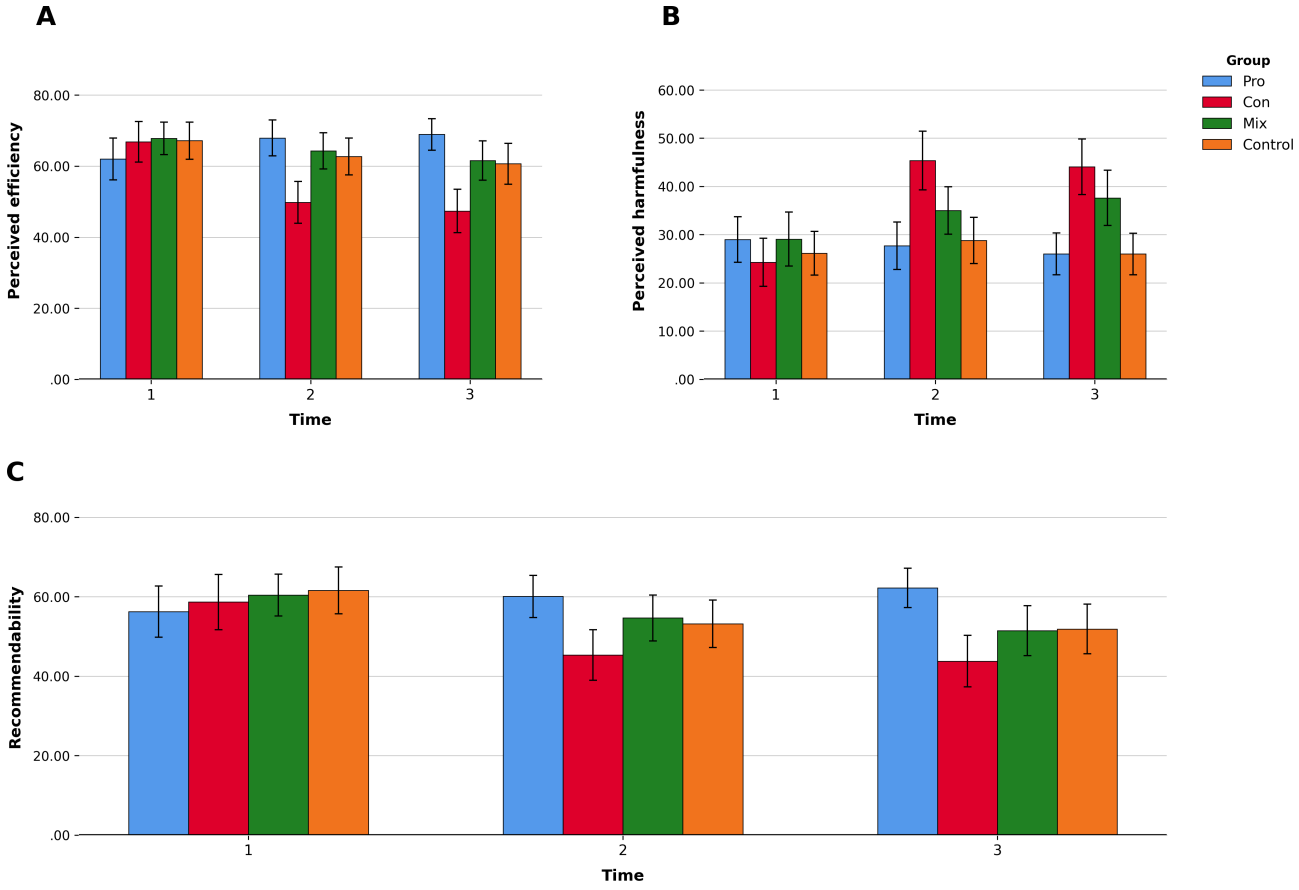
Perceived efficiency increased significantly in the PRO group from T0 to T2 (M_Δ=6.96, *d*=0.42; *P*=.02) and decreased significantly in the CON group (M_Δ=−19.49, *d*=0.83; *P*<.001). At T2, participants in the CON group rated supplements as less efficient than those in the other groups (*Ps*≤.006, *ds*=0.64-1.04), which did not differ from each other.

Harmfulness increased significantly in the CON group from T0 to T2 (M_Δ=19.82, *d*=0.84; *P*<.001) and to a smaller extent in the MIX group (M_Δ=8.56, *d*=0.48; *P*=.004). At T2, these groups judged supplements as more harmful than both the PRO (*Ps*≤.007, *ds*=0.60 and 0.94) and control groups (*Ps*≤.01, *ds*=0.61 and 0.94).

Recommendability decreased significantly in the CON group from T0 to T2 (M_Δ=−14.88, *d*=0.77; *P*<.001) and to a lesser degree in the MIX group (M_Δ=−9.00, *d*=0.52; *P*=.008). At T2, the CON group reported lower recommendability than the PRO group (*P*<.001; *d*=0.82).

XSL•FO
RenderX

Figure 4. Exploratory analyses of the components of explicit attitudes toward dietary supplements. (A) perceived efficiency, (B) perceived harmfulness, and (C) recommendability.



Discussion

Principal Findings

This study investigated how repeated exposure to polarized HRC influences explicit and implicit attitudes toward dietary supplements. Over 2 weeks of daily exposure to full-length articles, participants' explicit attitudes shifted in the direction of the presented valence (H1b, H2b, H3b, partial H4b, H5b), while implicit attitudes remained stable. The expected moderation by initial attitudes (H7), mediation through engagement (ie, time spent reading, H8), and the hypothesized moderation by initial attitudes of the mediated effect (moderated mediation, H9) were not supported. Exploratory analyses of the 3 explicit attitude components (ie, perceived efficiency, harmfulness, and recommendability) showed distinct valence-dependent changes, indicating that repetition differentially shaped specific evaluative dimensions rather than producing uniform shifts.

Consistent with predictions derived from the illusory truth effect [13,16] and the mere exposure effect [12,14], repeated exposure to polarized HRC systematically modified explicit evaluations. Participants repeatedly exposed to favorable articles developed more positive explicit attitudes toward dietary supplements (H1b), while those exposed to negative articles developed more negative attitudes (H2b). The control group, who read unrelated content, showed no change, confirming that attitude shifts were due to informational valence rather than time or repeated testing. Repeated reading of congruent information, even when spread across distinct articles, appears sufficient to increase the perceived validity and desirability of the presented perspective.

These results align with prior findings that repetition enhances processing fluency, which is then misattributed to truth [26]. Because all articles were matched in length, structure, and evidential strength (no difference in trust ratings between groups, $F_{3, 224}$=0.85; P=.47), the observed changes in explicit attitudes are unlikely to reflect persuasion via argument quality and instead point to fluency-driven propositional updating. The pattern observed in the mixed-exposure group provides further support for this interpretation. Participants in this condition consistently displayed negative shifts in their explicit attitudes (H3b), which, due to a technical ordering error, was the most recently encountered valence before each assessment point, suggesting that in contexts of informational ambivalence, recency effects can outweigh the cumulative balance of prior exposures. This is consistent with models of sequential belief updating showing that when conflicting information is presented sequentially and is comparable in evidentiary strength, recency exerts a disproportionate influence on judgments [43]. Nonetheless, because negative information is oriented toward more often and weighted more strongly than positive information (the "bad is stronger than good" phenomenon [44]), the present results cannot determine whether the observed pattern was driven primarily by recency, negativity dominance, or an interaction between the 2. Indeed, recent negative information may exert particularly strong influence because its temporal proximity and negative valence jointly increase its

XSL•FO
RenderX

impact on judgment, making these mechanisms difficult to disentangle within the current design.

The exploratory analyses revealed that the CON group exhibited the strongest and most consistent shifts across all three evaluative dimensions: efficiency, harmfulness, and recommendability. The MIX group showed smaller but still significant increases in perceived harmfulness and decreases in recommendability, which suggests that repetition produces cumulative effects even when exposures are heterogeneous. These changes may reflect partial counterbalancing from the positive articles, the recency of the last negative exposure, or both. Because both harmfulness and recommendability displayed a consistent trend from T0 to T1 to T2 (with a significant change from T0 to T1 only for harmfulness, M_{Δ}=–6.49; *P*=.04), the data suggest an additive negative exposure effect that was tempered, though not eliminated, by intermittent positive information. By contrast, efficiency increased modestly in the PRO group but was only slightly higher than in the control group at T2 (H4b, M_{Δ}=8.76; *P*=.12). This suggests that positive claims about supplement benefits may be less impactful than warnings about potential harm. This asymmetry aligns with extensive evidence showing that negative information typically produces stronger and more persistent evaluative shifts than positive information [45,46]. Practically, these findings suggest that health-related judgments, especially those concerning perceived harm and social recommendation, are highly susceptible to repeated negative framing, whereas positive framing appears weaker and more fragile, but nevertheless offers some degree of protection when mixed in an environment saturated with negative exposure.

The observed effects extend prior findings based on brief or decontextualized statements [23,24], showing that fluency-driven evaluative shifts are not limited to aphoristic or slogan-like content. Instead, similar shifts arise from repeated exposure to full-length, naturalistic articles, suggesting that real-world news formats, despite their complexity and narrative structure, can still produce the same fundamental repetition-based distortions in explicit evaluations.

As online news consumption is driven primarily by negativity [47], HRC may be especially vulnerable to this dynamic due to its high personal relevance and the emotional weight carried by risk-related information. When algorithms amplify engagement signals, users are more likely to encounter repeated negatively framed health content, which our findings suggest can disproportionately influence evaluations of harmfulness and recommendability. In real-world settings where user interests and algorithmic curation jointly shape exposure, this may create a self-reinforcing loop of negative information. Practically, this underscores the value of deliberately incorporating balanced or positively framed health information. Further research on recency and timing is required to determine whether such content should be introduced strategically before important health decisions, in order to counteract the cumulative impact of negativity-heavy feeds and support more proportionate and evidence-aligned judgments.

In contrast to the clear shifts observed in explicit evaluations, implicit attitudes toward dietary supplements remained unchanged across the 2-week exposure period (H1-6a). This pattern aligns with contemporary evidence showing that changes in implicit evaluations depend strongly on the type, diagnosticity, and consistency of new information, the strength and age of prior attitudes, and the timescale of exposure [48]. Implicit attitudes are most likely to shift when the target is novel or weakly evaluated, when new information is strongly diagnostic or affectively potent, or when repeated pairings occur under tightly controlled conditions.

The present attitude object, dietary supplements, is familiar, frequently discussed, and carries a predominantly positive prior valence for our target population [49]. Implicit attitudes toward well-known targets tend to show reduced malleability and may be buffered by entrenched associative structures [50]. Additionally, because the articles were balanced in structure and intentionally nonsensational, the affective input may have been insufficient to trigger strong evaluative conditioning effects [51]. Repeated article exposure may therefore have lacked the associative shock or diagnostic weight needed to update established evaluations [33]

Another possibility is methodological: the IAT may be less sensitive to short-term changes when the target category is broad, multifaceted, or already strongly encoded [49]. Tasks such as the affect misattribution procedure (AMP [52]) or evaluative priming (EPT [53]) can detect subtler shifts in implicit evaluations of familiar objects and may therefore be better suited for future work on HRC.

Taken together, these findings suggest that while explicit evaluations of dietary supplements readily shift under repeated exposure to polarized information, implicit attitudes may require either more potent affective signals, more diagnostic or surprising information, or longer and more diverse exposure histories to exhibit measurable change. In real-world settings, such changes may arise only when repetition is accompanied by credible social endorsement, identity relevance, or experiential consequences. Dietary supplements, being familiar and frequently discussed in public discourse, may therefore represent an attitude domain in which everyday news exposure alone is unlikely to yield measurable changes in automatic evaluations.

Contrary to predictions, the hypothesized mediation by engagement, operationalized as time spent reading the assigned articles, was not supported (H8). Participants did not differentially invest time depending on the informational valence of the content. Similarly, the preregistered moderated mediation models testing whether baseline implicit or explicit attitudes moderated the relationship between group-specific exposure and engagement were nonsignificant (H9). This indicates that prior beliefs did not systematically shape how long participants engaged with the material. Our expectation that counterattitudinal information would be perceived as more diagnostic, thereby eliciting deeper engagement and stronger propositional updating, was therefore not supported (H7). The lack of mediation by engagement is consistent with findings that exposure alone, not depth of processing, is sufficient to shift explicit evaluations, because illusory truth effects operate primarily through familiarity and fluency rather than

XSL•FO
RenderX

comprehension or elaboration [16,26]. However, our engagement measure was a coarse proxy, capturing only total reading time. More fine-grained measures such as eye-tracking, attention-based metrics (eg, attention allocation), or comprehension tests may better differentiate between superficial and elaborative processing in future work and offer clearer perspectives on their interaction with the illusory truth effect.

From a practical perspective, these findings suggest that stakeholders involved in health communication should consider patterns of repeated exposure in addition to the accuracy of individual messages. Public health agencies, healthcare professionals, and science communicators may benefit from disseminating evidence-based information repeatedly over periods of days or weeks, rather than relying on single-exposure campaigns. Examples include scheduled reposting of evidence-based content, recurring newsletter messages, and push notifications that direct users to trustworthy health information sources. Because repeated exposure influenced explicit evaluations even when participants read relatively lengthy articles, interventions aimed at improving public understanding of health topics may benefit from emphasizing sustained visibility rather than one-time corrections.

At the same time, platform designers and policymakers should recognize that recommendation systems may influence health attitudes through cumulative exposure patterns. In addition to content moderation and fact-checking, platforms could monitor whether users are repeatedly exposed to predominantly favorable or unfavorable health content and test strategies that increase informational diversity. For example, recommendation systems could periodically introduce evidence-based content presenting alternative perspectives, while A/B testing could be used to compare the effects of different exposure frequencies and content distributions on user attitudes and engagement. Such approaches may help reduce the formation of health-related echo chambers while preserving access to accurate information.

Limitations and Further Directions

Although several methodological limitations and directions for further research have already been noted in interpreting the findings, a few broader limitations and directions warrant explicit consideration.

This study sought to improve ecological validity in the investigation of repeated exposure by using full-length, naturalistic articles rather than brief statements or isolated claims. Although this approach provides a closer approximation to real-world health communication, the exposure environment still differed from genuine online news ecosystems. Participants read a single text-only article per day, whereas everyday exposure is rapid, interleaved, multimodal, and often shaped by social cues such as likes, comments, or peer sharing. Future research should incorporate more dynamic presentations, such as simulated algorithmic feeds, social endorsement signals, or mixed-content timelines, in order to capture the attentional competition and contextual layering characteristic of actual digital media environments.

A second limitation concerns the MIX condition. Although it was designed to provide balanced exposure, a technical error produced a fixed sequence in which negative content always immediately preceded each assessment. This systematic recency constraint limits our ability to interpret whether attitude trajectories reflected cumulative exposure, recency-driven updating, or an interaction of the two. Fully counterbalanced designs and stimuli that integrate both positive and negative information within the same article would allow future studies to disentangle additive effects from recency and negativity biases more precisely. A particularly informative design would orthogonally manipulate overall exposure balance and final exposure valence, for example by comparing sequences with the same number of positive and negative articles but ending either with a positive or a negative article. This would clarify whether mixed exposure effects are driven by cumulative balance, the valence of the most recent exposure, or whether negative final exposures exert disproportionate influence compared with positive final exposures.

A third limitation is that the sample consisted primarily of university students, which restricts generalizability. Young adults differ from older or more medically vulnerable populations in terms of prior familiarity with dietary supplements [49], susceptibility to misinformation [54], and online consumption patterns. Future work should recruit more diverse samples, including older adults, individuals with chronic health conditions, and frequent consumers of HRC on social media platforms.

Another limitation is that this study focused exclusively on attitudinal outcomes and did not assess behavioral consequences. Although repeated exposure influenced explicit evaluations of dietary supplements, it remains unclear whether these changes would translate into actual health-related behaviors. This question is particularly important given the absence of changes in implicit attitudes, which are often associated with more automatic forms of behavior. Future studies should incorporate behavioral measures, such as purchasing intentions, sharing intentions (eg, willingness to share supplement-related articles with friends or social media contacts), supplement consumption, or behavioral preference tasks. For example, participants could choose between a chance to win dietary supplements or an alternative reward of comparable value (eg, cereal bars), or between receiving a guide to high-quality supplement producers and a guide to healthy cooking. In addition, follow-up assessments conducted 1 and 3 months after the intervention could examine whether exposure predicts subsequent information-seeking behavior, supplement use, purchasing decisions, or the sharing of supplement-related content. Such measures would help determine whether repeated exposure to polarized HRC translates into meaningful behavioral outcomes.

A final limitation concerns the duration of the exposure period. Although the 2-week design allowed us to examine attitude trajectories at multiple assessment points while maintaining participant engagement with a demanding daily exposure protocol, it remains unclear whether the observed changes in explicit attitudes would persist, accumulate, or diminish over longer periods of exposure. Future studies should examine longer exposure schedules and include follow-up assessments to determine the durability of these effects and their relevance to real-world information environments.

XSL•FO
RenderX

Conclusions

The present findings show that repeated exposure to polarized HRC can shift explicit but not implicit attitudes, even when the content consists of long, naturalistic articles rather than isolated statements.

The explicit-implicit dissociation observed here offers a nuanced perspective on the potential risks of repeated exposure. On the one hand, the absence of changes in implicit attitudes suggests that underlying automatic associations remained stable, which may act as a buffer against drastic behavioral shifts. On the other hand, explicit attitudes guide reflective choices, risk assessments, and help-seeking behaviors, meaning that incremental changes in perceived harmfulness or recommendability could still influence consumer decisions, conversations with peers, or trust in official guidance.

In algorithmic news environments, where repetition is driven not by editorial intention but by engagement metrics, many such subtle explicit shifts may accumulate over time from various sources. This suggests that efforts to promote evidence-aligned health judgments should consider not only the accuracy of content but also the informational environments that shape familiarity and fluency. Supporting healthier digital ecosystems may require designing feeds that balance valence, introduce corrective or contextualizing information, and avoid reinforcing negativity-driven loops. In doing so, we may help preserve reflective health judgments without assuming that deeper evaluative change necessarily follows from repetition alone.

From an infodemiological perspective, the findings suggest that repeated exposure itself may represent an important mechanism through which online health information environments shape user evaluations. Consequently, exposure diversity and timing may deserve consideration alongside content accuracy in efforts to mitigate misinformation risks.

Acknowledgments

The authors used ChatGPT (OpenAI) to assist with language editing. The authors reviewed and take full responsibility for the content.

Funding

This research was supported by a PhD Scholarship provided by the Romanian Ministry of Education.

Data Availability

The datasets, stimuli, and code generated or analyzed during this study are available in the Open Science Framework repository [55] alongside the study preregistration.

Authors' Contributions

Conceptualization: ES, AO
Data curation: ES
Formal analysis: ES
Funding acquisition: AO, ES
Investigation: ES
Methodology: ES, AO
Project administration: ES, AO
Resources: AO
Supervision: AO
Validation: AO
Visualization: ES
Writing – original draft: ES
Writing – review and editing: ES (lead), AO (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Methodological information, descriptive statistics, and additional analyses.
[DOCX File, 75 KB - infodemiology_v6i1e88632_app1.docx]

Multimedia Appendix 2
Full text of the article stimuli used in the repeated exposure intervention.
[DOCX File, 119 KB - infodemiology_v6i1e88632_app2.docx]

XSL•FO
RenderX

References

1. In changing news landscape, even television is vulnerable. Pew Research Center. 2012. URL: https://www.pewresearch.org/journalism/2012/09/27/in-changing-news-landscape-even-television-is-vulnerable [accessed 2026-07-08]
2. Ali M, Sapiezynski P, Bogen M, Korolova A, Mislove A, Rieke A. Discrimination through optimization: how Facebook's ad delivery can lead to skewed outcomes. In: Proceedings of the ACM on Human-Computer Interaction 2019, Vol. 3:1-30. [doi: 10.1145/3359301]
3. Pariser E. How the new personalized web is changing what we read and how we think. In: The Filter Bubble: Penguin; 2011.
4. Nguyen CT. Echo chambers and epistemic bubbles. Episteme 2020 Jun;17(2):141-161. [doi: 10.1017/epi.2018.32]
5. Nyhan B, Settle J, Thorson E, et al. Like-minded sources on Facebook are prevalent but not polarizing. Nature 2023 Aug;620(7972):137-144. [doi: 10.1038/s41586-023-06297-w] [Medline: 37500978]
6. Guess AM, Malhotra N, Pan J, et al. How do social media feed algorithms affect attitudes and behavior in an election campaign? Science 2023 Jul 28;381(6656):398-404. [doi: 10.1126/science.abp9364] [Medline: 37498999]
7. Robertson RE, Green J, Ruck DJ, Ognyanova K, Wilson C, Lazer D. Users choose to engage with more partisan news than they are exposed to on Google search. Nature 2023 Jun;618(7964):342-348. [doi: 10.1038/s41586-023-06078-5] [Medline: 37225979]
8. Budak C, Nyhan B, Rothschild DM, Thorson E, Watts DJ. Misunderstanding the harms of online misinformation. Nature 2024 Jun;630(8015):45-53. [doi: 10.1038/s41586-024-07417-w] [Medline: 38840013]
9. Bhattacharya S, Singh A. Unravelling the infodemic: a systematic review of misinformation dynamics during the COVID-19 pandemic. Front Commun 2025;10:1560936. [doi: 10.3389/fcomm.2025.1560936]
10. Himelboim I, Borah P, Lee DKL, et al. What do 5G networks, Bill Gates, Agenda 21, and QAnon have in common? Sources, distribution, and characteristics. New Media & Society 2024 Oct;26(10):6019-6039. [doi: 10.1177/14614448221142800]
11. Dujeancourt E, Garz M. The effects of algorithmic content selection on user engagement with news on Twitter. The Information Society 2023 Oct 20;39(5):263-281. [doi: 10.1080/01972243.2023.2230471]
12. Zajonc RB. Attitudinal effects of mere exposure. J Pers Soc Psychol 1968;9(2, Pt.2):1-27. [doi: 10.1037/h0025848]
13. Hasher L, Goldstein D, Toppino T. Frequency and the conference of referential validity. J Verbal Learning Verbal Behav 1977 Feb;16(1):107-112. [doi: 10.1016/S0022-5371(77)80012-1]
14. Bornstein RF, Craver-Lemley C. Mere exposure effect. In: Cognitive Illusions: Routledge; 2022:241-258. [doi: 10.4324/9781003154730-18]
15. Udry J, Barber SJ. The illusory truth effect: a review of how repetition increases belief in misinformation. Curr Opin Psychol 2024 Apr;56:101736. [doi: 10.1016/j.copsyc.2023.101736] [Medline: 38113667]
16. Fazio LK, Rand DG, Pennycook G. Repetition increases perceived truth equally for plausible and implausible statements. Psychon Bull Rev 2019 Oct;26(5):1705-1710. [doi: 10.3758/s13423-019-01651-4] [Medline: 31420808]
17. De Keersmaecker J, Dunning D, Pennycook G, et al. Investigating the robustness of the illusory truth effect across individual differences in cognitive ability, need for cognitive closure, and cognitive style. Pers Soc Psychol Bull 2020 Feb;46(2):204-215. [doi: 10.1177/0146167219853844] [Medline: 31179863]
18. Stanley ML, Yang BW, Marsh EJ. When the unlikely becomes likely: qualifying language does not influence later truth judgments. J Appl Res Mem Cogn 2019;8(1):118-129. [doi: 10.1016/j.jarmac.2018.08.004]
19. Bell R, Nadarevic L, Mieth L, Buchner A. The illusory-truth effect and its absence under accuracy-focused processing are robust across contexts of low and high advertising exposure. Cogn Res Princ Implic 2025 May 13;10(1):21. [doi: 10.1186/s41235-025-00628-3] [Medline: 40358856]
20. Hassan A, Barber SJ. The effects of repetition frequency on the illusory truth effect. Cogn Res Princ Implic 2021 May 13;6(1):38. [doi: 10.1186/s41235-021-00301-5] [Medline: 33983553]
21. Udry J, White SK, Barber SJ. The effects of repetition spacing on the illusory truth effect. Cognition 2022 Aug;225:105157. [doi: 10.1016/j.cognition.2022.105157] [Medline: 35533415]
22. Garcia-Marques T, Silva RR, Reber R, Unkelbach C. Hearing a statement now and believing the opposite later. J Exp Soc Psychol 2015 Jan;56:126-129. [doi: 10.1016/j.jesp.2014.09.015]
23. Morgan JC, Cappella JN. The effect of repetition on the perceived truth of tobacco-related health misinformation among U.S. adults. J Health Commun 2023 Mar 4;28(3):182-189. [doi: 10.1080/10810730.2023.2192013] [Medline: 36938585]
24. Unkelbach C, Speckmann F. Mere repetition increases belief in factually true COVID-19-related information. J Appl Res Mem Cogn 2021;10(2):241-247. [doi: 10.1016/j.jarmac.2021.02.001]
25. Winkielman P, Schwarz N, Fazendeiro TA, Reber R. The hedonic marking of processing fluency: implications for evaluative judgment. In: Musch J, Klauer KC, editors. The Psychology of Evaluation: Affective Processes in Cognition and Emotion: Lawrence Erlbaum Associates Publishers; 2003:189-217. [doi: 10.4324/9781410606853-14]
26. Unkelbach C, Greifeneder R. A general model of fluency effects in judgment and decision making. In: Unkelbach C, Greifender R, editors. The Experience of Thinking: How the Fluency of Mental Processes Influences Cognition and Behaviour: Psychology Press; 2013:11-32. [doi: 10.4324/9780203078938-3]
27. Wilson TD, Lindsey S, Schooler TY. A model of dual attitudes. Psychol Rev 2000 Jan;107(1):101-126. [doi: 10.1037/0033-295x.107.1.101] [Medline: 10687404]

XSL•FO
RenderX

28. Greenwald AG, Banaji MR. Implicit social cognition: attitudes, self-esteem, and stereotypes. Psychol Rev 1995 Jan;102(1):4-27. [doi: 10.1037/0033-295x.102.1.4] [Medline: 7878162]
29. Rydell RJ, McConnell AR. Understanding implicit and explicit attitude change: a systems of reasoning analysis. J Pers Soc Psychol 2006 Dec;91(6):995-1008. [doi: 10.1037/0022-3514.91.6.995] [Medline: 17144760]
30. Ajzen I, Fishbein M, Albarracin D, Lohmann S. The influence of attitudes on behavior. In: Johnson B, editor. The Handbook of Attitudes, Volume 1: Basic Principles: Routledge:197-255. [doi: 10.4324/9781315178103-5]
31. Fazio RH, Olson MA. The MODE model: attitude-behavior processes as a function of motivation and opportunity. In: Sherman JW, Gawronski B, Trope Y, editors. Dual-Process Theories of the Social Mind: The Guilford Press; 2014:155-171.
32. Kurdi B, Banaji MR. Repeated evaluative pairings and evaluative statements: how effectively do they shift implicit attitudes? J Exp Psychol Gen 2017 Feb;146(2):194-213. [doi: 10.1037/xge0000239] [Medline: 28134543]
33. Cone J, Ferguson MJ. He did what? The role of diagnosticity in revising implicit evaluations. J Pers Soc Psychol 2015 Jan;108(1):37-57. [doi: 10.1037/pspa0000014] [Medline: 25365037]
34. Van Dessel P, Ye Y, De Houwer J. Changing deep-rooted implicit evaluation in the blink of an eye: negative verbal information shifts automatic liking of Gandhi. Soc Psychol Personal Sci 2019 Mar;10(2):266-273. [doi: 10.1177/1948550617752064]
35. Kurdi B, Mann TC, Ferguson MJ. Persuading the implicit mind: changing negative implicit evaluations with an 8-minute podcast. Soc Psychol Personal Sci 2022 Apr;13(3):688-697. [doi: 10.1177/19485506211037140]
36. Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, Evershed JK. Gorilla in our midst: an online behavioral experiment builder. Behav Res Methods 2020 Feb;52(1):388-407. [doi: 10.3758/s13428-019-01237-x] [Medline: 31016684]
37. What are dietary supplements? Buiced. 2022. URL: https://www.buiced.com/pages/dietary-supplements [accessed 2026-07-08]
38. Dietary supplements: what you need to know. Nova Southeastern University Florida. URL: https://nsunews.nova.edu/dietary-supplements-what-you-need-to-know/index.html [accessed 2026-07-08]
39. Preacher KJ, Rucker DD, Hayes AF. Addressing moderated mediation hypotheses: theory, methods, and prescriptions. Multivariate Behav Res 2007;42(1):185-227. [doi: 10.1080/00273170701341316] [Medline: 26821081]
40. Greenwald AG, McGhee DE, Schwartz JL. Measuring individual differences in implicit cognition: the implicit association test. J Pers Soc Psychol 1998 Jun;74(6):1464-1480. [doi: 10.1037//0022-3514.74.6.1464] [Medline: 9654756]
41. Greenwald AG, Nosek BA, Banaji MR. Understanding and using the implicit association test: I. An improved scoring algorithm. J Pers Soc Psychol 2003 Aug;85(2):197-216. [doi: 10.1037/0022-3514.85.2.197] [Medline: 12916565]
42. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach: Guilford Publications; 2017.
43. Hogarth RM, Einhorn HJ. Order effects in belief updating: the belief-adjustment model. Cogn Psychol 1992 Jan;24(1):1-55. [doi: 10.1016/0010-0285(92)90002-J]
44. Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD. Bad is stronger than good. Review of General Psychology 2001 Dec;5(4):323-370. [doi: 10.1037/1089-2680.5.4.323]
45. Soroka S, Fournier P, Nir L. Cross-national evidence of a negativity bias in psychophysiological reactions to news. Proc Natl Acad Sci U S A 2019 Sep 17;116(38):18888-18892. [doi: 10.1073/pnas.1908369116] [Medline: 31481621]
46. Watson J, van der Linden S, Watson M, Stillwell D. Negative online news articles are shared more to social media. Sci Rep 2024 Sep 16;14(1):21592. [doi: 10.1038/s41598-024-71263-z] [Medline: 39285221]
47. Robertson CE, Pröllochs N, Schwarzenegger K, Pärnamets P, Van Bavel JJ, Feuerriegel S. Negativity drives online news consumption. Nat Hum Behav 2023 May;7(5):812-822. [doi: 10.1038/s41562-023-01538-4] [Medline: 36928780]
48. Kurdi B, Charlesworth TES. A 3D framework of implicit attitude change. Trends Cogn Sci 2023 Aug;27(8):745-758. [doi: 10.1016/j.tics.2023.05.009] [Medline: 37270388]
49. Burcă V, Csorba LM, Munteanu FD, Rusu S. The Romanian consumers' perception regarding the safety of dietary supplements. J Public Health (Berl) 2022 Jun;30(6):1583-1602. [doi: 10.1007/s10389-020-01425-6]
50. Van Dessel P, Gawronski B, Smith CT, De Houwer J. Mechanisms underlying approach-avoidance instruction effects on implicit evaluation: results of a preregistered adversarial collaboration. J Exp Soc Psychol 2017 Mar;69:23-32. [doi: 10.1016/j.jesp.2016.10.004]
51. McConnell AR, Rydell RJ. The systems of evaluation model: a dual-systems approach to attitudes. In: Sherman JW, Gawronski B, Trope Y, editors. Dual-Process Theories of the Social Mind: The Guilford Press; 2014:204-218.
52. Payne BK, Cheng CM, Govorun O, Stewart BD. An inkblot for attitudes: affect misattribution as implicit measurement. J Pers Soc Psychol 2005 Sep;89(3):277-293. [doi: 10.1037/0022-3514.89.3.277] [Medline: 16248714]
53. Spruyt A, De Houwer J, Hermans D, Eelen P. Affective priming of nonaffective semantic categorization responses. Exp Psychol 2007;54(1):44-53. [doi: 10.1027/1618-3169.54.1.44] [Medline: 17341014]
54. Ma I, Sultan M, Kozyreva A, van den Bos W. Understanding the impact of misinformation on adolescents. Nat Hum Behav 2026 Jan;10(1):18-28. [doi: 10.1038/s41562-025-02338-8] [Medline: 41238751]
55. Epistemic bubbles: the effect of repeated exposure to polarized news on implicit and explicit attitudes. Open Science Framework (OSF). URL: https://osf.io/r7y8m [accessed 2026-07-08]

XSL•FO
RenderX

Abbreviations

AMP: affect misattribution procedure
ANCOVA: analysis of covariance
BootCI: bias-corrected bootstrapped confidence interval
CON: negative exposure condition/group
EPT: evaluative priming task
HRC: health-related content
IAT: Implicit Association Test
MIX: mixed exposure condition/group
PRO: positive exposure condition/group
VAS: visual analog scale

Edited by T Mackey; submitted 28.Nov.2025; peer-reviewed by C Sistiarani, J Malinowska; revised version received 14.Jun.2026; accepted 22.Jun.2026; published 27.Jul.2026.

Please cite as:
Secară EC, Opre NA
Impact of Repeated Exposure to Polarized Health-Related News on Explicit and Implicit Attitudes Toward Dietary Supplements: Online Experimental Study
JMIR Infodemiology 2026;6:e88632
URL: https://infodemiology.jmir.org/2026/1/e88632
doi:10.2196/88632

© Eugen-Călin Secară, Nicolae-Adrian Opre. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 27.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Quality, Reliability, and Dissemination of In Vitro Fertilization–Related Videos on Chinese Social Media: Cross-Sectional Analysis of 300 Short Videos

Xueyan Bai[1*], MD; Feng Guo[2*], BA; Dapeng Chu[1*], PhD

[1]Beijing Chao-Yang Hospital, Capital Medical University, Main Campus, Beijing Chao-Yang Hospital, 8 Gongren Tiyuchang Nanlu, Chaoyang District, Beijing, 100020, Beijing, China

[2]APUCH Innovation, Beijing, China

[*]all authors contributed equally

Corresponding Author:
Dapeng Chu, PhD
Beijing Chao-Yang Hospital, Capital Medical University, Main Campus, Beijing Chao-Yang Hospital, 8 Gongren Tiyuchang Nanlu, Chaoyang District, Beijing, 100020, Beijing, China

Related Article:

This is a corrected version. See correction statement: https://infodemiology.jmir.org/2026/1/e95354

Abstract

Background: Patients increasingly rely on short-video platforms for information regarding in vitro fertilization (IVF), yet the relationship between the scientific quality of this content and its algorithmic dissemination remains unclear.

Objective: This study aimed to assess the quality, reliability, and key drivers of dissemination of IVF-related short videos on major Chinese social media platforms.

Methods: A cross-sectional content analysis was conducted on 300 popular IVF-related videos (the top 100 results from each platform) retrieved from Douyin, Bilibili, and Xiaohongshu between January 10 and 15, 2025. Video quality and reliability were evaluated using the Global Quality Score and a modified DISCERN instrument. Predictors of video dissemination were identified using an Extreme Gradient Boosting machine learning model, with the number of "likes" serving as the primary outcome variable.

Results: Content produced by medical professionals demonstrated significantly higher quality and reliability (median mDISCERN 11.0, IQR 9.0-15.0) compared to non-medical sources (median mDISCERN 8.0, IQR 5.0-13.0; $P< .001$). However, the Extreme Gradient Boosting analysis identified the uploader's follower count as the most powerful predictor of video "likes." In contrast, quality metrics (Global Quality Score and modified DISCERN scores) had a negligible impact on dissemination.

Conclusions: In the current Chinese social media landscape, the dissemination of IVF-related videos is strongly associated with creator influence rather than scientific merit. This disconnect between engagement and quality poses a potential risk of misinformation, highlighting the need for medical professionals to adopt platform-native communication strategies to ensure that high-quality information reaches patients.

(JMIR Infodemiology 2026;6:e83900) doi:10.2196/83900

KEYWORDS

in vitro fertilization; social media; health communication; content quality; misinformation

Introduction

In vitro fertilization (IVF) offers profound hope to individuals and couples facing infertility, yet the journey is fraught with challenges. The complexity of the procedures, significant financial costs, and uncertain outcomes impose a substantial physiological, psychological, and economic burden on patients [1]. In navigating this demanding process, access to accurate, comprehensive, and understandable medical information is critical for informed decision-making, managing treatment expectations, and mitigating psychological distress [2-5]. Historically, this information was primarily disseminated by health care institutions. However, the digital era has precipitated a paradigm shift, with patients increasingly turning to the internet for more accessible and diverse sources of support [6-8]. While social media's impact on health behaviors is a global phenomenon, China's digital ecosystem offers a unique context for study. With the world's largest internet user base and high demand for assisted reproductive technology, platforms such

XSL•FO
RenderX

as Douyin (the Chinese counterpart to TikTok) provide a critical "natural laboratory" to understand algorithmic health communication patterns that are increasingly relevant worldwide.

In recent years, short-video platforms such as Douyin, Bilibili, and Xiaohongshu have emerged as dominant arenas for health information dissemination, distinguished by their algorithm-driven, highly engaging, and rapidly propagating nature [9,10]. While these platforms present unprecedented opportunities for medical education, they also introduce formidable challenges [11-13]. Unlike traditional medical websites, content generation is often spontaneous and lacks rigorous professional oversight, creating a "perfect storm of information" where quality is highly variable [14,15]. Furthermore, their personalized recommendation algorithms, while enhancing user experience, risk creating "information cocoons" that can amplify biased or inaccurate content [16,17], posing a potential hazard to patients seeking IVF treatment, especially concerning misinformation on reproductive health [18-20].

Despite growing analyses of social media health content, the IVF domain on Chinese short-video platforms remains understudied. Moreover, prior work has largely assessed quality in isolation, leaving unclear whether intrinsic quality or extrinsic platform factors (eg, creator influence) primarily drive dissemination. The few studies that have explored dissemination dynamics have relied on conventional linear models, which are ill-equipped to capture the complex, nonlinear factors driving content virality in sophisticated social networks. This leaves our understanding of the contemporary health information ecosystem fundamentally incomplete.

To address this gap, we conducted a cross-sectional analysis of the top-ranked IVF videos on Douyin, Bilibili, and Xiaohongshu. Our methodological approach consisted of three phases: (1) content analysis to classify uploader identity and topics, (2) assessment of information quality using the Global Quality Score (GQS) and the modified DISCERN (mDISCERN) instrument, and (3) machine learning analysis using Extreme Gradient Boosting (XGBoost) and Shapley Additive Explanations (SHAP) values to isolate independent predictors of video dissemination among metadata variables. We hypothesized that uploader influence (follower count), rather than content quality, would be the dominant predictor of engagement. The findings are intended to provide an evidence-based foundation for enhancing the effective communication of high-quality medical information and to offer actionable guidance for platforms, content creators, and public health authorities.

Methods

Study Design and Video Retrieval

A cross-sectional study was designed to evaluate the quality, reliability, and dissemination of IVF-related videos across three popular Chinese social media platforms: Xiaohongshu, Bilibili, and Douyin [21]. These platforms were selected based on their market dominance and distinct demographic profiles. Douyin (Chinese version of TikTok, (ByteDance, Beijing, China), with 766 million daily active users (DAUs) as of 2024 [22], represents China's short-video mainstream platform, with videos typically ranging from 15 to 60 seconds in length [23], Bilibili (Bilibili Inc., Shanghai, China), with an average of 104 million DAUs in 2024 [24], specializes in medium-to-long form video content (typically 3 - 30 min), with medium and long videos accounting for 70% of platform views [25]. The platform's user base is predominantly young, with nearly 70% of China's Generation Z population and an average user age of 25 years [24]. Xiaohongshu (RedNote, Xiaohongshu, Shanghai, China), with 143 million global DAUs by the end of 2024 [26], serves as a lifestyle-focused platform with a predominantly female user base (70% female) and functions as a primary search engine for lifestyle and health-related decisions among Chinese women [27].

Using the Chinese keyword "试管婴儿" (IVF), relevant videos were retrieved from each platform between January 10 and 15, 2025. To mitigate the influence of personalized recommendations, searches were conducted using newly created accounts with no prior viewing history. No filters or sorting mechanisms were applied, thereby simulating a typical user experience.

An initial systematic search on the 3 platforms identified 531 potentially relevant videos. These records were then screened for eligibility based on predefined inclusion and exclusion criteria. A total of 231 videos were excluded for the following reasons: duplicate content (n=127, 55.0%), non–Chinese-language content (n=14, 6.1%), being purely promotional without educational value (n=55, 23.8%), or having content irrelevant to the topic of IVF (n=35, 15.2%). This screening process yielded a final sample of 300 (56.5%) unique videos for analysis, comprising the top 100 eligible videos from each platform.

We used a quota sampling strategy based on platform search rankings. For each platform, videos were retrieved and screened sequentially, starting from the top-ranked search result. The screening process continued down the ranked list until a quota of 100 eligible videos meeting all inclusion and exclusion criteria was reached for each platform. This strategy ensures that the sample reflects the content most visible to users, as search rankings prioritize high-engagement content. These rankings are algorithmically driven and prioritize a synthesis of user engagement metrics (eg, likes, comments, and shares), topical relevance, and content freshness, thereby simulating the ecological search experience of a typical user. Selection was subject to the following criteria: (1) the video was in the Chinese language, (2) the video focused on IVF-related medical content, and (3) the video was publicly accessible. Videos were excluded if they contained (1) duplicate content, (2) pure advertisements without educational value, (3) videos unrelated to IVF, and (4) non–Chinese-language videos. For each video, basic information was documented, including title, upload date, duration, uploader identity, and engagement metrics (eg, likes, comments, shares, and saves). Regarding cross-platform posting, videos uploaded by the same creator to multiple platforms were treated as distinct analytical units. This approach was chosen because engagement metrics (eg, likes and comments) are platform specific and

XSL•FO
RenderX

reflect the unique algorithmic distribution and audience reaction within that specific ecosystem. However, intraplatform duplicates (the same video uploaded twice to the same platform) were excluded.

The diagram illustrates the selection process for the study. Initially, a keyword search for "试管婴儿" (IVF) was conducted on 3 platforms, namely, Douyin, Bilibili, and Xiaohongshu, identifying 231, 500, and 390 videos, respectively. Following the platforms' comprehensive ranking algorithms, the top-ranked videos were selected for screening (Douyin, n=116; Bilibili, n=120; and Xiaohongshu, n=122). During the screening phase, videos were excluded for being duplicates or thematically irrelevant (Douyin, n=16; Bilibili, n=20; and Xiaohongshu, n=22). Finally, 100 eligible videos from each platform were included, resulting in a total of 300 videos for the final analysis.

Video Classification

Videos were categorized by uploader and content type by 2 independent researchers, with disagreements resolved by a third reviewer, following established content analysis methodologies [28,29].

All included videos were systematically classified according to a predefined coding scheme focusing on two primary dimensions: uploader identity and content theme. The uploader of each video was categorized into one of five groups: medical professionals, which included verified IVF doctors, reproductive medicine institutions, or health care providers; health science communicators or key opinion leaders, defined as individuals known for disseminating health knowledge without formal medical credentials; patients and sharers, consisting of individuals sharing personal IVF treatment experiences; marketing promoters, identified as commercial entities promoting fertility services; and news and general content creators, such as media outlets providing general information.

Concurrently, the primary subject matter of each video was assigned to one of five content categories: medical knowledge, comprising scientific explanations, clinical guidelines, or technical information; fertility and lifestyle optimization, focusing on content related to lifestyle practices intended to enhance fertility; patient experience sharing, which covered personal narratives detailing individual IVF journeys; policy and ethical topics, which included discussions on regulations or social implications; and misleading or marketing content, which included promotional material or videos with verifiably inaccurate information.

Quality and Reliability Assessment

Video quality and reliability were assessed using the GQS [30,31] and the mDISCERN instrument [32,33].

The GQS is a 5-point scale evaluating overall quality, flow, and integrity of information (1=poor and 5=excellent) and has been validated in numerous studies of web-based health information [30,31]. The mDISCERN instrument was adapted from the original 16-item DISCERN tool [32,33] to specifically evaluate short-form digital health content. To accommodate the brevity of social media videos, the instrument was condensed into five core dimensions:

1. Reliability of information: assesses the evidentiary basis and accuracy of medical claims
2. Clarity of aims: evaluates whether the video's purpose and structure are clearly communicated
3. Relevance of sources: measures the transparency and authority of cited evidence (eg, clinical guidelines vs unverifiable anecdotes)
4. Balance and impartiality: assesses the extent of commercial bias or one-sided promotion
5. Presentation of uncertainty: evaluates the disclosure of risks, side effects, and biological variability

Items 9 to 15 of the original DISCERN tool, which focus on detailed treatment choices and shared decision-making, were excluded as they are rarely applicable to brief, nonconsultative video clips. Furthermore, the scoring system was modified from the original 1- to 5-point scale to a 0- to 5-point Likert scale. This modification allowed for a score of "0" to explicitly categorize content that was completely devoid of sources, reliability, or clear aims, which is a common characteristic of low-quality user-generated content.

Two experienced reproductive medicine specialists independently scored all videos. Initial disagreements were resolved through discussion to reach a consensus. If a consensus could not be reached, the score from the third senior specialist was used as the final arbitration score. Interrater reliability was assessed and found to be high (Cohen κ coefficient >0.80), indicating a strong degree of agreement [34,35].

Dissemination Analysis With XGBoost

An XGBoost regression model was used to identify factors influencing video dissemination, with the number of "likes" as the primary outcome variable [36,37]. Input features included platform, uploader category, content category, GQS score, mDISCERN score, video length, days since publishing, and the presence of background music or subtitles. To investigate the primary drivers of raw engagement, the model was first trained on the untransformed "likes" count.

Furthermore, to account for the highly skewed distribution of the "likes" variable and to build a more stable model for confirming feature contributions, a secondary analysis was performed using a logarithmic transformation (log1p) on the target variable before model training. This standard preprocessing technique helps mitigate the influence of outlier videos with extremely high engagement.

Hyperparameter tuning was performed using grid search with 3-fold cross-validation to optimize model performance [38,39]. The importance of features was assessed using SHAP values to provide transparent and intuitive insights into how each factor influences dissemination [40,41].

Statistical Analysis

All statistical analyses were conducted using R (version 4.2.3; R Foundation for Statistical Computing). Owing to the nonnormal distribution, nonparametric tests were used, including the Kruskal-Wallis test for multigroup comparisons and the Dunn test for post-hoc analysis. The Spearman correlation

XSL•FO
RenderX

coefficient was used to assess relationships between variables. Statistical significance was defined as $P<.05$.

Ethical Considerations

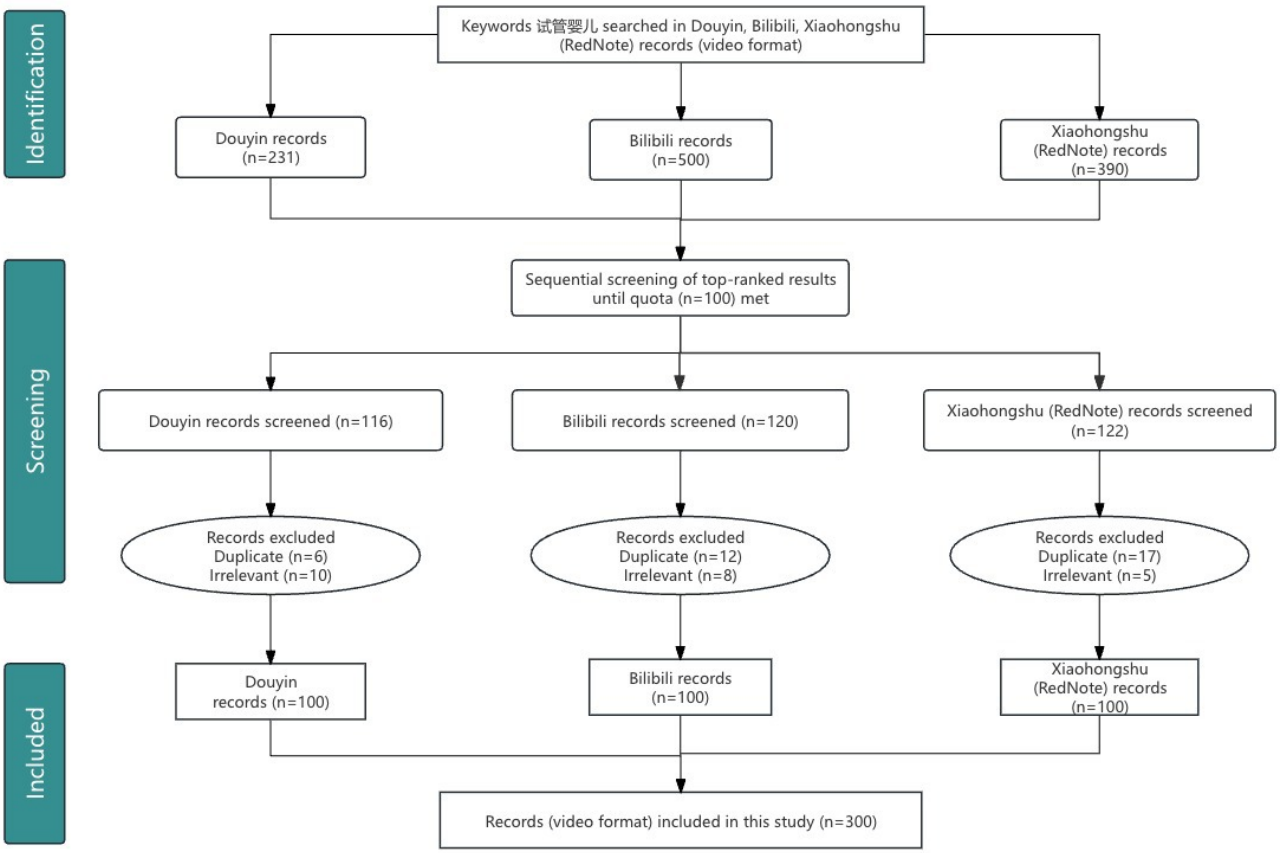
This study analyzed publicly available, user-generated IVF-related videos and their metadata from Chinese social media platforms. The research involved no interaction or intervention with individuals and did not collect, store, or process identifiable personal data. In accordance with institutional and national guidelines, the use of publicly accessible, aggregate data does not constitute human subjects research; therefore, ethics approval and consent to participate were not required.

Results

Overview

The complete selection process is detailed in the video selection flow diagram (Figure 1). The study's findings reveal a fundamental paradox in the digital health ecosystem: while the quality of IVF-related information is critically dependent on the professional identity of its creator, its dissemination is overwhelmingly governed by the creator's platform influence, not the quality of the content itself.

Figure 1. Flow diagram of the video identification, screening, and inclusion process.



Platform Ecosystems Exhibit Profound Heterogeneity

Analysis of 300 videos revealed significant heterogeneity in content strategy, user base, and engagement dynamics across platforms ($P<.001$; Table 1). On Douyin and Xiaohongshu, medical professionals were the dominant uploaders (87/100, 87%, and 89/100, 89% respectively), and content was primarily "medical knowledge" (71/100, 71% on both). In contrast, Bilibili featured a more diverse creator base, including patient sharers (28/100, 28%) and health science communicators (22/100, 22%), with "patient experience sharing" (31/100, 31%) being more prevalent (Table 2).

XSL•FO
RenderX

Table . Baseline of in vitro fertilization–relevant videos.

Characteristics	Douyin (n=100)	Bilibili (n=100)	Xiaohongshu (n=100)	*P* value
Likes, median (IQR)	539 (81 - 2986.5)	138 (16 - 844.5)	277.5 (66 - 925.25)	<.001
Saves, median (IQR)	145 (11-732)	78 (12-415)	134 (40 - 525.75)	.31
Comments, median (IQR)	30 (6.75 - 327.25)	17 (3 - 149.5)	51 (7.75 - 126.5)	.16
Shares, median (IQR)	161 (9.5 - 785)	56 (5-310)	96.5 (26.5 - 437.75)	.10
Days since uploading, median (IQR)	116 (100.75 - 155)	788.5 (293.25 - 1300.75)	218.5 (125 - 394.25)	<.001
Length, median (IQR)	41.5 (25.75 - 59)	325 (169-692)	53 (37.75 - 94)	<.001
Followers, median (IQR)	36000 (7469.25 - 273,500)	4679.5 (322.25 - 35,500)	8769.5 (3010.75 - 33,000)	<.001
Total video count, median (IQR)	364 (167-560)	201 (66.5 - 488.5)	309.5 (191.25 - 758.25)	.004
Type of video, n				<.001
Medical knowledge	71	43	71	
Fertility & lifestyle optimization	0	10	3	
Patient experience sharing	19	31	18	
Policy & ethical topics	8	12	3	
Misleading or marketing content	2	4	5	
Type of uploader, n				<.001
Medical professionals	87	24	89	
Health science communicators/medical key opinion leaders (KOLs)	1	22	0	
Patients and fertility journey sharers	7	28	5	
Marketing promoters	1	11	6	
News and general interest content creators	4	15	0	
BGM[a], n				<.001
Without BGM	38	46	19	
With BGM	62	54	81	
Subtitle, n				.02
Without subtitle	0	6	1	
With subtitle	99	94	99	
GQS[b] score, median (IQR)	2 (2-3)	2 (2-3)	3 (2-3)	.04
DISCERN score, median (IQR)	10 (7-14)	10 (6 - 14.25)	12 (9-15)	.02

[a]BGM: background music.

[b]GQS: Global Quality Score.

XSL•FO
RenderX

Table . Characteristics of uploaders and video content across the 3 platforms (N=300).

Category	Douyin (n=100), n (%)	Bilibili (n=100), n (%)	Xiaohongshu (RedNote) (n=100), n (%)	*P* value
Uploader profile				<.001
Medical professionals	87 (87)	24 (24)	89 (89)	
Patient voices	1 (1)	28 (28)	5 (5)	
Health communicators or KOLs[a]	7 (7)	22 (22)	6 (6)	
Marketing and promoters	4 (4)	11 (11)	0 (0)	
News and general content	1 (1)	15 (15)	0 (0)	
Video content type				<.001
Medical knowledge	71 (71)	43 (43)	71 (71)	
Patient experience	19 (19)	31 (31)	18 (18)	
Fertility and lifestyle	8 (8)	12 (12)	3 (3)	
Policy and ethics	0 (0)	10 (10)	3 (3)	
Misleading and marketing	2 (2)	4 (4)	5 (5)	

[a]KOL: key opinion leader.

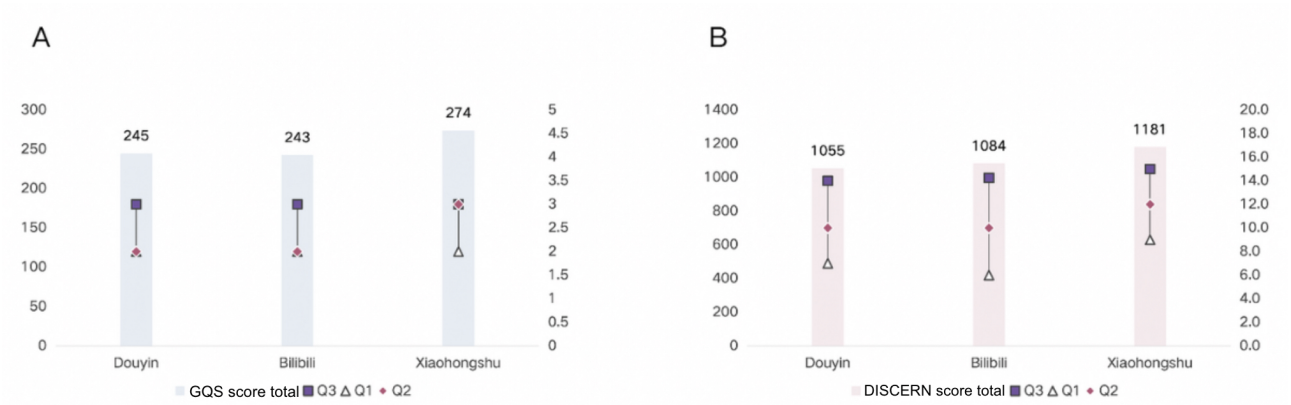
Video attributes also differed significantly. Bilibili hosted older (median 788.5, IQR 293.3-1300.8 d) and longer videos (median 325, IQR 169-692 s), whereas Douyin featured more recent, shorter-form content (median 41.5, IQR 25.8-59 s; *P*<.001 for both). Douyin creators had the largest median follower counts (n=36,000) and generated the highest median "likes" (n=539), significantly outperforming the other platforms. These baseline differences in creator influence and content strategy precede the analysis of dissemination drivers.

Quality and Reliability Assessment

Overall, video quality and reliability were moderate. Interrater agreement for the scoring was high (weighted κ_{GQS}=0.82, 95% CI 0.76 - 0.87; weighted $\kappa_{mDISCERN}$=0.79, 95% CI 0.72 - 0.85; International Code Council$_{GQS}$=0.90, 95% CI 0.86 - 0.93; International Code Council$_{mDISCERN}$=0.88, 95% CI 0.84 - 0.91).

Platform-level analysis showed that Xiaohongshu videos achieved a statistically significant higher quality, with a median GQS of 3.0 (IQR 2.0-3.0; *P*=.04) and a median mDISCERN score of 12.0 (IQR 9.0-15.0; *P*=.02) compared to Douyin and Bilibili (Figure 2).

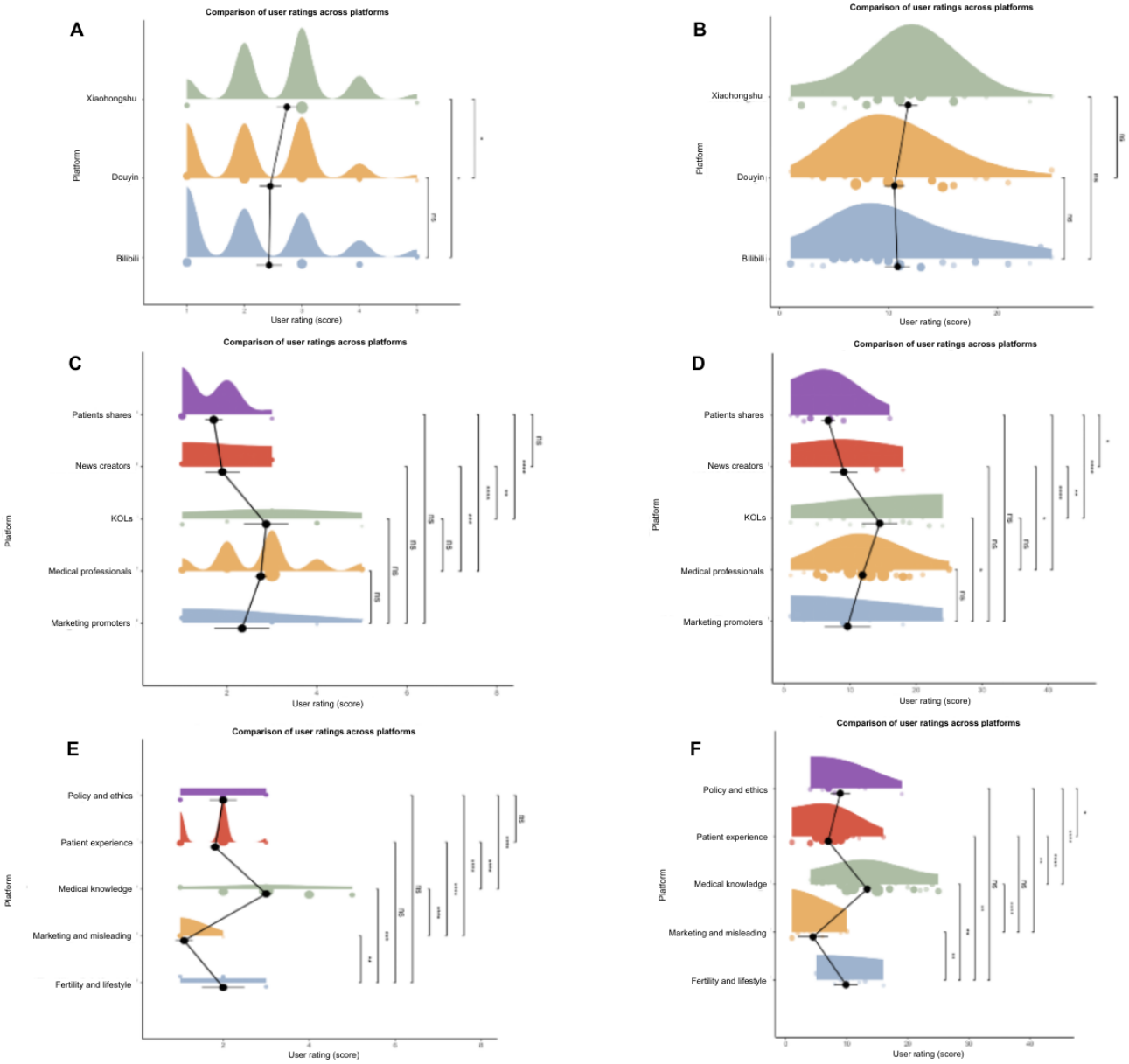
Figure 2. Comparison of Global Quality Score (GQS) and DISCERN score distributions across 3 platforms. (A) Distribution of GQS scores across the 3 platforms; (B) distribution of DISCERN scores across the 3 platforms.



Subgroup analysis revealed a strong association between uploader identity and content quality (*P*<.001). Our data highlights a distinct "competence hierarchy": videos from medical professionals consistently achieved the highest reliability scores (median mDISCERN 11.0, IQR 9.0-15.0), reflecting adherence to clinical guidelines. Conversely, patient sharers and marketing promoters scored significantly lower. While patient sharers provide emotional value, their content often lacked medical accuracy (median GQS 2.0, IQR 1.0-2.0), suggesting that the "lived experience" often comes at the expense of clinical precision. Videos categorized as "medical knowledge" were rated significantly higher than all other content themes (*P*<.001; Figure 3).

XSL•FO
RenderX

Figure 3. Comparison of video quality and reliability scores across different subgroups. (A) Distribution of Global Quality Score (GQS) scores across the 3 platforms; (B) distribution of DISCERN scores across the 3 platforms; (C) distribution of GQS scores across the 5 uploader types; (D) distribution of DISCERN scores across the 5 uploader types; (E) distribution of GQS scores across the 5 content types; and (F) distribution of DISCERN scores across the 5 content types. Each subplot displays the data distribution using a violin plot and a box plot. Asterisks indicate the level of statistical significance from pairwise statistical tests. *P<.05, **P<.01, ***P<.001. KOL: key opinion leader.



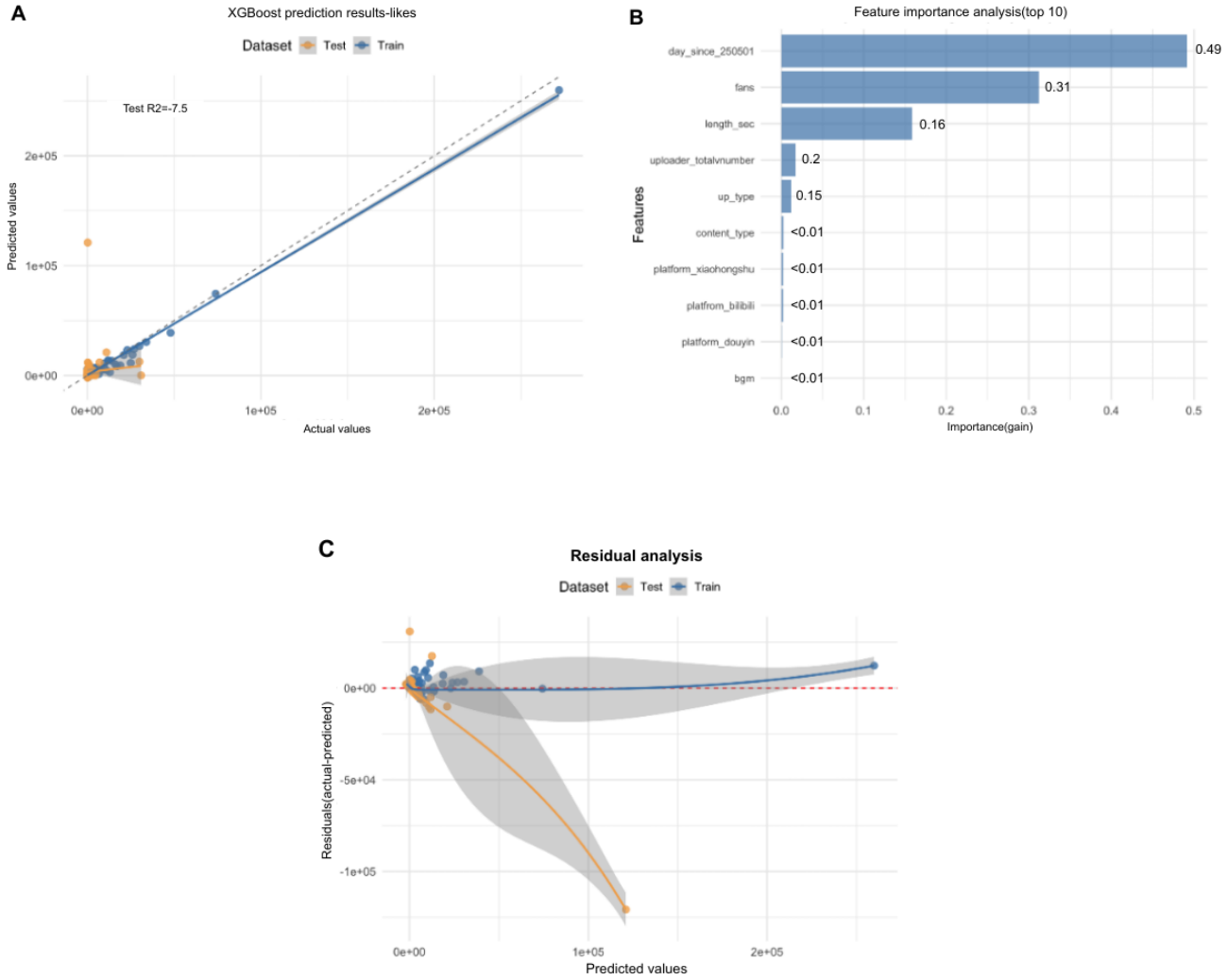
Predictors of Video Dissemination

To resolve the disconnect between content creation and consumption, an XGBoost machine learning model was developed to identify the primary drivers of video dissemination ("likes"). Initial bivariate correlation analysis found no significant relationship between a video's like count and its GQS or mDISCERN scores, providing a preliminary suggestion that quality was not a key factor for engagement.

The initial XGBoost model, trained directly on the untransformed "likes" count, yielded a negative R^2 of −7.5 (Figure 4A). This result confirms that raw social media engagement follows a nonlinear, heavy-tailed distribution that cannot be modeled by standard additive regression. Consequently, feature importance rankings from this initial model were disregarded to avoid spurious conclusions (Figure 4). We therefore relied exclusively on the secondary log1p-transformed model for identifying predictors.

XSL•FO
RenderX

Figure 4. Results of the XGBoost model interpreting the factors influencing video likes: (A) the goodness of fit of the XGBoost model (R^2=−7.5); (B) the importance of each variable on the number of video likes; and (C) the Shapley Additive Explanation summary plot showing the impact of important variables on the model output. XBoost: Extreme Gradient Boosting.

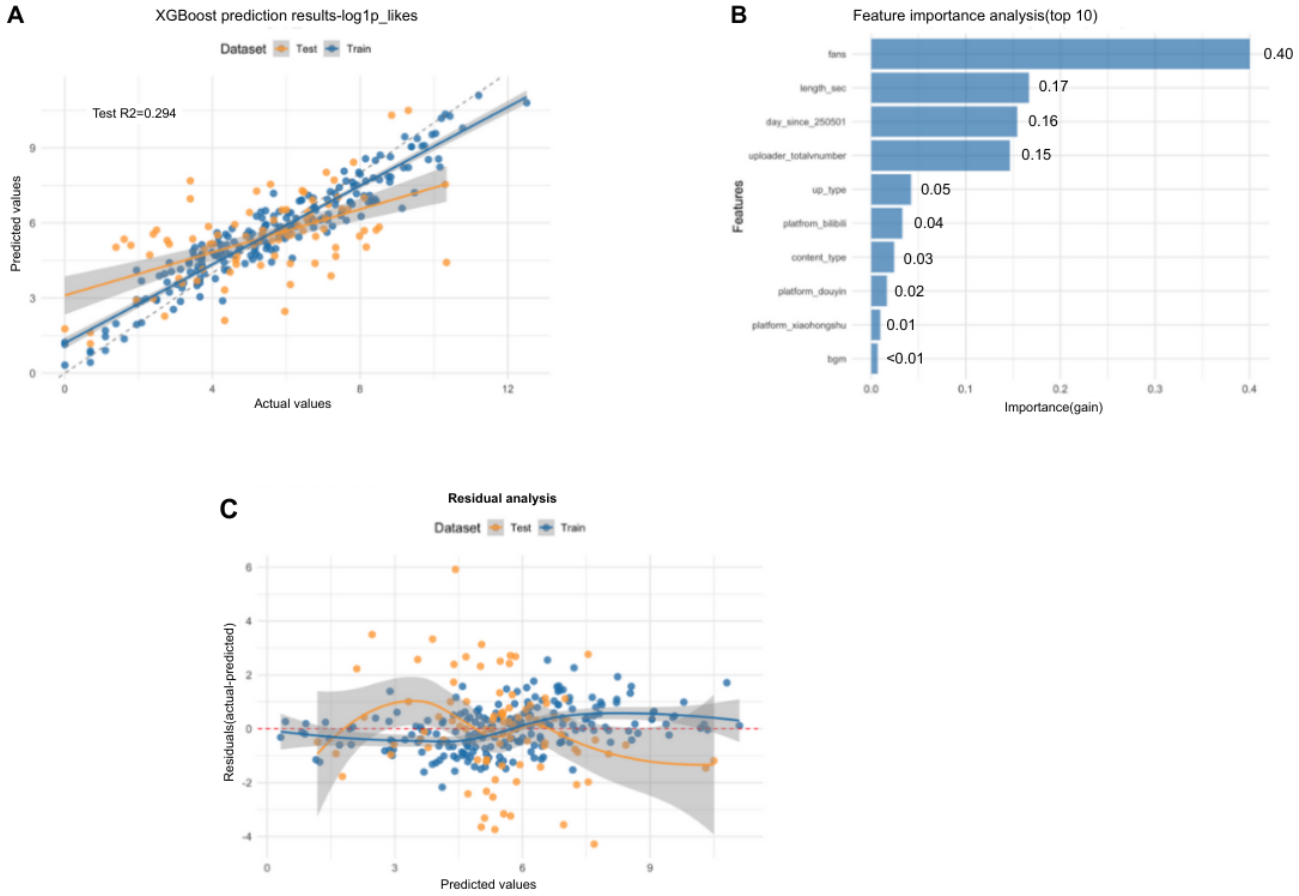


A closer examination of the SHAP summary plot (Figure 4) clarifies how these top features influence predictions. For the "fans" feature, a clear positive trend is visible: high feature values (represented by red dots) are strongly associated with high positive SHAP values, confirming that a larger follower count directly contributes to a higher prediction of likes. In contrast, for metrics such as GQS and mDISCERN, the points are clustered vertically around the zero-line with no discernible color gradient, visually confirming their negligible impact on the model's output and reinforcing the quality-impact gap.

To ensure these findings were not an artifact of the target variable's extreme skewness and to robustly validate the feature hierarchy, we developed a second XGBoost model by applying a logarithmic transformation (log1p) to the "likes" count. This standard data preprocessing step resulted in a more stable model with a positive predictive fit, achieving an R^2 of 0.294 on the test set (Figure 5A). The corresponding residual plot confirmed a more balanced and desirable error distribution (Figure 5C).

XSL•FO
RenderX

Figure 5. Performance and interpretation of the XGBoost model for predicting video likes. (A) Prediction results of the XGBoost model, plotting actual versus predicted values for the log1p_likes target. The coefficient of determination for the test set is R^2=0.294. (B) Feature importance analysis showing the top 10 predictors ranked by their gain score. (C) Residual analysis plotting residuals against predicted values to assess model fit and error distribution. XGBoost: Extreme Gradient Boosting.



Crucially, despite the improved model performance, the feature importance analysis remained remarkably consistent (Figure 5B). The uploader's follower count was once again the most dominant predictor. In stark contrast, the validated metrics for content quality and reliability—GQS and mDISCERN scores—remained at the bottom of the feature importance hierarchy, exerting negligible influence. This dual-model approach provides robust, 2-fold evidence that in the current algorithmic landscape, a video's reach is driven primarily not by its scientific merit but by the preexisting social capital of its creator.

Discussion

The "Quality-Impact Gap" in Digital Health

This study provides the first systematic evaluation of IVF-related health information on major Chinese short-video platforms, revealing a significant and troubling paradox at the heart of the modern digital health landscape. Our analysis empirically demonstrates a divergence between content quality and engagement. While quality metrics (GQS and mDISCERN) track closely with medical expertise, dissemination metrics (likes) track with uploader influence. This suggests that high-quality medical information does not automatically generate high engagement. This "quality-impact gap" [42] implies that scientific accuracy is not the primary driver of algorithmic visibility. This phenomenon is not merely an algorithmic quirk; it reflects a fundamental tension between the clinical nature of information and the socioemotional needs of patients. Patients navigating the arduous IVF journey are not just passive consumers of data; they are actively seeking hope, validation, and a sense of community. Consequently, low-quality but emotionally resonant content—such as unverified "miracle baby" testimonials—may be perceived as more valuable than dry, technically accurate explanations, leading to higher engagement.

Clinically, this gap risks therapeutic misconception and spending on unproven add-ons among IVF patients. To mitigate harm, clinicians and fertility centers should coproduce platform-native content—short, narrative-driven videos that embed evidence (success rates and indications or contraindications), use on-screen references, and include myth-fact segments—and deploy them via verified accounts with regular posting cadence and call-to-action links to authoritative resources [11,43].

Our central finding—that an uploader's follower count is the most potent predictor of reach—must be interpreted with nuance. In the fast-paced digital environment, follower count acts as a powerful cognitive heuristic for trust. Lacking the time or expertise to critically appraise every video, users subconsciously substitute "popularity" for "credibility," operating under the assumption that a large following implies authority and trustworthiness [43]. This dynamic, where social capital eclipses scientific capital, aligns perfectly with findings from Western

XSL•FO
RenderX

platforms such as YouTube [44,45] and extends recent analyses of other medical topics on Chinese platforms [46,47]. Our research confirms that this algorithmic prioritization of engagement over evidence is a universal feature of contemporary social media architecture [48,49], creating a global challenge for evidence-based health communication.

Platform Ecosystems: Expert-Led Versus Community-Driven

Furthermore, our 3-platform comparison revealed distinct "platform personalities" that shape this information flow. Douyin and Xiaohongshu function primarily as expert-led, knowledge-dissemination channels, yet they are still subject to the influencer dynamic. In contrast, Bilibili operates as a community-driven, experience-sharing hub (with 28% patient sharers vs ≤7% on other platforms), hosting longer narratives that fulfill patients' documented need for peer support [50]. While valuable for emotional well-being, this "experiential" content was found to be of significantly lower quality, posing a risk of normalizing anecdotal advice over clinical guidelines.

The implications of this ecosystem for patient care and public health are profound. For a vulnerable population already facing immense emotional and financial stress, the stakes are exceptionally high. Exposure to misinformation or low-quality content can foster therapeutic misconceptions, leading patients to pursue unproven and costly adjunct therapies. It can create unrealistic expectations about success rates, leading to deeper psychological distress when treatments fail. Moreover, the prevalence of marketing content masquerading as educational material exposes patients to potential financial exploitation [51]. The "attention economy" of social media is thus not a neutral marketplace of ideas; for IVF patients, it is a high-risk environment where the most visible information is often the least reliable.

Clinical Implications and Future Directions

Therefore, a paradigm shift is imperative for medical professionals and health care organizations. A passive approach of simply producing high-quality content and expecting it to be discovered is destined to fail. A proactive, 2-pronged strategy is required. First, proactive content creation demands that clinicians become platform-native communicators [11,52,53]. This means moving beyond static informational videos and embracing storytelling, patient-centered narratives, and visually compelling formats developed in collaboration with communication experts, without compromising scientific integrity [54]. Second, reactive engagement is equally crucial. Medical professionals and institutions should consider themselves "digital first responders," actively identifying and correcting high-reach misinformation through comments, response videos, or collaborations with platforms—a practice shown to be effective in other health contexts [43].

Looking forward, a clear agenda for future research emerges from this work. While our quantitative model identified what drives dissemination, qualitative studies are now needed to understand why. In-depth interviews with IVF patients could illuminate the specific motivations and cognitive processes behind their engagement with different types of content. Furthermore, interventional research is urgently needed to design and test the efficacy of novel communication strategies. Randomized controlled trials could compare the reach and impact of standard informational videos against narrative-based, emotionally resonant, yet scientifically accurate content. Finally, longitudinal studies are required to track the real-world impact of social media exposure on patient decision-making, treatment adherence, and clinical outcomes over time.

This study has several notable strengths. It is the first to systematically analyze IVF content across China's 3 dominant short-video platforms. By using a robust content analysis methodology underpinned by validated instruments—GQS and the mDISCERN tool with high interrater reliability (κ>0.80)—we provided a rigorous assessment of content quality. Methodologically, our use of a dual XGBoost modeling strategy provides a particularly robust analysis. We first demonstrated the model's inability to predict absolute "likes" (R^2=−7.5), empirically confirming the highly stochastic nature of social media virality [55]. We then solidified our conclusions by using a second, logarithmically transformed model, which, despite a better predictive fit (R^2=0.294), reproduced the exact same feature importance hierarchy. This confirmatory step ensures that our central finding is robust and not an artifact of data skewness. Third, while our optimized XGBoost model achieved a robust R^2 of 0.294, approximately 70% of the variance in dissemination remains unexplained. This suggests that engagement is influenced by unmeasured "soft" factors extrinsic to medical quality or uploader status, such as thumbnail aesthetics, opening "hooks" (the first 3 s of video), emotional delivery, and platform-specific trending audio. Future studies should use computer vision and sentiment analysis to quantify these variables.

Limitations

However, the study's limitations must be acknowledged. First, our sampling strategy restricted analysis to the top-ranked videos. While this design validly represents the "information diet" of a typical user who rarely scrolls beyond the first page, it introduces selection bias. Our findings characterize the most visible content ecosystem rather than the entire universe of IVF-related videos. Additionally, its cross-sectional design precludes any causal inference. The findings are specific to the Chinese social media context and may not be generalizable. Finally, our analysis used "likes" as the primary proxy for dissemination. Future research could conduct a more granular analysis exploring the distinct drivers of other interaction types, such as comments or shares, which may reflect different dimensions of user engagement [56].

Conclusions

In conclusion, our research paints a stark picture of the IVF information landscape on Chinese social media, where the mechanisms of dissemination are dangerously decoupled from the principles of evidence-based medicine. Our dual-model analysis robustly demonstrates that the digital health ecosystem does not inherently reward quality; it rewards influence. To bridge the gap between what is popular and what is reliable, the medical community must not only produce trustworthy

XSL•FO
RenderX

information but also master the art and science of platform-native communication to ensure that their expertise can successfully navigate the algorithmic currents and reach the patients who need it most.

Funding

This work was supported by the National Natural Science Foundation of China (grant 82301803). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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. Wyns C, Bergh C, Calhaz-Jorge C, et al. Assisted reproductive technology in Europe, 2016: results generated from European registries by ESHRE. Hum Reprod Open 2020(3):hoaa032. [doi: 10.1093/hropen/hoaa032]
2. Leone D, Menichetti J, Barusi L, et al. Breaking bad news in assisted reproductive technology: a proposal for guidelines. Reprod Health 2017 Jul 20;14(1):87. [doi: 10.1186/s12978-017-0350-1] [Medline: 28728610]
3. Mayette E, Scalise A, Li A, McGeorge N, James K, Mahalingaiah S. Assisted reproductive technology (ART) patient information-seeking behavior: a qualitative study. BMC Womens Health 2024 Jun 15;24(1):346. [doi: 10.1186/s12905-024-03183-z] [Medline: 38877503]
4. Lemoine ME, O'Connell SBL, Grunberg PH, Gagné K, Ells C, Zelkowitz P. Information needs of people seeking fertility services in Canada: a mixed methods analysis. Health Psychol Behav Med 2021 Feb 11;9(1):104-127. [doi: 10.1080/21642850.2021.1879650] [Medline: 34104552]
5. Karami NA, Latifi M, Berahmand N, Eini F, Al-Suqri MN. The impact of individual factors on health information-seeking behavior of infertile couples undergoing assisted reproductive technologies: Longo model. Adv Biomed Res 2023;12(1):68. [doi: 10.4103/abr.abr_181_22] [Medline: 37200740]
6. Slauson-Blevins KS, McQuillan J, Greil AL. Online and in-person health-seeking for infertility. Soc Sci Med 2013 Dec;99:110-115. [doi: 10.1016/j.socscimed.2013.10.019] [Medline: 24355477]
7. Jones CA, Mehta C, Zwingerman R, Liu KE. Fertility patients' use and perceptions of online fertility educational material. Fertil Res Pract 2020;6:11. [doi: 10.1186/s40738-020-00083-2] [Medline: 32695432]
8. Moran G, Muzellec L, Johnson D. Message content features and social media engagement: evidence from the media industry. J Prod Brand Manag 2019 Oct 16;29(5):533-545. [doi: 10.1108/JPBM-09-2018-2014]
9. Zeng F, Zhang W, Wang M, Zhang H, Zhu X, Hu H. Douyin and Bilibili as sources of information on lung cancer in China through assessment and analysis of the content and quality. Sci Rep 2024 Sep 4;14(1):20604. [doi: 10.1038/s41598-024-70640-y] [Medline: 39232044]
10. Gao H, Yin H, Peng L, Wang H. Effectiveness of social video platforms in promoting COVID-19 vaccination among youth: a content-specific analysis of COVID-19 vaccination topic videos on Bilibili. Risk Manag Healthc Policy 2022;15:1621-1639. [doi: 10.2147/RMHP.S374420] [Medline: 36071816]
11. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520. [Medline: 25083128]
12. Cain J. Social media in health care: the case for organizational policy and employee education. Am J Health Syst Pharm 2011 Jun 1;68(11):1036-1040. [doi: 10.2146/ajhp100589] [Medline: 21593233]
13. Chretien KC, Kind T. Social media and clinical care: ethical, professional, and social implications. Circulation 2013 Apr 2;127(13):1413-1421. [doi: 10.1161/CIRCULATIONAHA.112.128017] [Medline: 23547180]
14. Fernández-Luque L, Bau T. Health and social media: perfect storm of information. Healthc Inform Res 2015 Apr;21(2):67-73. [doi: 10.4258/hir.2015.21.2.67] [Medline: 25995958]
15. Laranjo L, Arguel A, Neves AL, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
16. Sommariva S, Vamos C, Mantzarlis A, Đào LUL, Martinez Tyson D. Spreading the (fake) news: exploring health messages on social media and the implications for health professionals using a case study. Am J Health Educ 2018 Jul 4;49(4):246-255. [doi: 10.1080/19325037.2018.1473178]
17. Tasnim S, Hossain MM, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 2020 May;53(3):171-174. [doi: 10.3961/jpmph.20.094] [Medline: 32498140]
18. Abbasi J. Widespread misinformation about infertility continues to create COVID-19 vaccine hesitancy. JAMA 2022 Mar 15;327(11):1013-1015. [doi: 10.1001/jama.2022.2404] [Medline: 35191947]

XSL•FO
RenderX

19. John JN, Gorman S, Scales D, Gorman J. Online misleading information about women's reproductive health: a narrative review. J Gen Intern Med 2025 Apr;40(5):1123-1131. [doi: 10.1007/s11606-024-09118-6] [Medline: 39511120]
20. Zaila KE, Osadchiy V, Shahinyan RH, Mills JN, Eleswarapu SV. Social media sensationalism in the male infertility space: a mixed methodology analysis. World J Mens Health 2020 Oct;38(4):591-598. [doi: 10.5534/wjmh.200009] [Medline: 32378368]
21. Olsen C, George DMM. Cross-sectional study design and data analysis. : College Entrance Examination Board; 2004 URL: http://www.yes-competition.org/media.collegeboard.com/digitalServices/pdf/yes/4297_MODULE_05.pdf [accessed 2025-12-24]
22. TikTok revenue and usage statistics (2025). Business of Apps. 2024. URL: https://www.businessofapps.com/data/tik-tok-statistics/ [accessed 2025-12-24]
23. Lu X, Lu Z. Fifteen seconds of fame: a qualitative study of Douyin, a short video sharing mobile application in China. Presented at: Social Computing and Social Media Design, Human Behavior and Analytics: 11th International Conference, SCSM 2019, Held as Part of the 21st HCI International Conference, HCII 2019; Jul 26-31, 2019. [doi: 10.1007/978-3-030-21902-4_17]
24. Annual report. : Bilibili Inc; 2024 URL: https://ir.bilibili.com/media/jesfytmo/bilibili-inc-2024-annual-report.pdf [accessed 2025-12-24]
25. Bilibili to change video tracking to viewing duration from number of hits to avoid short clip bias. Yicai Global. 2023 Jun 27. URL: https://www.yicaiglobal.com/news/20230627-bilibili-to-change-video-tracking-to-viewing-duration-from-number-of-hits-to-avoid-short-clip-bias [accessed 2025-12-24]
26. Xiaohongshu (Rednote) after the TikTok refugees' dust has settled. TechBuzz China. 2025. URL: https://techbuzzchina.substack.com/p/xiaohongshu-rednote-after-the-tiktok [accessed 2025-12-24]
27. Hall C, Yu S. China's instagram-like Xiaohongshu making inroads with e-commerce sales. Reuters. 2024. URL: https://www.reuters.com/technology/chinas-instagram-like-xiaohongshu-making-inroads-with-e-commerce-sales-2024-12-16/ [accessed 2025-12-24]
28. Tian Y, Robinson JD. Content analysis of health communication. In: Research Methods in Health Communication: Principles and Application: Routledge; 2014:196-212. [doi: 10.4324/9780203115299-13]
29. Fazeli S, Sabetti J, Ferrari M. Performing qualitative content analysis of video data in social sciences and medicine: the visual-verbal video analysis method. Int J Qual Methods 2023 Oct;22:1-18. [doi: 10.1177/16094069231185452]
30. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007 Sep;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
31. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. Asso for Info Science & Tech 2015 Oct;66(10):2071-2084. [doi: 10.1002/asi.23311]
32. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111. [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
33. Kaicker J, Borg Debono V, Dang W, Buckley N, Thabane L. Assessment of the quality and variability of health information on chronic pain websites using the DISCERN instrument. BMC Med 2010 Oct 12;8:59. [doi: 10.1186/1741-7015-8-59] [Medline: 20939875]
34. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [Medline: 23092060]
35. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013 Apr 29;13:61. [doi: 10.1186/1471-2288-13-61] [Medline: 23627889]
36. Ghosal S, Jain A. Depression and suicide risk detection on social media using fastText embedding and XGBoost classifier. Procedia Comput Sci 2023;218:1631-1639. [doi: 10.1016/j.procs.2023.01.141]
37. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. IEEE Access 2017;5:8869-8879. [doi: 10.1109/ACCESS.2017.2694446]
38. Belete DM, Huchaiah MD. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. International Journal of Computers and Applications 2022 Sep 2;44(9):875-886. [doi: 10.1080/1206212X.2021.1974663]
39. Adnan M, Alarood AAS, Uddin MI, Ur Rehman I. Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. PeerJ Comput Sci 2022;8:e803. [doi: 10.7717/peerj-cs.803] [Medline: 35494796]
40. Prendin F, Pavan J, Cappon G, Del Favero S, Sparacino G, Facchinetti A. The importance of interpreting machine learning models for blood glucose prediction in diabetes: an analysis using SHAP. Sci Rep 2023 Oct 6;13(1):16865. [doi: 10.1038/s41598-023-44155-x] [Medline: 37803177]
41. Wang H, Liang Q, Hancock JT, Khoshgoftaar TM. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. J Big Data 2024;11(1):45. [doi: 10.1186/s40537-024-00905-w]

XSL•FO
RenderX

42. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
43. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
44. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. JAMA 2007 Dec 5;298(21):2482-2484. [doi: 10.1001/jama.298.21.2482] [Medline: 18056901]
45. Azak M, Şahin K, Korkmaz N, Yıldız S. YouTube as a source of information about COVID-19 for children: Content quality, reliability, and audience participation analysis. J Pediatr Nurs 2022;62:e32-e38. [doi: 10.1016/j.pedn.2021.06.024] [Medline: 34247879]
46. Wang M, Yao N, Wang J, Chen W, Ouyang Y, Xie C. Bilibili, TikTok, and YouTube as sources of information on gastric cancer: assessment and analysis of the content and quality. BMC Public Health 2024 Jan 2;24(1):57. [doi: 10.1186/s12889-023-17323-x] [Medline: 38166928]
47. Liu H, Peng J, Li L, et al. Assessment of the reliability and quality of breast cancer related videos on TikTok and Bilibili: cross-sectional study in China. Front Public Health 2024;11:1296386. [doi: 10.3389/fpubh.2023.1296386] [Medline: 38317686]
48. Aldous KK, An J, Jansen BJ. View, like, comment, post: analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. Presented at: 13th International Conference on Web and Social Media, ICWSM 2019; Jun 11-14, 2019. [doi: 10.1609/icwsm.v13i01.3208]
49. Lim MSC, Wright CJC, Carrotte ER, Pedrana AE. Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites. Health Promot J Austr 2016 Feb;27(3):187-197. [doi: 10.1071/HE16057] [Medline: 27719734]
50. Broughton DE, Schelble A, Cipolla K, Cho M, Franasiak J, Omurtag KR. Social media in the REI clinic: what do patients want? J Assist Reprod Genet 2018 Jul;35(7):1259-1263. [doi: 10.1007/s10815-018-1189-2] [Medline: 29766400]
51. Carneiro MM, Koga CN, Mussi MCL, Fradico PF, Ferreira MCF. Quality of information provided by Brazilian Fertility Clinic websites: compliance with Brazilian Medical Council (CFM) and American Society for Reproductive Medicine (ASRM) Guidelines. JBRA Assist Reprod 2023 Jun 22;27(2):169-173. [doi: 10.5935/1518-0557.20220026] [Medline: 35916465]
52. Richardson MA, Park W, Bernstein DN, Mesfin A. Analysis of the quality, reliability, and educational content of youtube videos concerning spine tumors. Int J Spine Surg 2022 Apr;16(2):278-282. [doi: 10.14444/8215] [Medline: 35444036]
53. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188. [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
54. Murphy D, Balka E, Poureslami I, Leung DE, Nicol AM, Cruz T. Communicating health information: the community engagement model for video production. Can J Commun 2007 Nov 12;32(3-4):383-400. [doi: 10.22230/cjc.2007v32n3a1966]
55. Poecze F, Ebster C, Strauss C. Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. Procedia Comput Sci 2018;130:660-666. [doi: 10.1016/j.procs.2018.04.117]
56. Tenenboim O. Comments, shares, or likes: what makes news posts engaging in different ways. Soc Media Soc 2022 Oct;8(4):1-15. [doi: 10.1177/20563051221130282]

Abbreviations

DAU: daily active user
GQS: Global Quality Score
IVF: in vitro fertilization
mDISCERN: modified DISCERN
SHAP: Shapley Additive Explanations
XGBoost: Extreme Gradient Boosting

Edited by T Mackey; submitted 10.Sep.2025; peer-reviewed by A Famotire, Y Jiang; revised version received 09.Dec.2025; accepted 09.Dec.2025; published 28.Jan.2026.

Please cite as:
Bai X, Guo F, Chu D
Quality, Reliability, and Dissemination of In Vitro Fertilization–Related Videos on Chinese Social Media: Cross-Sectional Analysis of 300 Short Videos
JMIR Infodemiology 2026;6:e83900
URL: https://infodemiology.jmir.org/2026/1/e83900
doi:10.2196/83900

XSL•FO
RenderX

© Xueyan Bai, Feng Guo, Dapeng Chu. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 28.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Japanese-Language AI Agent System for Human Papillomavirus Vaccine Infoveillance and Public Communication: Development and Feasibility Evaluation

Junyu Liu[1], ME; Siwen Yang[2], BEng; Dexiu Ma[3], MS; Qian Niu[4], PhD; Zequn Zhang[5], MS; Momoko Nagai-Tanima[1], PhD; Tomoki Aoyama[1], MD, PhD

[1]Graduate School of Medicine, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, Japan

[2]David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada

[3]Department of Computer Science, Whitacre College of Engineering, Texas Tech University, Lubbock, TX, United States

[4]Graduate School of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

[5]Department of EEIS, University of Science and Technology of China, Hefei, Anhui, China

Corresponding Author:
Tomoki Aoyama, MD, PhD
Graduate School of Medicine, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, Japan

Related Article:

This is a corrected version. See correction statement: https://infodemiology.jmir.org/2026/1/e109982

Abstract

Background: Human papillomavirus (HPV) vaccine hesitancy remains a significant public health challenge in Japan, where proactive vaccination recommendations were suspended between 2013 and 2021. The resulting information gap between medical institutions and vaccine-hesitant populations is exacerbated by misinformation on social media platforms. Traditional public health communication strategies cannot address individual queries while simultaneously monitoring population-level discourse.

Objective: This study aimed to develop and conduct a feasibility evaluation of a dual-purpose artificial intelligence agent system that delivers verified HPV vaccine information to the public through a conversational interface while generating infoveillance reports for medical institutions based on user interactions and social media discourse.

Methods: We implemented a system with 3 components: a vector database integrating 139,803 documents, including academic papers, Japanese government sources, news media, and social media posts; a retrieval-augmented generation chatbot using a ReAct agent architecture with iterative multitool orchestration across 5 specialized knowledge sources; and an automated report generation system with modules for news analysis, research synthesis, social media sentiment analysis, including stance classification and topic modeling, and user interaction pattern identification. System performance was assessed using both automated and manual evaluation protocols on a scale from 0 to 5.

Results: The entire system functioned as expected. For single-turn evaluation, the chatbot achieved mean scores of 4.83 (SD 0.67; 95% CI 4.71-4.93) for relevance, 4.89 (SD 0.53; 95% CI 4.79-4.97) for routing, 4.50 (SD 1.29; 95% CI 4.27-4.70) for reference quality, 4.90 (SD 0.62; 95% CI 4.78-4.99) for correctness, and 4.88 (SD 0.54; 95% CI 4.78-4.96) for professional identity, with an overall mean of 4.80. Multiturn evaluation yielded higher mean scores: 4.94 for context memory (SD 0.25; 95% CI 4.84-5.00) and an overall mean of 4.98, with topic centering and identity achieving 5.00. The report generation system achieved high scores across all sections: 4.83 for completeness (SD 0.37; 95% CI 4.73-4.94), 4.88 for correctness (SD 0.33; 95% CI 4.77-4.96), and 4.12 for helpfulness (SD 0.48; 95% CI 3.98-4.27). Reference validity achieved perfect scores (5.00) across all periods, with citation correctness averaging 4.21 (SD 0.58; 95% CI 3.96-4.46).

Conclusions: This feasibility study demonstrated that an integrated artificial intelligence agent system can support both public HPV vaccine communication and social media infoveillance in a Japanese-language context. Prospective deployment with real users is needed to assess actual public health impact.

(JMIR Infodemiology 2026;6:e90295) doi:10.2196/90295

KEYWORDS

human papillomavirus; HPV; artificial intelligence agent; AI agent; large language model; stance analysis; topic modeling; artificial intelligence; AI

XSL•FO
RenderX

Introduction

Human papillomavirus (HPV) is a significant public health concern that caused 662,044 new cervical cancer cases and 348,709 deaths worldwide in 2022 [1]. HPV vaccines have demonstrated high efficacy in preventing HPV-related diseases [2], and numerous countries have implemented national vaccination programs since their introduction in 2006. However, vaccine hesitancy remains a persistent challenge [3], particularly in countries such as Japan, where HPV vaccination rates have dropped dramatically owing to safety concerns and media coverage of alleged adverse events [4].

The spread of misinformation regarding HPV vaccines through social media platforms has exacerbated public concern [5], creating a complex information landscape in which accurate medical information competes with anecdotal reports and unverified claims. Traditional public health communication strategies face significant challenges in addressing health misinformation at scale because responding effectively requires simultaneously countering individual-level psychological barriers and monitoring population-level misinformation dynamics across diverse platforms [6]. Medical institutions require timely insights into public discourse to develop effective communication strategies; however, manual analysis of vast amounts of social media data and public inquiries is resource-intensive and time-consuming.

Recent advances in large language model (LLM) and retrieval-augmented generation (RAG) systems offer promising solutions to bridge this information gap [7]. LLMs demonstrate remarkable capabilities for natural language understanding and generation across multiple languages [8], including Japanese, which presents unique challenges owing to its complex writing system and grammatical structure [9]. RAG systems combine the generative capabilities of LLMs with retrieval from curated knowledge bases, enabling responses grounded in verified information sources while maintaining conversational fluency.

Previous studies have applied natural language processing to HPV-related social media analysis, primarily focusing on sentiment analysis and topic modeling [10]. However, these approaches typically operate as passive analytical tools instead of as active information dissemination systems. Chatbot systems for health information have been developed for various domains [11]; however, few integrate multisource retrieval from academic literature, official guidelines, news media, and social media discourse while simultaneously providing bidirectional communication between the public and health institutions.

The Japanese context poses unique challenges to and opportunities for such systems. Japan experienced a dramatic suspension of proactive HPV vaccination recommendations from 2013 to 2021 because of safety concerns, resulting in vaccination rates falling below 1% and creating a substantial gap in population immunity [4]. The government's 2022 resumption of vaccination recommendations necessitates renewed public education efforts [12]. Furthermore, Japanese-language health information systems face technical challenges, including multiscript processing (hiragana, katakana, and kanji), medical terminology localization, and culturally appropriate communication styles.

In this study, we developed and implemented a comprehensive artificial intelligence (AI) agent system designed to address both public information needs and institutional monitoring requirements for the HPV vaccine discourse in Japan. Our system has two main features: (1) a RAG-based chatbot that answers public queries by retrieving and synthesizing information from academic papers, official documents, news articles, and social media posts; and (2) an analytics dashboard that generates reports for medical institutions based on aggregated chat histories and social media data. The system uses multisource data collection, semantic search with vector embeddings, intelligent query routing, and automated evaluation frameworks.

Methods

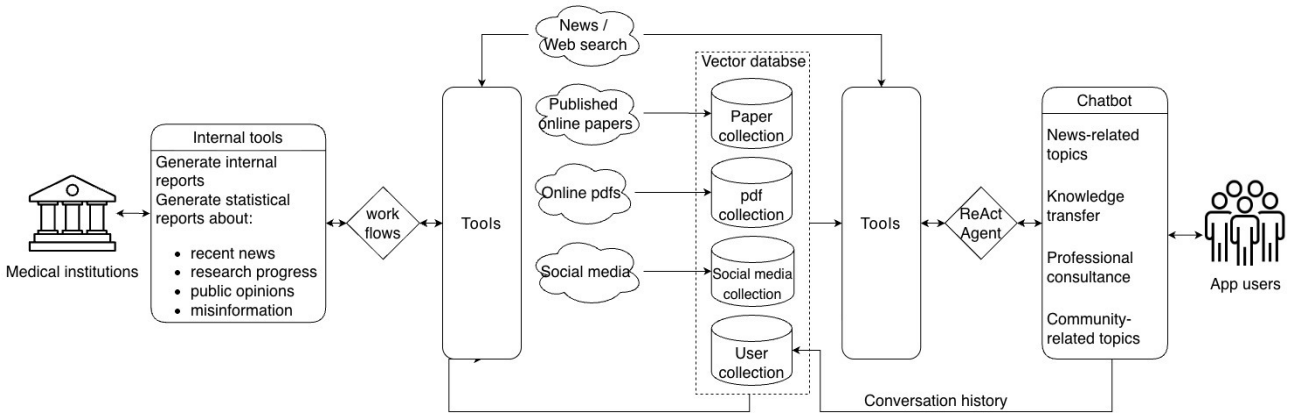
System Architecture

We developed a multicomponent AI agent system for HPV vaccine information dissemination and public opinion analysis. The system comprises 3 main modules: a multisource data collection and storage system, a ReAct agent–based chatbot [13] for public information queries, and a report generation system for medical institutions.

The overall architecture follows a distributed design pattern with a centralized vector database (Qdrant) [14] serving as a knowledge repository (Figure 1). Data flow from multiple external sources through specialized collectors in the database, where they are indexed using semantic embeddings. The chatbot and report generation modules both query this database but serve different end users with distinct interfaces and functionalities. The system implements a bidirectional information flow. The chatbot provides HPV vaccine information to the public while simultaneously collecting user inquiries with consent, and the report generator aggregates these interactions with social media data to produce actionable insights for medical institutions.

XSL•FO
RenderX

Figure 1. Overall system architecture showing the integration of the data collection, vector database, chatbot interface, and report generation components. HPV: human papillomavirus.



Data Collection and Database

We implemented a vector database infrastructure as the central knowledge repository managing 4 distinct collections: academic papers, official documents, social media posts, and chat histories. Each document was represented as a 2048D vector using embedding models optimized for Japanese-language processing (PLaMo-Embedding-1B; Preferred Networks, Inc) [15]. This database uses cosine similarity metrics for semantic search operations [16,17], supporting efficient retrieval with customizable parameters and metadata preservation.

Data were collected from 4 heterogeneous sources to construct a comprehensive knowledge base spanning scientific evidence, official guidance, media coverage, and public discourse. Academic papers were retrieved from PubMed [18] through keyword-based searches with temporal filtering, capturing abstracts, MeSH (Medical Subject Headings) terms, journal information, and DOIs. Official documents and web content were collected from authoritative sources, including the World Health Organization [19] and the Japanese Ministry of Health, Labor, and Welfare (MHLW) [20] through multiple complementary methods: intelligent query analysis for information synthesis, filtered web searches targeting official sources, online PDF document discovery and extraction, and specialized scraping of government meeting records and reference materials. News articles were aggregated from multiple news sources using keyword-based searches in Japanese and English, and deduplication was used to ensure unique coverage. Social media data from X (formerly known as Twitter) were collected through daily automated harvesting using Tweepy [21] with temporal specifications. Rate limit handling was implemented to ensure comprehensive data capture across extended periods.

Chatbot Implementation

We implemented a ReAct agent–based chatbot using LlamaIndex [22] through an iterative multitool orchestration architecture in which a single intelligent controller dynamically selects and combines information from multiple specialized data sources across sequential decision-making iterations. The system addresses the challenge of answering diverse user queries by enabling flexible, multisource information gathering while maintaining conversational coherence and citation quality assurance.

Architecture

The chatbot uses a single controller agent with 5 specialized tools: papers (academic literature), the web (official documents and guidelines), social media (public discourse), news (media coverage), and chitchat (casual conversations). A citation validation tool ensures the response quality. Unlike conventional routing architectures, this design enables the controller to select and combine multiple tools iteratively for a single query, synthesizing information across heterogeneous sources.

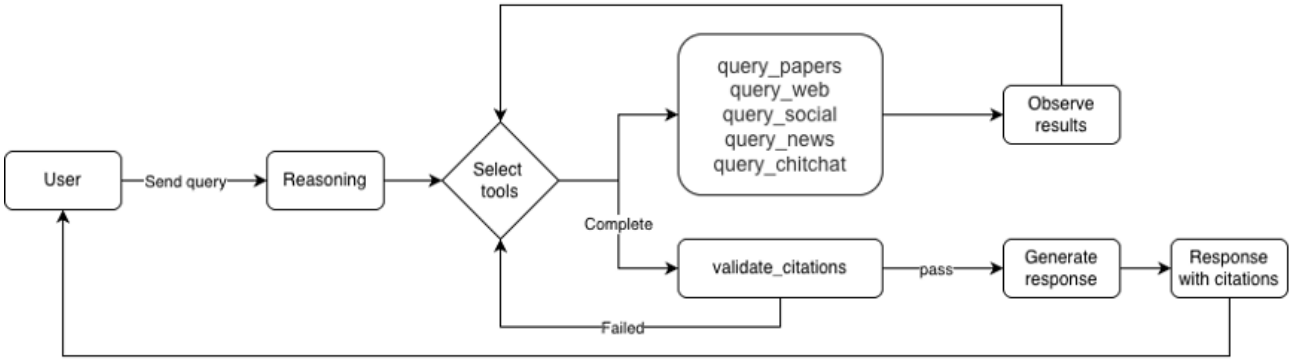
Each tool performs a semantic similarity search against its respective vector database collection, retrieving the relevant documents that the controller assembles into responses using the proper source attribution. The controller analyzes queries in the conversation context, determines appropriate information sources, and iteratively gathers evidence until it is sufficient for comprehensive response synthesis. A web-based Streamlit interface (Snowflake Inc) [23] presents conversations with integrated citations, whereas tool use metadata are stored with user consent to inform institutional reporting.

Operational Workflow

The query processing follows an iterative orchestration loop (Figure 2). Upon receiving a user message (query), the controller examines the question along with the recent conversation history to assess the information requirements. The controller then enters a decision cycle: (1) analyze information gaps, (2) select the most appropriate tool, (3) retrieve results via a semantic similarity search, (4) review relevance, and (5) determine whether sufficient evidence exists or additional retrieval is required. This process continues until comprehensive information is gathered for response generation.

XSL•FO
RenderX

Figure 2. Chatbot operational workflow showing the iterative ReAct agent architecture. The user query flows through reasoning and tool selection, with the controller dynamically selecting from 5 specialized tools (papers, the web, social media, news, and chitchat). Results are observed and validated through a citation validation mechanism before generating the final response with proper source attribution.



The system generates responses using inline citation markers corresponding to the retrieved documents, enabling users to trace claims to their original sources. A 2-level citation validation mechanism ensures quality: individual tools validate their own citations, and a dedicated validation tool examines the entire response for citation completeness before delivery.

Privacy protection is implemented for social media queries, synthesizing themes and sentiment patterns without attributing statements to individual users. Stateful conversation management maintains dialogue context through a windowed history approach, enabling interpretation of follow-up questions with implicit references (eg, “What about side effects?” following a vaccine efficacy discussion) while maintaining topical continuity.

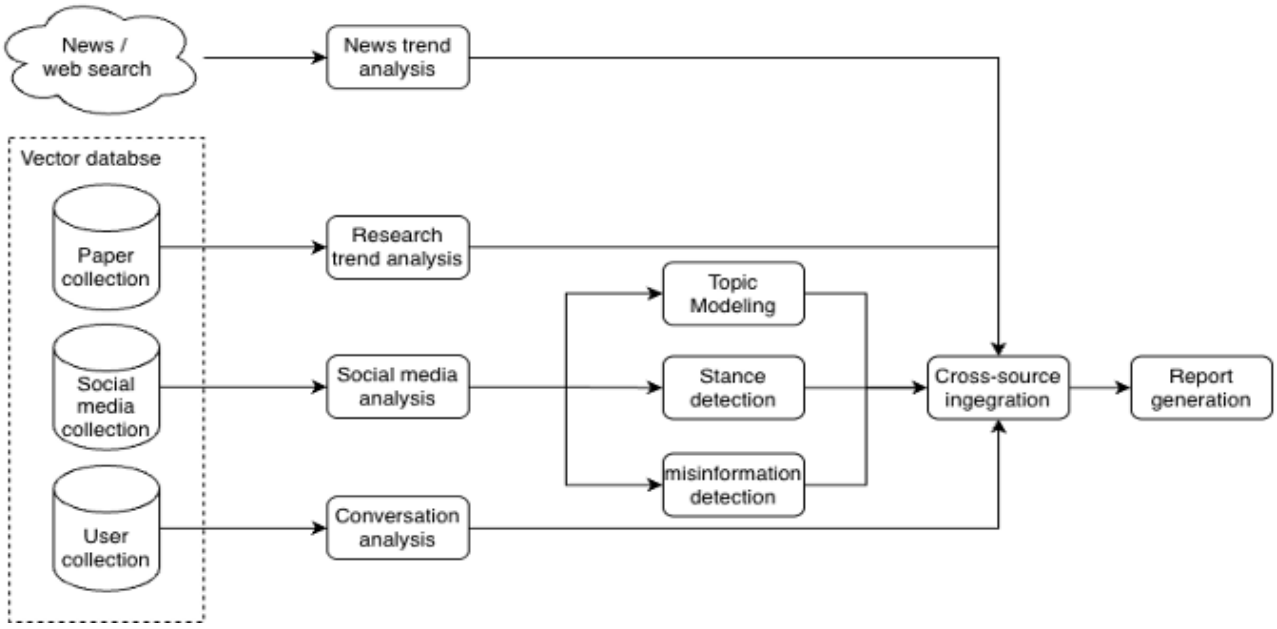
Report Generation System

We developed an automated report generation system that synthesizes data from multiple sources to produce comprehensive PDF reports for medical institutions and policymakers. This system uses LLMs for intelligent analysis and generates professional documents with academic-style citations, visualizations, and actionable insights.

System Architecture

The report generation system (Figure 3) comprises four specialized analysis modules coordinated by a central orchestrator: (1) news analyzer for recent news, (2) paper analyzer for recent academic research, (3) social media analyzer for public sentiment analysis, and (4) chat analyzer for user interaction pattern identification. Each module queries the vector database for documents within a configurable time window, performs domain-specific analysis using an LLM-based inference, and generates a structured output with properly formatted citations. The orchestrator coordinates module execution, manages data flow between components, aggregates results, and assembles the final PDF document with bilingual support (Japanese and English).

Figure 3. Report generation system architecture. Data flow from external news sources and vector database collections (papers, social media, and user conversations) through specialized analysis modules. The social media analyzer performs topic modeling, stance detection, and misinformation detection. All analysis results are integrated through cross-source aggregation before final report generation.



XSL•FO
RenderX

Social Media Analysis

Social media platforms have emerged as critical channels for public health discourse, serving as real-time indicators of population-level attitudes toward vaccination [24]. For medical institutions and policymakers, a systematic analysis of social media content provides valuable insights into public concerns, emerging misinformation narratives, and temporal shifts in vaccine sentiment [25].

The social media analyzer performs a multidimensional public opinion assessment through 4 complementary analytical processes. Stance classification categorizes each post as supportive, opposed, or neutral regarding HPV vaccination using batch LLM inference with a temporal context, thereby aggregating daily counts to track sentiment evolution. Topic modeling uses a hybrid approach that combines statistical latent Dirichlet allocation [26] with LLM-based semantic interpretation in which Japanese text undergoes morphological analysis [27], term frequency–inverse document frequency weighting [28], and model training via Gensim (RARE Technologies Ltd) [29] to extract interpretable topic labels from keyword distributions. Misinformation detection uses LLM-based analysis to identify posts containing claims that contradict the established scientific consensus, recording for each flagged post a short statement of the claim for institutional awareness. Visualization generation produces temporal trend graphics and thematic distribution charts that are embedded directly in the reports to enhance interpretability for nontechnical stakeholders.

PDF Report Assembly

The report generator produces professional bilingual documents (in Japanese and English) structured into five main sections: (1) news trends, presenting influential media coverage with relevance assessments; (2) research progress, synthesizing recent academic literature; (3) social media analysis, containing sentiment trends, topic distributions, and visualizations; (4) chat analysis, identifying user information needs and knowledge gaps; and (5) overall summary, providing an executive synthesis across all data sources.

Each section includes inline citations with source-appropriate formatting, which enables medical institutions to verify information and assess evidence quality independently. The orchestrator synthesizes the findings from all the analysis modules into an executive summary that provides a comprehensive overview of the reporting period. This multisource integration approach captures diverse perspectives, enabling stakeholders to develop informed HPV vaccination communication strategies and policy interventions.

Evaluation Framework

We developed a multifaceted evaluation framework comprising complementary assessment protocols for chatbot performance and automated report generation. The framework uses an LLM-based evaluation for scalable assessment and human expert validation for quality assurance.

Chatbot Evaluation Methodology

The chatbot evaluation framework assesses system performance through 2 complementary protocols: single-turn evaluation for individual question-answer exchanges and multiturn evaluation for complete conversation quality. Both protocols use LLM-based judges who receive conversation context, tool use information, and scoring rubrics, generating scores on a scale from 0 to 5 (details are provided in Multimedia Appendices 1 and 2) for each dimension, along with written evaluation notes.

To collect test data, 3 volunteers simulated diverse user personas (Multimedia Appendix 3) to create realistic conversations. They posed questions across different personas and topics, and multiturn conversations with the production chatbot system were conducted and stored for subsequent evaluations.

Using the test data, a single-turn evaluation assesses 5 dimensions: relevance, measuring whether the response addressed the question; routing, evaluating the appropriate tool selection for the query type; reference, assessing citation validity and proper source attribution; correctness, verifying factual accuracy against established guidelines; and identity, examining professional medical communication tone. Multiturn evaluation extends these 5 dimensions through 2 additional metrics for conversational coherence: context memory, which assesses the appropriate use of information from previous turns, and topic centering, which evaluates natural conversation flow with logical transitions between related topics.

We randomly selected 20 question-answer pairs for manual scoring by the 3 domain experts to validate the reliability of the automated evaluation. The correlation between the expert and LLM-generated scores was analyzed to assess whether the automated metrics accurately reflected human judgment.

Report Generation Evaluation Methodology

Report quality assessment uses 2 complementary evaluation protocols: main text evaluation for content quality and reference evaluation for citation validity. Both protocols were applied by 3 domain experts using the standardized scoring rubrics provided in Multimedia Appendices 4 and 5, assigning scores on a scale from 0 to 5 for each dimension. Each expert scored every report section independently. The main text evaluation assesses 3 dimensions: completeness, measuring structural integrity and whether sections contain well-developed content; correctness, evaluating factual accuracy and proper interpretation of source materials; and helpfulness, examining practical utility and actionable insights for institutional stakeholders. Reference evaluation validates citation quality across 2 dimensions: reference validity, measuring the proportion of cited sources that are accessible and exist in the underlying database, and citation correctness, assessing whether citations properly support the claims made in the report text.

Temporal analysis generates reports for multiple periods to assess system robustness across varying conditions. This approach evaluates both the system consistency and the ability to capture temporal variations in public discourse. Three volunteers read and scored each report independently.

Ethical Considerations

The knowledge base was constructed exclusively from publicly available data: PubMed abstracts, World Health Organization and Japanese MHLW documents, news articles, and public X

XSL•FO
RenderX

posts. Social media data were analyzed only in aggregate; no individual users were identified or quoted.

During chatbot evaluation, 3 volunteer members of the research team generated simulated conversations with informed consent. No personally identifiable information or personal health data were collected. Conversation data and metadata (time stamps and tool use logs) were stored in a secured database accessible only to the research team.

This study was determined to not require institutional review board review in accordance with the Ethical Guidelines for Medical and Biological Research Involving Human Subjects (Ministry of Education, Culture, Sports, Science, and Technology; MHLW; and Ministry of Economy, Trade, and Industry, 2021) [30] because (1) all external data were publicly available and analyzed without individual identification; (2) evaluation participants were consenting research team members, not patients or vulnerable populations; and (3) no clinical interventions or health outcomes were measured.

Results

Database

This database integrates 139,803 documents across 4 specialized collections (Table 1). This document volume and distribution enables a realistic simulation of operational use cases.

Table . Database collection statistics.

Collection	Documents (n=139,803), n (%)	Unique source	Primary language
Papers	51 (0.04)	37 journals	English (100%)
Official documents	6739 (4.82)	MHLW[a]	Japanese (100%)
Social media	132,874 (95.04)	X (formerly known as Twitter)	Japanese (100%)
Chat conversations[b]	139 (0.10)	37 sessions	English (98.6%); Japanese (1.4%)

[a]MHLW: Ministry of Health, Labor, and Welfare.

[b]The chat row describes the conversation set used for the chatbot evaluation: 139 turns, each stored as one document, from 37 conversations. Language proportions were computed over the 139 user messages in this set. Primary language for the Papers, Official documents and Social media rows was assigned on the basis of the source of each collection and was not derived from automated language detection.

Chatbot

On the basis of the large-scale database, we applied simulation of operational use cases.

Example Response

Textbox 1 presents a representative question-answer exchange that demonstrates the response characteristics of the chatbot.

Textbox 1. Example chatbot response demonstrating professional medical tone, structured information delivery, evidence-based recommendations with citations, and personalized guidance.

User query

"Are there any specific HPV-related health risks or symptoms that become more common or concerning for people over 65?"

Chatbot response

"Thank you for asking such an important question about HPV-related health risks for people over 65. As a healthcare provider, I want to address your concerns with both the latest medical evidence and an understanding of the unique considerations for older adults.

Key Health Risks and Symptoms After 65

While HPV infections can occur at any age, there are some important patterns to

understand for people over 65. Research shows that while new HPV infections become less common with age, when cervical abnormalities do occur in older adults, they tend to be more significant [1].

...

References:

[1] Prevalence of cervical cytology abnormalities and its correlation with human papillomavirus infection. Medicinski glasnik (2025). DOI: 10.17392/1968-22-02"

Evaluation Results

The chatbot system was evaluated using simulated conversations generated by 3 volunteers. Conversations spanned diverse user personas of varying age groups, occupations, and levels of concern about HPV vaccination. The questions covered multiple information domains, including vaccine safety, efficacy, eligibility criteria, and procedural guidelines. Table 2 summarizes the single-turn evaluation results (Multimedia Appendix 6) for all 5 assessment dimensions. Average scores ranged from 4.50 to 4.90 on the scale from 0 to 5, with correctness (4.90, 95% CI 4.78-4.99) and routing (4.89, 95% CI 4.79-4.97) achieving the highest scores. Across all dimensions, 90% (125/139) to 99% (137/139) of responses received scores of 4 or higher.

XSL•FO
RenderX

Table . Single-turn evaluation results (n=139 question-answer pairs).

Dimension	Score (0-5), mean (SD; 95% CI)[a]	Score (0-5), median (IQR)	Score of 5, %	Score of ≥4, %	Score of ≤3, %
Relevance	4.83 (4.71-4.93)	5.00 (5.00-5.00)	91	96	4
Routing	4.89 (4.79-4.97)	5.00 (5.00-5.00)	94	97	3
Reference	4.50 (4.27-4.70)	5.00 (5.00-5.00)	81	90	10
Correctness	4.90 (4.78-4.99)	5.00 (5.00-5.00)	96	99	1
Identity	4.88 (4.78-4.96)	5.00 (5.00-5.00)	93	99	1

[a]Overall mean 4.80.

A comparison of multiturn with single-turn evaluations revealed consistent improvements (Table 3; details are provided in Multimedia Appendix 7); the overall average increased from 4.80 to 4.98 (+0.18). Topic centering and identity both achieved perfect scores of 5.00 in multiturn settings, indicating that the chatbot maintained natural conversation flow and a consistent professional tone across extended dialogues. Across all dimensions, all responses received scores of 4 or higher.

Table . Multiturn evaluation results (n=31 conversations).

Dimension	Score (0-5), mean (SD; 95% CI)[a]	Score (0-5), median (IQR)	Score of 5, %	Score of ≥4, %	Score of ≤3, %
Context memory	4.94 (4.84-5.00)	5.00 (5.00-5.00)	94	100	0
Topic centering	5.00 (5.00-5.00)	5.00 (5.00-5.00)	100	100	0
Identity	5.00 (5.00-5.00)	5.00 (5.00-5.00)	100	100	0

[a]Overall mean 4.98.

We compared the automated scores with human expert assessments to validate the reliability of the LLM-based evaluation. Three domain experts independently scored randomly selected subsets of conversations (n=20 question-answer pairs for single-turn evaluation and n=8 conversations for multiturn evaluation). We report 3 complementary agreement metrics: mean absolute difference (MAD) between averaged expert and LLM scores, Spearman rank correlation coefficient (ρ) for item-level ranking agreement, and intraclass correlation coefficient (ICC(3,1)) for interrater reliability among the 3 experts. Table 4 lists the validation results.

XSL•FO
RenderX

Table . Agreement with human experts and interrater reliability.

Check type and dimension	Rater 1	Rater 2	Rater 3	MAD[a]	Spearman ρ (95% CI)	ICC[b](3,1) (95% CI)
Single-turn evaluation (n=20 conversations)						
Relevance	0	0.10	0.05	0.05	0.609 (0.227 to 0.828)[c]	0.203 (−0.057 to 0.511)
Routing	0.40	1.25	0	0.55	__[d]	0.279 (0.008 to 0.575)
Reference	0.60	0.75	0.05	0.47	0.068 (−0.386 to 0.496)	0.170 (−0.085 to 0.481)
Correctness	0.05	0.20	0	0.08	—	−0.025 (−0.229 to 0.281)
Identity	0.15	0.05	0.55	0.25	—	0.314 (0.040 to 0.603)
Multiturn evaluation (n=8 conversations)						
Context memory	0.13	0.13	0.13	0.13	—	__[e]
Topic centering	0	0.13	0	0.04	—	0.000 (−0.307 to 0.545)
Identity	0	0.13	0.50	0.21	—	0.174 (−0.208 to 0.684)

[a]MAD: mean absolute difference. Overall MAD: 0.28 for single-turn evaluation and 0.13 for multiturn evaluation.

[b]ICC: intraclass correlation coefficient.

[c]$P<.01$.

[d]The large language model scores were all the same, and correlation could not be calculated.

[e]The 3 experts' scores were all the same, and the ICC could not be calculated.

For single-turn evaluation, relevance (MAD=0.05) and correctness (MAD=0.08) exhibited close LLM-human alignment, whereas routing (MAD=0.55) and reference (MAD=0.47) showed larger deviations. The overall MAD of 0.28 on a scale from 0 to 5 represents a deviation of less than 6% from human judgment. The Spearman ρ was computable only for relevance (ρ=0.609, 95% CI 0.227-0.828; P=.004) and reference (ρ=0.068; P=.78) as the LLM assigned uniform perfect scores for the remaining dimensions, leaving no variance for correlation analysis. This ceiling tendency of the LLM judge is itself a notable finding. Interrater reliability among the 3 experts was low to moderate (ICC range −0.025 to 0.314), indicating limited consensus even among human raters, particularly for correctness (ICC=–0.025) and reference (ICC=0.170). Per-rater analysis revealed that individual LLM-expert disagreement varied substantially (eg, routing MAD ranged from 0.00 to 1.25 across raters), suggesting that the observed LLM-human discrepancies partly reflect genuine interexpert disagreement rather than systematic LLM bias.

The multiturn evaluation exhibited closer overall alignment (MAD=0.13; 2.6% deviation), with topic centering achieving near-perfect agreement (MAD=0.04). All 3 experts assigned identical scores to every conversation for context memory, whereas topic centering (ICC=0.000) and identity (ICC=0.174) showed that the small deviations among raters did not follow consistent patterns—a consequence of near-ceiling score distributions. These results suggest that LLM-based evaluation serves as a reasonable proxy for human judgment on dimensions with clear-cut criteria (relevance and correctness) but should be interpreted cautiously for dimensions involving subjective judgment (routing and reference quality), where both the LLM and human raters exhibited greater variability.

Report Generation

The report generation system was evaluated for 4 distinct periods: January 2020, July 2020, September 2020, and October 2020. We provide an example report in Multimedia Appendix 8. For each period, the system generated complete reports by analyzing 30 days of data from all source collections. Three evaluators independently scored each report section, and we reported the mean with 95% bootstrap CIs across evaluators. Both main text assessment (completeness, correctness, and helpfulness) and reference validation (reference validity and citation correctness) protocols were applied to each report section. The detailed scoring results are presented in Multimedia Appendix 9.

Table 5 summarizes the main text and reference evaluation results. Completeness and correctness both exhibited strong ceiling effects, with medians of 5.00 (IQR 5.00-5.00) and 100% (48/48) of responses achieving scores of 4 or above. Helpfulness showed lower performance (median 4.00, IQR 4.00-4.00; mean 4.12, 95% CI 3.98-4.27), with only 19% (9/48) of responses achieving the maximum score, suggesting room for improvement in generating actionable insights. Reference validity achieved perfect scores across all evaluations (mean 5.00, 95% CI 5.00-5.00). Citation correctness was the most variable dimension (median 4.00, IQR 4.00 - 5.00; mean 4.21, 95% CI 3.96 - 4.46), with scores spanning the range of 3 to 5

XSL•FO
RenderX

and only 29% (7/24) of responses achieving perfect scores, indicating that the accuracy of citation-to-claim matching is the primary area for improvement in the report generation pipeline.

Table . Report main text and reference evaluation results.

Dimension	Score (0-5), mean (SD; 95% CI)	Score (0-5), median (IQR)	Score of 5, %	Score of ≥4, %	Score of ≤3, %
Completeness	4.83 (4.73-4.94)	5.00 (5.00-5.00)	83	100	0
Correctness	4.88 (4.77-4.96)	5.00 (5.00-5.00)	88	100	0
Helpfulness	4.12 (3.98-4.27)	4.00 (4.00-4.00)	19	94	6
Reference validity	5.00 (5.00-5.00)	5.00 (5.00-5.00)	100	100	0
Citation correctness	4.21 (3.96-4.46)	4.00 (4.00-5.00)	29	92	8

Discussion

Principal Findings

This study demonstrated the feasibility of a dual-purpose AI agent system for HPV vaccine information communication in Japan. This system integrates heterogeneous data sources (academic literature, government documents, news media, and social media) into a unified retrieval infrastructure that supports public-facing conversational interfaces and institutional analytical reporting.

In this preliminary evaluation, the chatbot achieved single-turn medians of 5.00 across all 5 dimensions (IQR 5.00-5.00), with means ranging from 4.50 (reference) to 4.90 (correctness). The pronounced ceiling effects indicate generally high performance but limit fine-grained differentiation between dimensions. Top-box analysis revealed that reference quality received lower scores (14/139, 10% at or below a score of 3 compared to a range of 2/139, 1% to 6/139, 4% for other dimensions), identifying citation handling as the primary area for improvement. Multiturn evaluation showed even stronger ceiling effects (overall mean 4.98), with topic centering and identity both achieving perfect scores across all 31 conversations. The overall mean improvement from single-turn to multiturn evaluation (+0.18) should be interpreted cautiously as it primarily reflects the elimination of low-scoring outliers rather than a broad performance shift. These preliminary results suggest that the iterative multitool orchestration architecture may maintain factual accuracy while delivering appropriately toned medical communication, although validation with real users is needed.

The report generation system maintained consistent quality across 4 temporal evaluation periods. Completeness and correctness exhibited strong ceiling effects (median 5.00, IQR 5.00-5.00), confirming reliable document structure and factual accuracy regardless of data availability fluctuations. Helpfulness was the most variable main text dimension (median 4.00, IQR 4.00-4.00; 9/48, 19% of responses achieving perfect scores), suggesting that, while the system reliably produces structurally complete and accurate reports, generating actionable institutional insights—particularly for chat analysis sections—remains a challenge. Citation correctness showed the widest spread among all report dimensions (median 4.00, IQR 4.00-5.00; 7/24, 29% of responses achieving perfect scores), indicating that the system consistently identifies valid sources (reference validity of median 5.00, IQR 5.00-5.00 across all periods) but is less consistent in matching citations to the specific claims they support. This distinction highlights an important design consideration for automated reporting systems: source retrieval and citation-claim alignment require different optimization strategies.

Interpretation

These findings suggest that LLM-based RAG systems may address the information asymmetry between medical institutions and vaccine-hesitant populations. The chatbot architecture differs from traditional static frequently asked question systems in that it dynamically selects and combines information from specialized knowledge sources, enabling responses that integrate academic evidence with official guidelines and contemporary public discourse.

The observed improvements in the multiturn evaluation merit consideration. The high context memory scores suggest that the controller incorporates information from previous turns, such as retaining user demographic information when providing age-specific recommendations. Topic centering scores suggest smooth transitions between related topics, resembling the natural progression of clinical consultations from symptoms to screening and prevention. These patterns indicate that the windowed conversation history approach provides sufficient context for a coherent extended dialogue. The low interrater reliability (Table 4) among human experts in single-turn evaluation (ICC range −0.025 to 0.314) may partly reflect the inherent ambiguity of judging individual question-answer pairs in isolation, where raters lack conversational context to disambiguate routing and reference quality assessments.

A 2-level citation validation mechanism may contribute to maintaining response quality. Reference scores confirm consistent source attribution, addressing a key concern in health information systems where users must verify claims independently. Although prospective studies are required to confirm this relationship, transparency may also contribute to user trust.

For institutional stakeholders, the report generation system offers capabilities that would otherwise require substantial manual effort. The consistent structural completeness across the evaluation periods and sections demonstrated reliable document generation regardless of data availability fluctuations. Per-section analysis (Multimedia Appendix 10) reveals that paper sections consistently achieved the highest scores, whereas

XSL•FO
RenderX

chat sections exhibited more variability in helpfulness. The September 2020 social media section maintained acceptable quality despite sparser data, suggesting robustness to temporal variation.

Japanese-language implementation addresses challenges specific to this context: multiscript processing, medical terminology localization, and culturally appropriate formal communication. The successful integration of Japanese government documents with English-language research literature provides initial evidence of the viability of this approach in settings where scientific evidence and public health communication occur in different languages.

Comparison With Prior Work

This work extends previous HPV vaccine natural language processing research, which has primarily focused on passive social media analysis [10], by implementing a bidirectional information flow. Prior health chatbots have been typically retrieved from single knowledge sources [11], whereas our iterative multitool architecture integrates 4 heterogeneous collections, enabling responses that synthesize information across source types.

The evaluation framework extends beyond typical single-turn RAG assessments [31,32] by incorporating a multiturn conversation analysis with simulated users. This approach captures the dimensions of conversational coherence, context memory, and topic continuity that single exchange evaluations overlook.

Limitations

This study has several limitations that warrant consideration. Most fundamentally, this work represents a proof-of-concept system evaluation rather than a clinical or public health impact study. The system was not tested with real patients, members of the public, or clinical end users in an operational setting. The evaluation relied on simulated conversations generated by 3 volunteers and report assessments by 3 evaluators over 4 periods, which can be insufficient to support strong generalizable conclusions. Additionally, the social media data were exclusively derived from Japanese X users, potentially underrepresenting the older adult population and those with limited digital access. In addition, the LLM-based evaluation may introduce biases that differ from human judgment, particularly for nuanced routing decisions in which multiple valid tool selections exist. The evaluation datasets (139 single-turn exchanges and 31 multiturn conversations) may not capture all real-world interaction patterns, and keyword-based data collection introduces potential selection bias. Geographic specificity to the Japanese government constrains transferability to national contexts with different regulatory frameworks and vaccination policies.

Future Directions

Several directions warrant further investigation. Future prospective studies with real users and measurable health outcomes are needed to assess the system's impact on vaccine knowledge, trust, and decision-making, whereas A/B testing will assess whether citation transparency affects user trust. Expansion to other languages and comparative studies across vaccine types would test the generalizability and inform the design principles for vaccine information communication systems facing similar challenges in other contexts.

Conclusions

This study demonstrated the initial feasibility of an AI agent system that simultaneously addresses public HPV vaccine information needs and institutional discourse monitoring in Japan. The integrated architecture may enable bidirectional information flow—providing verified information with transparent source attribution to users while generating analytical reports for institutional stakeholders—creating links between public concerns and communication strategies. Although this evaluation relied on simulated users, this study established a proof of concept for an AI-augmented vaccine information communication infrastructure, with the transferable architecture and evaluation frameworks providing foundations for adaptation to other vaccines and health conditions and multilingual public health contexts.

Acknowledgments

The authors used the generative artificial intelligence tool Claude [33] by Anthropic to refine the writing and structure of the manuscript, which was further reviewed and revised by the study group. The original Claude transcripts are available in Multimedia Appendix 11.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

Snapshots of data collection for chats, papers, and official documents or web sources used for retrieval-augmented generation, as well as social media IDs, are available in the Hugging Face repository [34].

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Single-turn chat scoring metrics.
[TXT File, 7 KB - infodemiology_v6i1e90295_app1.txt]

Multimedia Appendix 2
Multiple-turn chat scoring metrics.
[TXT File, 4 KB - infodemiology_v6i1e90295_app2.txt]

Multimedia Appendix 3
Guidance for the volunteers to chat with the chatbot with different personas.
[DOCX File, 15 KB - infodemiology_v6i1e90295_app3.docx]

Multimedia Appendix 4
Main text scoring metrics for the generated report.
[TXT File, 2 KB - infodemiology_v6i1e90295_app4.txt]

Multimedia Appendix 5
Reference scoring metrics for the generated report.
[TXT File, 1 KB - infodemiology_v6i1e90295_app5.txt]

Multimedia Appendix 6
Chat single-turn evaluation results.
[TXT File, 521 KB - infodemiology_v6i1e90295_app6.txt]

Multimedia Appendix 7
Chat multiple-turn evaluation results.
[TXT File, 397 KB - infodemiology_v6i1e90295_app7.txt]

Multimedia Appendix 8
Example of a generated report.
[PDF File, 537 KB - infodemiology_v6i1e90295_app8.pdf]

Multimedia Appendix 9
Scores for the generated reports.
[XLSX File, 17 KB - infodemiology_v6i1e90295_app9.xlsx]

Multimedia Appendix 10
Section-level evaluation results for the generated reports.
[TXT File, 1 KB - infodemiology_v6i1e90295_app10.txt]

Multimedia Appendix 11
Record of chat with artificial intelligence tool.
[DOCX File, 2310 KB - infodemiology_v6i1e90295_app11.docx]

References

1. Wu J, Jin Q, Zhang Y, et al. Global burden of cervical cancer: current estimates, temporal trend and future projections based on the GLOBOCAN 2022. J Natl Cancer Cent 2025;5(3):322-329. [doi: 10.1016/j.jncc.2024.11.006] [Medline: 40693230]
2. Drolet M, Bénard É, Boily MC, et al. Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. Lancet Infect Dis 2015 May;15(5):565-580. [doi: 10.1016/S1473-3099(14)71073-4] [Medline: 25744474]
3. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: definition, scope and determinants. Vaccine (Auckl) 2015 Aug 14;33(34):4161-4164. [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]
4. Simms KT, Hanley SJB, Smith MA, Keane A, Canfell K. Impact of HPV vaccine hesitancy on cervical cancer in Japan: a modelling study. Lancet Public Health 2020 Apr;5(4):e223-e234. [doi: 10.1016/S2468-2667(20)30010-4] [Medline: 32057317]

XSL•FO
RenderX

5. Johnson NF, Velásquez N, Restrepo NJ, et al. The online competition between pro- and anti-vaccination views. Nature 2020 Jun;582(7811):230-233. [doi: 10.1038/s41586-020-2281-1]
6. Sylvia Chou WY, Gaysynsky A, Cappella JN. Where we go from here: health misinformation on social media. Am J Public Health 2020 Oct;110(S3):S273-S275. [doi: 10.2105/AJPH.2020.305905] [Medline: 33001722]
7. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Advances in Neural Information Processing Systems: Curran Associates; 2020:9459-9474.
8. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
9. Kudo T, Yamamoto K, Matsumoto Y. Applying conditional random fields to Japanese morphological analysis. Presented at: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing p. 206-213.
10. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. JMIR Public Health Surveill 2020 May 29;6(2):e19273. [doi: 10.2196/19273] [Medline: 32427106]
11. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258. [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
12. Ujiie M. Resumption of active recommendation of the human papillomavirus vaccine in Japan and future challenges for the National Immunization Program. Hum Vaccin Immunother 2022 Nov 30;18(6):2090777. [doi: 10.1080/21645515.2022.2090777] [Medline: 35767827]
13. Yao S, Zhao J, Yu D, et al. React: synergizing reasoning and acting in language models. Presented at: 11th International Conference on Learning Representations, ICLR 2023; May 1-5, 2023 URL: https://collaborate.princeton.edu/en/publications/react-synergizing-reasoning-and-acting-in-language-models/ [accessed 2026-05-03]
14. Qdrant. URL: https://qdrant.tech/ [accessed 2025-11-29]
15. PLaMo-embedding-1B. Hugging Face. URL: https://huggingface.co/pfnet/plamo-embedding-1b [accessed 2025-11-29]
16. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. Presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); Nov 3-7, 2019. [doi: 10.18653/v1/D19-1410]
17. Johnson J, Douze M, Jegou H. Billion-scale similarity search with GPUs. IEEE Trans Big Data 2021;7(3):535-547. [doi: 10.1109/TBDATA.2019.2921572]
18. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2025-11-29]
19. World Health Organization. URL: https://www.who.int/ [accessed 2025-11-29]
20. Ministry of Health, Labour and Welfare, Japan. URL: https://www.mhlw.go.jp/ [accessed 2025-11-29]
21. Tweepy. URL: https://www.tweepy.org/ [accessed 2025-11-29]
22. LlamaIndex. URL: https://www.llamaindex.ai/ [accessed 2024-10-18]
23. Streamlit. URL: https://streamlit.io/ [accessed 2025-11-29]
24. Singh T, Roberts K, Cohen T, et al. Social media as a research tool (SMaaRT) for risky behavior analytics: methodological review. JMIR Public Health Surveill 2020 Nov 30;6(4):e21660. [doi: 10.2196/21660] [Medline: 33252345]
25. Broniatowski DA, Jamison AM, Qi S, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
26. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022. [doi: 10.5555/944919.944937]
27. Kudo T, Yamamoto K, Matsumoto Y. Applying conditional random fields to Japanese morphological analysis. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing: Association for Computational Linguistics; 2004:230-237.
28. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Inf Process Manag 1988;24(5):513-523. [doi: 10.1016/0306-4573(88)90021-0]
29. Řehůřek R, Sojka P. Software framework for topic modelling with large corpora. Presented at: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks; May 22, 2010. [doi: 10.13140/2.1.2393.1847]
30. Ethical guidelines for medical and biological research involving human subjects. Ministry of Education, Culture, Sports, Science and Technology, Ministry of Health, Labour and Welfare, Ministry of Economy, Trade and Industry, Japan. 2021. URL: https://www.mext.go.jp/content/20250325-mxt_life-000035486-01.pdf [accessed 2026-05-03]
31. Es S, James J, Espinosa-Anke L, Schockaert S. Ragas: automated evaluation of retrieval augmented generation. arXiv. Preprint posted online on Sep 26, 2023. [doi: 10.48550/arXiv.2309.15217]
32. Yu H, Gan A, Zhang K, Tong S, Liu Q, Liu Z. Evaluation of retrieval-augmented generation: a survey. arXiv. Preprint posted online on May 13, 2024. [doi: 10.48550/arXiv.2405.07437]
33. Claude. URL: https://www.anthropic.com/claude [accessed 2024-10-18]
34. Humanalysis-square/HPV_agent_data. Hugging Face. URL: https://huggingface.co/datasets/humanalysis-square/HPV_agent_data [accessed 2025-12-25]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
HPV: human papillomavirus
ICC: intraclass correlation coefficient
LLM: large language model
MAD: mean absolute difference
MeSH: Medical Subject Headings
MHLW: Ministry of Health, Labor, and Welfare
RAG: retrieval-augmented generation

Edited by T Mackey; submitted 25.Dec.2025; peer-reviewed by K Kim, SF Quadri, WJ Lo; revised version received 21.Mar.2026; accepted 21.Apr.2026; published 21.May.2026.

Please cite as:
Liu J, Yang S, Ma D, Niu Q, Zhang Z, Nagai-Tanima M, Aoyama T
Japanese-Language AI Agent System for Human Papillomavirus Vaccine Infoveillance and Public Communication: Development and Feasibility Evaluation
JMIR Infodemiology 2026;6:e90295
URL: https://infodemiology.jmir.org/2026/1/e90295
doi:10.2196/90295

© Junyu Liu, Siwen Yang, Dexiu Ma, Qian Niu, Zequn Zhang, Momoko Nagai-Tanima, Tomoki Aoyama. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 21.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review of the Quality and Reliability of Online Arabic Content on Diabetic Retinopathy: Infodemiological Study

Abdullah Ahmed Alabdrabulridha, MBBS; Dalal Mahmoud Alabdulmohsen, MBBS; Maryam Abdullah AlNajjar, MBBS; Ibtisam Ahmed Algouf, MBBS; Abdullah Mohammed Al-Omair, MBBS; Omaima Muneer Alyahya, MBBS; Mesa Ahmed Almahmudi, MBBS; Abdulaziz Ahmed Al Taisan, MBBS

Department of Ophthalmology, College of Medicine, King Faisal University, P.O. 380 Ahsaa, Hofuf, Saudi Arabia

Corresponding Author:
Abdullah Mohammed Al-Omair, MBBS
Department of Ophthalmology, College of Medicine, King Faisal University, P.O. 380 Ahsaa, Hofuf, Saudi Arabia

Abstract

Background: Diabetic retinopathy (DR) is a leading cause of vision loss, particularly in the Middle East. With the rise of online health information, many patients turn to the internet for knowledge about health conditions. However, the accuracy and quality of this information can be questionable, particularly in languages other than English.

Objective: We sought to evaluate the quality and reliability of Arabic websites on DR to address this knowledge gap and improve patient care.

Methods: The first 100 Arabic search results for DR were examined on Google, focusing on patient education websites in Arabic. Content was assessed using a 20-question model, quality was evaluated with the DISCERN instrument, and reliability was measured using the *Journal of the American Medical Association* (*JAMA*) benchmark. Two independent raters conducted evaluations, and data were analyzed with SPSS (IBM Corp). Descriptive statistics were used for website characteristics, and the first 10 Google web pages were compared to others using bivariate analysis with a significance level of $P<.05$.

Results: A Google search yielded 178,000 websites, and the first 100 were examined, with 29 meeting inclusion criteria. Most were hospital or medical center sites (n=20, 69%). The DISCERN assessment showed a low mean score of 36.59(SD 9.32) out of 80 points, with most rated “poor” or “very poor.” The *JAMA* benchmarks indicated low reliability, with 62% (18/29) failing to meet any criteria.

Conclusions: This study identified significant failings in the content, quality, and reliability of Arabic websites on diabetic retinopathy, highlighting the need for stronger evidence-based online resources focused on early disease prevention.

(JMIR Infodemiology 2026;6:e70514) doi:10.2196/70514

KEYWORDS

diabetic retinopathy; diabetes; online health information; arabic content; reliability; internet; online; quality; retinopathy; website; JAMA; DISCERN; health education; Journal of the American Medical Association

Introduction

Diabetic retinopathy (DR) is a major microvascular complication of diabetes mellitus, and it is considered one of the primary causes of permanent vision loss in adults and older individuals across the globe [1]. According to the World Health Organization, the Eastern Mediterranean region has the highest prevalence of diabetes mellitus worldwide. Moreover, it is estimated that DR cases are reaching up to 31% in the Eastern Mediterranean region, which is considered to be higher than the other regions on the globe [2].

Since online-based medical health information has become easily accessible to the population, it facilitates the searching and understanding of disease symptoms, risk factors, and treatment choices. It has been implemented that browsing the internet for health information has become a widespread part of the daily routine of individuals of all ages [3]. Interestingly, in today’s digital age, online search engines are the first resource for almost 75% of patients for medical conditions [4]. People tend to seek medical online information due to the low costs, being less time-consuming, and anonymity; however, online literature can provide mixed results that can be true or misleading, and this can be due to multiple factors, such as the accuracy, readability, and quality of the reported information [5]. Patients are increasingly using the internet to find health-related information that could influence medical decisions; however, there is a risk of encountering commercially influenced content. Hence, these findings indicate that the online information available on DR may not offer sufficient guidance for medical purposes [6].

The quality of online-based ophthalmological diseases was evaluated by several studies. Unfortunately, only a few studies

XSL•FO
RenderX

have documented the quality and reliability of the online content of DR [6]. The studies conducted so far have mainly centered on health-related material available in English, overlooking the need to evaluate the reliability and quality of online information in other languages. It must be mentioned that the main spoken language in the Middle East, which encounters a high rate of DR cases, is Arabic [7,8]. Due to the lack of highly standardized literature on online health content about DR in Arabic, this study seeks to address this gap by conducting a qualified evaluation of Arabic-language websites focusing on DR.

Methods

Study Aim, Design, and Setting

This cross-sectional website analysis was designed to evaluate the reliability and quality of Arabic online information about DR. Google.com, the most widely used search engine worldwide, held a 90.29% market share in Asia and a 97.19% market share in Africa as of May 2024. Arabic-speaking countries in the Middle East span both continents. For example, Saudi Arabia has a market share of 95.60% while Egypt has 97.38% [9]. The engine was used on May 1, 2024, to search for the Arabic term for DR, “ ,” in incognito mode using a new account to avoid browser bias. The first 100 search results were examined, simulating a patient’s or a general reader’s search behavior.

Inclusion and Exclusion Criteria

Websites were included if they were written primarily in Arabic and focused on providing educational content about DR for patients or the general public. Eligible websites were required to contain written text that addressed DR-specific topics, such as causes, symptoms, diagnosis, treatment options, and preventive strategies. The inclusion was limited to the first 100 search results to reflect typical patient behavior when searching for online health information.

Websites were excluded if they were not written in Arabic, targeted health care professionals (such as academic articles or clinical guidelines), or consisted solely of multimedia content like videos or audio recordings without accompanying text. In addition, websites that required login credentials, subscription access, or were otherwise inaccessible were excluded. Duplicate URLs among the first 100 search results were also removed. Finally, websites with irrelevant content, such as general diabetes pages lacking specific focus on DR, social media posts, news articles, blogs, forums, or advertisements, were excluded. The searching process is further illustrated in Figure 1.

Figure 1. Flowchart of the search process and results.

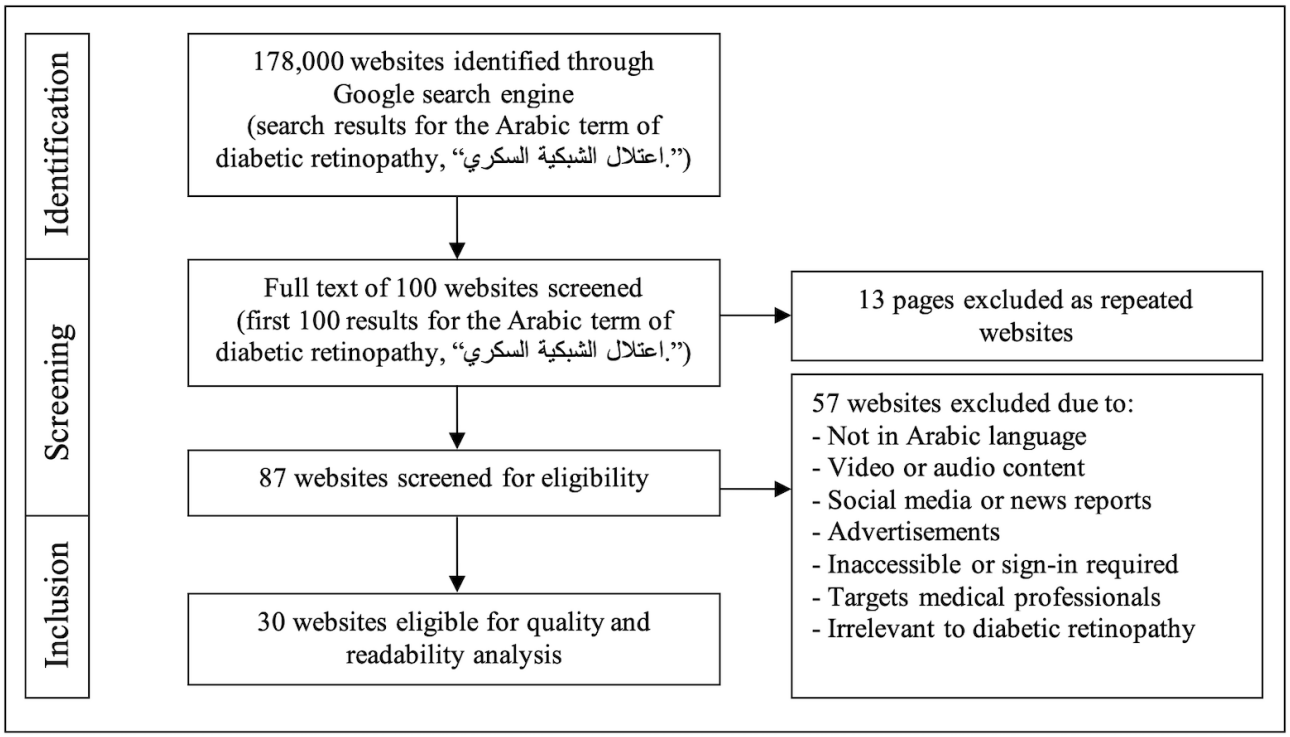


Table 1 exhibits the list of websites eligible for evaluation. Some websites, such as Mayo Clinic, have identical URLs in different languages but automatically localize content to Arabic. Websites were classified into four main categories: hospitals and medical centers, health portals (websites dedicated to health information), commercial (websites selling a service or product), and nonprofit organizations.

XSL•FO
RenderX

Table . Websites eligible for evaluation.

Website type	Website name	URL
Nonprofit organization	Mayo Clinic [10]	https://www.mayoclinic.org
Health portals	Webteb [11]	https://www.webteb.com
Health portals	MSD Manuals [12]	https://www.msdmanuals.com
Commercial	Altibbi [13]	https://altibbi.com
Hospitals and medical centers	Cleveland Clinic Abu Dhabi [14] https://www.msdmanuals.com	https://www.clevelandclinicabudhabi.ae
Hospitals and medical centers	Institut Català De Retina [15]	https://icrcat.com
Hospitals and medical centers	Barraquer UAE Eye Hospital [16]	https://www.barraquer.com
Hospitals and medical centers	Dünyagöz [17]	https://www.dunyagoz.com
Hospitals and medical centers	Bangkok Hospital [18]	https://www.bangkokhospital.com
Hospitals and medical centers	Med Care [19]	https://www.medcare.ae
Hospitals and medical centers	Northwest Eye Surgeons [20]	https://www.nweyes.com
Hospitals and medical centers	Dr. Haifa Eye Hospital [21]	https://www.drhaifaeyehospital.com
Hospitals and medical centers	King Khaled Eye Specialist Hospital [22]	https://pep.kkesh.med.sa
Hospitals and medical centers	Royal Spanish Center [23]	https://www.royalspanishcenter.com
Hospitals and medical centers	Jgemc [24]	http://www.jgemc.com
Commercial	Vezeeta [25]	https://www.vezeeta.com
Commercial	Ilajak [26]	https://www.ilajak.com
Hospitals and medical centers	Dr Mahmoud Hassaan [27]	https://www.drmahmoud-hassaan.com
Hospitals and medical centers	Eye City Center [28]	https://www.eyecitycenter.com
Nonprofit organization	Gulf Health Council [29]	https://www.ghc.sa
Health portals	Elconsolto [30]	https://www.elconsolto.com
Hospitals and medical centers	Alkahhal [31]	https://alkahhal.com.sa
Hospitals and medical centers	Ebsaar [32]	https://ebsaar.com
Hospitals and medical centers	Smart Laser Eye Center [33]	https://www.smartlasereyecenter.com
Hospitals and medical centers	Jordan Finland Modern Hospital [34]	https://jfmhospital.com
Hospitals and medical centers	Novomed [35]	https://www.novomed.com
Commercial	Tebcan [36]	https://tebcan.com
Hospitals and medical centers	Andalusia Clinic [37]	https://www.andalusiaclinic.com
Hospitals and medical centers	Dr-Oyoun [38]	https://dr-oyoun.com
Hospitals and medical centers	Magrabi Hospital [39]	https://www.magrabi.com

Assessment Tools

Content Assessment

To describe the content characteristics of the assessed websites, 20 questions were adapted from an established model by Kloosterboer et al [6] specific to DR websites. We used the model's questions to create a scoring system ranging from 0 (no content criteria fulfilled) to 100 (all content questions covered). Of the original 26 questions, 20 were selected for their relevance to the Arabic language content and patient-oriented information. Questions that were not directly applicable were excluded.

Quality Assessment

The DISCERN instrument, a well-known tool for evaluating patient-targeted content quality, was also used. This questionnaire contains 16 questions with clear instructions for objective assessment, with answers ranging from 1 (no sufficient answer) to 5 (efficiently answered) [40]. The total score ranges from 16 to 80 total points. The score categories were defined by Novin et al [40], who used this tool to assess DR online information for patients, to be as follows: "Excellent (75 - 63 points), Good (62 - 51 points), Average (50 - 39 points), Poor (38 - 28 points), and Very Poor (<28 points)."

XSL•FO
RenderX

Reliability Assessment

The *Journal of the American Medical Association* (*JAMA*) benchmark, used to evaluate website reliability, has 4 standards, namely authorship, attribution, currency, and disclosure. Each website receives 1 point when it fulfills a standard's criteria, with a maximum total score of 4 points. Websites scoring 3 or higher are considered highly reliable, while those scoring lower are considered low reliability [41].

Evaluation Process

The evaluation was performed by 2 independent raters (Evaluators A and B), followed by a shared revision session to determine a final rating and resolve any disagreements from the initial evaluation. The 2 raters were senior medical students trained in the use of the DISCERN and *JAMA* benchmarks, following a structured protocol to ensure consistency and objectivity. While not board-certified ophthalmologists, the raters applied the tools according to standardized instructions that do not require clinical expertise. Raters followed the instructions of the tools to objectively determine appropriate scores for each website. The final rating was derived from the initial independent evaluations and did not substantially deviate from either rater's original score. In all instances, the final rating either corresponded to one of the initial assessments or reflected a minor adjustment reached through consensus during the revision session.

Data Management and Analysis

All data was organized in an Excel (Microsoft) spreadsheet and imported into IBM SPSS (version 21; IBM Corp). Descriptive statistics, such as means and SDs, were used to summarize website characteristics and evaluation scores. Normality tests were conducted to determine the data distribution (normal or nonparametric), and appropriate statistical tests were applied. Because internet users are more likely to view websites on the first page of Google (the first 10 websites) [42], these were considered to be the most viewed and therefore grouped together to be compared with the websites of other pages using both the Mann-Whitney *U* test and independent 2-tailed *t* test according to the distribution of the data. The direction of association between the DISCERN and *JAMA* scores was examined using nonparametric measures, more specifically Spearman rank correlation. A *P* value below .05 was considered statistically significant. This study was reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [43].

Results

Website Selection Process and Overview

Our search for the Arabic term for DR using the Google search engine yielded 178,000 websites. Examining the first 100 revealed 13 duplicates, which were excluded, and further filtering of the remaining 87 for eligibility led to the exclusion of pages in languages other than Arabic, video or audio content only, inaccessible sites requiring visitors to sign in, those targeting medical professionals only, irrelevant topics, advertisements, social media posts, and news reports. Among the 30 pages that were eligible for further analysis, one more website was excluded and removed from the dataset as it became unavailable during the analysis process.

Of the 29 websites finally included, the majority (n=20, 69%) were classified as hospitals and medical centers. Commercial pages made up 4 (13.8%) websites. Around 3 (10.3%) websites fell under the health portal classification. The remaining 2 (6.9%) websites belonged to the foundation or nonprofit organization category.

Quality Assessment by DISCERN

The overall mean score of DISCERN for all websites was low (mean 36.59,SD 9.322), ranging from 17 to 61 points. None of the websites reached the "excellent" category (≥63 points). Only 2 of 29 (6.9%) websites scored in the "good" range (≥51 points). A total of 7 of 29 (24.1%) pages fell into the "average" category with a score ranging from 39 to 50 points. The majority of the included websites (16/29, 55.2%) had a "poor" score (38 - 28 points), and the remaining 4 of 29 (13.8%) pages achieved a "very poor" score of less than 28 points.

The overall quality score (DISCERN item 16) of the identified 29 websites had an average of 2.45 (SD 0.910). The DISCERN items with the highest scores were related to the relevancy of the content, the variety of treatment options, and how each treatment works (mean 4.59, SD 0.983; mean 4.00, SD 1.000; and mean 4.10, SD 1.263, respectively; Table 2).

XSL•FO
RenderX

Table . Average scores for each DISCERN item in evaluating 29 websites, ranging from 1 to 5.

Item	Score
1. Are the aims clear?	1.38
2. Does it achieve its aims?	1.55
3. Is it relevant?	4.59
4. Is it clear what sources of information were used to compile the publication (other than the author or producer)?	1.45
5. Is it clear when the information used or reported in the publication was produced?	1.48
6. Is it balanced and unbiased?	3.62
7. Does it provide details of additional sources of support and information?	1.66
8. Does it refer to areas of uncertainty?	1.10
9. Does it describe how each treatment works?	4.10
10. Does it describe the benefits of each treatment?	2.90
11. Does it describe the risks of each treatment?	1.90
12. Does it describe what would happen if no treatment is used?	1.69
13. Does it describe how the treatment choices affect overall quality of life?	1.17
14. Is it clear that there may be more than one possible treatment choice?	4.00
15. Does it provide support for shared decision-making?	1.55
16. Based on the answers to all of the above questions, rate the overall quality of the publication as a source of information about treatment choices	2.45

Several websites scored low in relation to revealing the explicit aims of their content and their source of information (mean 1.38,SD 1.015 and mean 1.45, SD 1.055, respectively). In addition, not many pages reported the date their content was created or published (mean score 1.48, SD 0.949). Few websites mentioned the possible risks of each treatment (mean score 1.90, SD 1.345), and almost none of the websites (1/29, 3.4%) highlighted any “gray” areas or uncertainties about the outcomes of the treatments (mean score 1.10, SD 0.557). Nearly all websites (27/29, 93.1%) scored 1 point out of 5 in the question related to the effect of treatment on quality of life (mean 1.17, SD 0.658; Table 2).

Reliability Assessment by *JAMA* Benchmarks

None of the websites obtained a score higher than 2 on the *JAMA* benchmarks tool (mean 0.45, SD 0.632), placing them all in the “low reliability” category. Over two-thirds (18/29, 62.1%) of the websites failed to meet any criteria, while the remaining websites (11/29, 37.9%) met only 1 or 2. The most common criterion met was currency, with 10 of 29 (34.5%) websites complying with it and an average score of 0.34 (SD 0.484). Authorship (mean 0.03, SD 0.186) and disclosure (mean 0.07, SD 0.258) were only displayed on 1 and 2 web pages (1/29, 3.4% and 2/29, 6.9%, respectively). No website has fulfilled the attribution benchmark. A very weak positive correlation between the DISCERN and *JAMA* scores was observed (ρ=0.130; *P*=.50).

Content Assessment

Out of 20 questions, a single website (1/29, 3.4%) answered the most, covering 75% of the questions. On the other hand, (2/29, 6.9% of the group) websites answered the least number of questions, only covering 15%. While most websites (27/29, 93.1%) explained what DR was and how it was treated, very few mentioned the screening period for DR (8/29, 27.6%). Only 2 of 29 (6.9%) websites discussed the reversibility of vision loss caused by DR. Moreover, 5 of 29 (17.2%) of the included websites covered the surgical options to treat DR and its possible risks. The vast majority (25/29, 86.2%) lacked images of DR (Table 3).

XSL•FO
RenderX

Table . Twenty questions about the content related to diabetic retinopathy with frequencies and percentage.

Questions	Yes, n (%)	No, n (%)
What is diabetic retinopathy?	27 (93.1)	2 (6.9)
What are the symptoms of diabetic retinopathy?	22 (75.9)	7 (24.1)
What is the difference between nonproliferative and proliferative diabetic retinopathy?	18 (62.1)	11 (37.9)
How is diabetic retinopathy diagnosed?	19 (65.5)	10 (34.5)
When should screening start?	8 (27.6)	21 (72.4)
What are the risk factors for diabetic retinopathy?	19 (65.5)	10 (34.5)
Can anything be done to reverse diabetic retinopathy?	9 (31)	20 (69)
What percentage of patients become legally blind from diabetic retinopathy?	2 (6.9)	27 (93.1)
How can vision loss be prevented?	11 (37.9)	18 (62.1)
Is vision loss reversible?	2 (6.9)	27 (93.1)
How is diabetic retinopathy treated?	27 (93.1)	2 (6.9)
What is panretinal photocoagulation, and what are the complications associated with it?	8 (27.6)	21 (72.4)
What is an anti-VEGF[a] injection and what are the complications associated with anti-VEGF therapy?	6 (20.7)	23 (79.3)
Are anti-VEGF injections or laser a cure or do they need to be repeated?	8 (27.6)	21 (72.4)
What are the surgical treatments for diabetic retinopathy and what are the potential complications?	5 (17.2)	24 (82.8)
What is tractional retinal detachment?	12 (41.4)	17 (58.6)
What is diabetic macular edema?	12 (41.4)	17 (58.6)
Are there any oral medications that can alter the progression of diabetic retinopathy?	1 (3.4)	28 (96.6)
Which age group is most commonly affected by diabetic retinopathy?	2 (6.9)	27 (93.1)
Does the source show pictures of diabetic retinopathy?	4 (13.8)	25 (86.2)

[a]VEGF: vascular endothelial growth factor.

XSL•FO
RenderX

Comparison of the Websites on the First Page and Additional Pages

The Mann-Whitney *U* test revealed no significant difference in *JAMA* scores between the first 10 websites (displayed on the first page) and those on subsequent pages (1-tailed *P*=.12). Similarly, bivariate analysis showed no significant difference in the DISCERN and content scores between the websites of the first page and other pages (*P*=.72). Detailed individual website scores for quality, reliability, and content are presented in Table 4.

XSL•FO
RenderX

Table . Included websites' scores on quality, reliability, and content.

Website name	Quality assessment by DISCERN (16-80)	Class	Reliability assessment by *JAMA*[a] (0-4)	Class	Content score (out of 20)	Content score (out of 100)
Mayo Clinic	55	Good	2	Low reliability	15	75
Webteb	37	Poor	1	Low reliability	11	55
MSD Manuals	38	Poor	2	Low reliability	9	45
Altibbi	44	Average	0	Low reliability	9	45
Cleveland Clinic Abu Dhabi	37	Poor	0	Low reliability	10	50
Institut Català De Retina	29	Poor	1	Low reliability	4	20
Barraquer Uae Eye Hospital	25	Very poor	0	Low reliability	7	35
Dünyagöz	17	Very poor	1	Low reliability	3	15
Bangkok Hospital	35	Poor	0	Low reliability	9	45
Med Care	35	Poor	0	Low reliability	9	45
Northwest Eye Surgeons	22	Very poor	0	Low reliability	3	15
Dr. Haifa Eye Hospital	38	Poor	0	Low reliability	10	50
King Khaled Eye Specialist Hospital	36	Poor	0	Low reliability	11	55
Royal Spanish Center	41	Average	0	Low reliability	13	65
Jgemc	30	Poor	0	Low reliability	5	25
Vezeeta	27	Very poor	0	Low reliability	6	30
Ilajak	41	Average	0	Low reliability	6	30
Dr Mahmoud Hassaan	41	Average	1	Low reliability	7	35
Eye City Center	48	Average	1	Low reliability	7	35
Gulf Health Council	29	Poor	0	Low reliability	5	25
Elconsolto	61	Good	0	Low reliability	5	25
Alkahhal	41	Average	0	Low reliability	12	60
Ebsaar	28	Poor	0	Low reliability	8	40
Smart Laser Eye Center	37	Poor	1	Low reliability	8	40
Jordan Finland Hospital	34	Poor	1	Low reliability	8	40
Novomed	35	Poor	1	Low reliability	5	25
Tebcan	37	Poor	1	Low reliability	5	25
Andalusia Clinic	34	Poor	0	Low reliability	6	30
Dr-Oyoun	49	Average	0	Low reliability	6	30

[a]*JAMA*: *Journal of the American Medical Association.*

XSL•FO
RenderX

Discussion

Principal Findings

Our study assessed the reliability and quality of the contents of 29 Arabic websites covering DR based on DISCERN and *JAMA* benchmarks. Using the DISCERN tool to evaluate the quality of each website has shown an overall poor quality of all websites evaluated, with an average score of 36.59 (SD 9.322). None of the 29 websites assessed achieved "excellent," with the majority of the websites falling under the "poor" (16/29, 55.2%) and "very poor" (4/29, 13.8%) categories. The content of the websites seems to falter in the first section of the DISCERN instrument focusing on the trustworthiness of the content presented (questions 1 - 8), which reflects the unreliability of the websites offering information on DR in Arabic. In addition, comprehensive reporting of treatment options with their associated benefits and risks was inconsistent between the websites, compromising the overall quality of the content presented. Our study found that the average score of the overall quality on the DISCERN tool is 2.45 (SD 0.9), which falls under the category of having "potentially important but not serious shortcomings." This outcome is similar to the results found by Novin et al [40], who evaluated the information about DR on US-based online websites and observed a mean score of 2.09 (SD 0.594), and similar to our results, no website was classified as "excellent" in their analysis. Moreover, both studies observed a deficiency in directing the readers to discuss their conditions with their physicians. While a perfect score of 5 signifies the highest quality, none of the analyzed websites achieved this benchmark.

Given the consistently poor quality and reliability scores across websites, these findings may be explained by several underlying factors. Hypotheses include (1) the absence of standardized quality controls in Arabic web content, (2) limited contributions from medical institutions in developing and maintaining patient education resources, and (3) a general lack of regulatory oversight governing the accuracy and reliability of health information published online. Such systemic gaps likely underlie the consistently low quality and reliability scores observed in our assessment.

Similar to the DISCERN assessment of the websites, assessing the 29 websites included with the *JAMA* benchmarks tool has revealed that all the websites are of low reliability. None of the websites evaluated obtained a score higher than 2, with an average score of 0.45 (SD 0.632). Comparing our results to the study done on dry eye disease websites, their assessment using *JAMA* has found that all the websites failed to meet half the benchmarks as well, with an average of 1.9 (SD 0.1) [44]. The low average in our study is due to more than two-thirds (18/29, 62.1%) of the websites assessed failing to meet any of the 4 *JAMA* benchmark criteria for a reliable website. A study conducted by Kloosterboer et al [6] on assessing online information regarding DR has also found that none of the websites evaluated has achieved all 4 *JAMA* benchmarks.

Upon evaluating each website's content, it was observed that no website has covered all questions, with the highest-scoring website only fulfilling 75% of the questions (n=15). While most websites included information on what DR is and how it is treated, many of them (20/29, 69%) lacked information about when screening should begin. The study by Novin et al [6] also noted this, pointing out that there was insufficient content on the DR screening intervals for type 1 and type 2 diabetes. A substantial number of the evaluated websites did not thoroughly address every therapeutic choice with its correlated risks and benefits. This was also a common finding in another study assessing DR-related internet resources for patients, where most websites scored in the lower range in questions related to treatment options.

The study found the average score for the balance and unbiasedness item of the DISCERN instrument to be 3.62 (SD 1.24) for Arabic websites covering DR. This rating may be attributed to the inclusion of commercial websites in the evaluation, despite excluding those explicitly labeled as "advertisement" or "ad" in Google search results. These commercial sites, while offering health services or products, also provide health educational content. The presence of commercial links and advertisements can hinder users' ability to locate reliable information, contributing to a "misinfodemic" and making it challenging for patients to access trustworthy health resources [45].

We believe that the gap found in the quality of content reported on all the assessed websites reflects the current health care practices in the region, where collaborative decision-making and thorough explanation of treatment risks and benefits may not always be given priority. A systematic literature review of patient-centered care (PCC) in the Middle East concluded that while there is support for adopting PCC in the Middle East and North African region, its implementation is still limited [46]. Webair identified barriers to adopting a PCC approach at multiple levels, mostly related to communication, suggesting a preference for a physician-driven approach that may not place as much emphasis on discussing treatment alternatives with patients [47].

Our study has its own set of limitations that require addressing for future investigation and research. First, Google was the primary search engine used in this study. Although it is a well-renowned and widely used search engine, the use of other search engines could provide more websites that were not taken into consideration by relying solely on Google [9]. The second limitation is the exclusive focus on written Arabic-language content. Video-based or multimedia educational resources, which are increasingly used by patients, were excluded from this analysis. Evaluating such content could provide further insights into the quality of health information consumed by the general population through popular platforms, such as WhatsApp (Meta Platforms Inc), YouTube (Google), or TikTok (ByteDance). While the quality of the content was assessed thoroughly in our study, assessing the readability of the websites included could provide insight into whether the information provided can be adequately comprehended by patients or not. Further studies should be performed in this area to assess whether websites are within the acceptable reading levels of online educational materials.

XSL•FO
RenderX

The study shows the need to improve the quality and reliability of Arabic-language online content on DR. Content creators should follow best practices in health communication, including clear authorship, proper source attribution, transparency regarding sponsorship, and regular content updates. Information should also be patient-centered, discussing treatment options, associated risks, and preventive measures like regular screening.

At the policy level, national guidelines are needed to ensure the quality of Arabic online health information. Policymakers may consider implementing standardized accreditation systems like the Health on the Net Code of Conduct, adapted to the region's linguistic and cultural context, to combat misinformation and enhance public health literacy.

Conclusion

Our study's findings disclose that Arabic-language websites providing information on DR treatment are significantly deficient in quality, reliability, and content. The DISCERN assessment tool displayed a "poor" score for most of the analyzed websites (16/29, 55.2%). In addition, according to the *JAMA* benchmark criteria, all of the websites showed low reliability, and none of them met the attribution benchmark. These findings highlight the necessity of enhancing the Arabic-language websites discussing DR treatment due to the high prevalence of this condition among patients in the Middle East and the insufficient high-quality online resources available. Middle Eastern health care organizations should collaborate to provide reliable, evidence-based online resources addressing this serious condition. While most websites discuss the treatment of DR, more information is needed on the associated risks and benefits, as well as a stronger focus on joint decision-making. In addition, Arabic websites should encourage screening and preventive measures to enhance patient outcomes.

Acknowledgments

The authors would like to express deep appreciation for all those who willingly contributed their time to this research.

Funding

We acknowledge the financial support for this research provided by the Deanship of Scientific Research at King Faisal University in Al-Ahsa, Kingdom of Saudi Arabia (grant number KFU241612).

Data Availability

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

Authors' Contributions

Conceptualization: AAA, DMA, MAA, IAA, AMA, OMA
Methodology: AAA, DMA, MAA, IAA, AMA, OMA
Investigation: AAA, DMA, MAA, IAA, AMA, OMA
Software: AAA, DMA, MAA, IAA, AMA, OMA
Writing–original draft: AAA, DMA, MAA, IAA, AMA, OMA
Writing–review & editing: AAAT
Supervision: AAAT
Project administration: AAAT
Funding acquisition: AAAT

Conflicts of Interest

None declared.

References

1. Chaudhari YS, Matkar SS, Bhujbal SS, Walunj VA, Bhor NS, Vyavhare RD. Diabetes mellitus: a review. Int J Adv Res Sci Commun Technol 2023 Mar 1:16-22. [doi: 10.48175/IJARSCT-8551]
2. Heiran A, Azarchehry SP, Dehghankhalili S, Afarid M, Shaabani S, Mirahmadizadeh A. Prevalence of diabetic retinopathy in the Eastern Mediterranean Region: a systematic review and meta-analysis. J Int Med Res 2022 Oct;50(10):3000605221117134. [doi: 10.1177/03000605221117134] [Medline: 36314851]
3. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790. [doi: 10.2196/34790] [Medline: 35171099]
4. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online health information seeking among us adults: measuring progress toward a healthy people 2020 Objective. Public Health Rep 2019;134(6):617-625. [doi: 10.1177/0033354919874074] [Medline: 31513756]
5. Singh S, Saini R, Sagar R. Quality of web-based information on attention deficit hyperactivity disorder. Asian J Psychiatr 2022 May;71:103071. [doi: 10.1016/j.ajp.2022.103071] [Medline: 35303590]

XSL•FO
RenderX

6. Kloosterboer A, Yannuzzi NA, Patel NA, Kuriyan AE, Sridhar J. Assessment of the quality, content, and readability of freely available online information for patients regarding diabetic retinopathy. JAMA Ophthalmol 2019 Nov 1;137(11):1240-1245. [doi: 10.1001/jamaophthalmol.2019.3116] [Medline: 31436789]
7. Abdulaziz Al Dawish M, Alwin Robert A, Braham R, et al. Diabetes mellitus in Saudi Arabia: a review of the recent literature. Curr Diabetes Rev 2016 Oct 26;12(4):359-368. [doi: 10.2174/1573399811666150724095130]
8. Albadrani MS, Alrehaili AM, Alahmadi SH, et al. Awareness of diabetic retinopathy among patients with type 2 diabetes mellitus in primary healthcare centers in Madinah, Saudi Arabia: a cross-sectional study. Cureus 2023 Nov;15(11):e49718. [doi: 10.7759/cureus.49718] [Medline: 38033448]
9. Browser market share worldwide. StatCounter Global Stats. URL: https://gs.statcounter.com [accessed 2024-06-28]
10. Diabetic retinopathy. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611 [accessed 2025-10-08]
11. Diabetic retinopathy. WebTeb. URL: https://www.webteb.com/diabetes/diseases/ - - [accessed 2025-10-08]
12. Mehta S. Diabetic retinopathy. MSD Manuals. URL: https://www.msdmanuals.com/home/eye-disorders/retinal-disorders/diabetic-retinopathy [accessed 2025-10-08]
13. Diabetic retinopathy. Altibbi. URL: https://altibbi.com/ - / - / - - [accessed 2025-10-08]
14. Diabetic retinopathy. Cleveland Clinic Abu Dhabi. URL: https://www.clevelandclinicabudhabi.ae/ar-ae/health-hub/health-resource/diseases-and-conditions/diabetic-retinopathy [accessed 2025-10-08]
15. Diabetic retinopathy. Institut Català de Retina. URL: https://icrcat.com/ar/ - - / - - / [accessed 2025-10-08]
16. Diabetic retinopathy. Barraquer UAE Eye Hospital. URL: https://www.barraquer.com/ar-uae/pathology/tll-lshbky-lskwry [accessed 2025-10-08]
17. Diabetic retinopathy. Dünyagöz. URL: https://www.dunyagoz.com/ar/medical-units/retina/diabetic-retinopathy [accessed 2025-10-08]
18. Do not overlook diabetic retinopathy as it can lead to complete vision loss!. Bangkok Hospital. URL: https://www.bangkokhospital.com/ar/bangkok/content/diabetic-retinopathy [accessed 2025-10-08]
19. Diabetic retinopathy. Medcare. URL: https://www.medcare.ae/ar/services/view/ophthalmology/diabetic-retinopathy.html [accessed 2025-10-08]
20. Diabetic retinopathy. Northwest Eye Surgeons. URL: https://web.archive.org/web/20231110015353mp_/https://www.nweyes.com/retina-vitreous-seattle/diabetic-retinopathy/ [accessed 2025-10-08]
21. Diabetic retinopathy. Dr Haifa Eye Hospital. URL: https://www.drhaifaeyehospital.com/ar/services/diabetic-retinopathy [accessed 2025-10-08]
22. Diabetic retinopathy. King Khaled Eye Specialist Hospital. URL: https://pep.kkesh.med.sa/?p=1446 [accessed 2025-10-08]
23. Learn about diabetic retinopathy: its causes and treatment methods. Royal Spanish Center. URL: https://royalspanishcenter.com/ - - - - - - /?lang=ar [accessed 2025-10-08]
24. Diabetic retinopathy. Jordan German Eye Center. URL: https://web.archive.org/web/20230321201534/http://www.jgemc.com/index.php/ar/consultancy-clinics/2016-07-23-22-08-43/8-department/54-diabetic-retinopathy [accessed 2025-10-08]
25. Diabetic retinopathy. Vezeeta. URL: https://www.vezeeta.com/ar/ - / / - - [accessed 2025-10-08]
26. Diabetic retinopathy: a comprehensive guide to symptoms, causes, and prevention. Ilajak Medical. URL: https://www.ilajak.com/ar/blog/diabetic-retinopathy [accessed 2025-10-08]
27. Prevention of diabetic retinopathy in 4 steps. Dr Mahmoud Hassaan. URL: https://www.drmahmoud-hassaan.com/blog/ - - - - - [accessed 2025-10-08]
28. Diabetic retinopathy. Al Sharq City Eye Center. URL: https://www.eyecitycenter.com/ar/blogs/23/aaatlal-alshbky-alskry [accessed 2025-10-08]
29. Diabetic retinopathy. Gulf Health Council. URL: https://yourhealthguide.ghc.sa/subjects/ - - -diabetic-retinopathy/ [accessed 2025-10-08]
30. Najah S. Neglecting it leads to blindness — 3 ways to treat diabetic retinopathy. Elconsolto. URL: https://www.elconsolto.com/eye-clinic/eye-clinic-news/details/2022/10/30/2315005/ - - -3- - - - - [accessed 2025-10-08]
31. Learn about how diabetic retinopathy. Alkahhal. URL: https://alkahhal.com.sa/ar/advices/ - - [accessed 2025-10-08]
32. Complications of diabetes. Ebsaar. URL: https://ebsaar.com/ar/services/diabetes-complications/ [accessed 2025-10-08]
33. Diabetic retinopathy. Smart Laser Eye Center. URL: https://www.smartlasereyecenter.com/ar/diabetic-retinopathy-disease/ [accessed 2025-10-08]
34. Diabetic retinopathy. Jordan Finland Modern Hospital. URL: https://web.archive.org/web/20240617224414/https://jfmhospital.com/ - - -diabetic-retinopathy/ [accessed 2025-10-08]
35. Diabetic retinopathy treatment in Dubai. Novomed. URL: https://www.novomed.com/services/specialized-clinics/ophthalmology/diabetic-retinopathy/ [accessed 2025-10-08]
36. Eye diseases: diabetic retinopathy. Tebcan. URL: https://tebcan.com/ar/Jordan/tebline/Admin/ /Articles/ - - - - - 2041 [accessed 2025-10-08]
37. Diabetic retinopathy. Andalusia Clinic. URL: https://andalusiaksa.com/ - - [accessed 2025-10-08]
38. Diabetic retinopathy. Dr-Oyoun. URL: https://dr-oyoun.com/ - - / [accessed 2025-10-08]

XSL•FO
RenderX

39. Diabetic retinopathy. Magrabi Health. URL: https://www.magrabihealth.com/blog/diabetic-retinopathy [accessed 2025-10-08]
40. Novin S, Konda SM, Xing B, Bange M, Blodi B, Burckhard B. Diabetic retinopathy online: a powerful opportunity for revision. J Consum Health Internet 2020 Jul 2;24(3):251-268. [doi: 10.1080/15398285.2020.1791668]
41. Ozduran E, Büyükçoban S. Evaluating the readability, quality and reliability of online patient education materials on post-COVID pain. PeerJ 2022;10:e13686. [doi: 10.7717/peerj.13686] [Medline: 35880220]
42. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577. [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
43. Elm EV, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808. [doi: 10.1136/bmj.39335.541782.AD]
44. Oydanich M, Kuklinski E, Asbell PA. Assessing the quality, reliability, and readability of online information on dry eye disease. Cornea 2022 Aug 1;41(8):1023-1028. [doi: 10.1097/ICO.0000000000003034] [Medline: 35344972]
45. Cai HC, King LE, Dwyer JT. Using the Google search engine for health information: is there a problem? Case study: supplements for cancer. Curr Dev Nutr 2021 Feb;5(2):nzab002. [doi: 10.1093/cdn/nzab002] [Medline: 33937613]
46. Alkhaibari RA, Smith-Merry J, Forsyth R, Raymundo GM. Patient-centered care in the Middle East and North African region: a systematic literature review. BMC Health Serv Res 2023 Feb 9;23(1):135. [doi: 10.1186/s12913-023-09132-0] [Medline: 36759898]
47. Laher I, editor. Handbook of Healthcare in the Arab World: Springer Nature; 2021. [doi: 10.1007/978-3-030-36811-1]

Abbreviations

DR: diabetic retinopathy
***JAMA*:** *Journal of American Medical Association*
PCC: patient-centered care
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by R Cuomo; submitted 23.Dec.2024; peer-reviewed by A Jafarizadeh, G Lim; revised version received 31.Aug.2025; accepted 16.Sep.2025; published 07.Jan.2026.

Please cite as:
Alabdrabulridha AA, Alabdulmohsen DM, AlNajjar MA, Algouf IA, Al-Omair AM, Alyahya OM, Almahmudi MA, Al Taisan AA
Review of the Quality and Reliability of Online Arabic Content on Diabetic Retinopathy: Infodemiological Study
JMIR Infodemiology 2026;6:e70514
URL: https://infodemiology.jmir.org/2026/1/e70514
doi:10.2196/70514

© Abdullah Ahmed Alabdrabulridha, Dalal Mahmoud Alabdulmohsen, Maryam Abdullah AlNajjar, Ibtisam Ahmed Algouf, Abdullah Mohammed Al-Omair, Omaima Muneer Alyahya, Mesa Ahmed Almahmudi, Abdulaziz Ahmed Al Taisan. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 7.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Shifting Public Discourse on Life-Sustaining Treatment in South Korea From 2014 to 2025: Longitudinal Social Media Analysis

Miryoung Kim[1,2,3], PhD; Hae Sun Suh[4,5,6], PhD

[1]College of Pharmacy, Sunchon National University, 255, Jungang-ro, Suncheon, Republic of Korea

[2]Research Institute of Life and Pharmaceutical Sciences, Sunchon National University, Suncheon, Republic of Korea

[3]AI Bio Autonomous Research Core Facility Center, Sunchon National University, Suncheon, Republic of Korea

[4]College of Pharmacy, Kyung Hee University, Seoul, Republic of Korea

[5]Department of Regulatory Science, Graduate School, Kyung Hee University, Seoul, Republic of Korea

[6]Institute of Regulatory Innovation through Science, Kyung Hee University, Seoul, Republic of Korea

Corresponding Author:
Miryoung Kim, PhD
College of Pharmacy, Sunchon National University, 255, Jungang-ro, Suncheon, Republic of Korea

Abstract

Background: Public discourse plays an important role in shaping how end-of-life care and medical decision-making are understood within society. In South Korea, the implementation of the Life-Sustaining Treatment Decision Act in 2018 established a legal framework for withholding or withdrawing life-sustaining treatment for patients at the end of life. However, relatively little is known about how public discourse surrounding life-sustaining treatment has evolved over time, particularly before and after the act's implementation.

Objective: This study aimed to examine how online discourse related to life-sustaining treatment in South Korea evolved using large-scale social media data by identifying dominant themes, tracing their temporal changes from 2014 to 2025, and assessing shifts in sentiment before and after the act.

Methods: A total of 16,191 text documents, comprising posts from Naver blogs, Naver cafes, and comments from YouTube, were collected using keywords related to life-sustaining treatment and end-of-life decision-making. Text was preprocessed using the Kiwi tokenizer for Korean morphological analysis, and latent Dirichlet allocation topic modeling was applied to extract underlying themes. A 6-topic solution was selected based on topic coherence and interpretability. Temporal changes in topic prevalence were examined across the study period, and topic distributions were compared between the prelaw (2014 - 2017) and postlaw (2018 - 2025) periods. Sentiment analysis was conducted using the Kunsan National University Korean sentiment lexicon to classify documents into positive, neutral, and negative categories. A study-defined descriptive net sentiment change index was calculated to summarize changes in positive and negative sentiment following policy implementation.

Results: Six major themes were identified: death with dignity, hospice and palliative care, aging and social care, end-of-life decision-making, euthanasia debate, and withdrawal of treatment. Discourse volume increased more than 17-fold over the study period. Prior to 2018, discussions were dominated by end-of-life decision-making (255/573, 44.5%) and withdrawal of treatment (178/573, 31.1%). Around the time of policy implementation, ethical and legal debates became more prominent. In subsequent years, discourse increasingly reflected care-oriented perspectives, including hospice and palliative care and aging and social care. Sentiment improved most markedly for end-of-life decision-making and hospice and palliative care following the act, whereas aging and social care showed minimal change, and death with dignity exhibited a decline in net sentiment.

Conclusions: Public discourse on life-sustaining treatment in South Korea evolved from a clinical focus to ethical debate and, more recently, to care-oriented perspectives. These changes were temporally associated with the implementation period of the Life-Sustaining Treatment Decision Act but should not be interpreted as causal effects of the legislation. The findings highlight the value of social media data for monitoring policy-associated changes in public discourse on end-of-life care.

(JMIR Infodemiology 2026;6:e95918) doi:10.2196/95918

KEYWORDS

life-sustaining treatment; terminal care; palliative care; social media; natural language processing; sentiment analysis; public opinion; health policy

XSL•FO
RenderX

Introduction

End-of-life care and life-sustaining treatment decisions lie at a critical intersection of medicine, ethics, law, and culture. As populations age, questions surrounding the appropriate use of life-prolonging interventions and dignified dying have increasingly entered the public sphere [1,2]. Public discourse on these topics both reflects and shapes societal values, informs policy development, and influences how individuals and families navigate end-of-life decisions [3].

In South Korea, the Life-Sustaining Treatment Decision Act, enacted in 2018, established the first comprehensive national framework for the withholding or withdrawal of life-sustaining treatment at the end of life [4,5]. The act formally recognized patients' right to autonomous end-of-life decisions; introduced structured mechanisms, including advance directives and a designated decision-making system; and defined the legal conditions under which treatment withdrawal could be authorized, marking a fundamental shift in the institutional approach to end-of-life care in South Korea. Following full implementation in 2018, the actual withdrawal of life-sustaining treatment in clinical settings increased markedly, and public discussion of death, dying, and end-of-life decision-making became more visible in the same period [6-8].

Social media platforms have emerged as important venues for capturing naturalistic public discourse on sensitive health topics [9,10]. Large-scale text data from platforms such as Naver and YouTube, when analyzed using computational methods including topic modeling and sentiment analysis, can reveal shifts in thematic focus and emotional tone that conventional surveys cannot [10-12]. However, despite growing interest in end-of-life attitudes in South Korea, the longitudinal evolution of online discourse—before and after the 2018 act—has received limited scholarly attention [13,14]. Sentiment analysis in health-related social media research can be broadly classified into categorical and dimensional approaches [15]. Categorical sentiment analysis characterizes emotional tone using discrete classes such as positive, neutral, and negative sentiment, whereas dimensional sentiment analysis characterizes affective meaning along continuous dimensions such as valence, arousal, and dominance. Although categorical approaches are useful for interpretable comparisons across large longitudinal corpora, dimensional approaches may provide a more nuanced representation of complex emotional states. This distinction is particularly relevant to end-of-life discourse, where expressions of sadness, fear, relief, acceptance, dignity, and moral concern may coexist in the same discussion.

To address this gap in the longitudinal understanding of online discourse on life-sustaining treatment before and after the act, we analyzed social media data from 2014 to 2025 using topic modeling and lexicon-based sentiment analysis. We aimed to (1) identify dominant themes in public discourse on life-sustaining treatment, (2) trace temporal changes in topic prevalence across the study period, and (3) assess sentiment shifts before and after the implementation of the act. Findings are discussed in a comparative context with end-of-life discourse patterns observed in Western countries.

Methods

Study Design

This study used a computational text analysis approach to examine changes in online discourse related to the life-sustaining treatment decision system in South Korea. We focused on how discourse evolved before and after the implementation of the Life-Sustaining Treatment Decision Act in 2018, which introduced a national framework for end-of-life treatment decisions. Drawing on large-scale social media data, we explored changes in both the thematic structure and emotional tone of public discourse over time. The analysis covered the period from January 2014 to December 2025, allowing for a longitudinal comparison of discourse patterns before and after the policy was implemented.

Data Sources

Text data were collected from 3 widely used Korean online platforms—Naver blogs, Naver cafes, and YouTube comments. Naver is a major Korean web portal and search ecosystem that integrates searches, news, blogs, online communities, maps, shopping, and other services. Because Naver functions not only as a search engine but also as a platform for user-generated content, Naver blogs and Naver cafes have been widely used in South Korea for personal narratives, informational posts, and community-based discussions. Naver blogs generally contain longer personal or informational posts, whereas Naver cafes are topic- or community-based online spaces in which users discuss shared interests, experiences, or concerns. YouTube comments, in contrast, are typically shorter and more reactive, reflecting responses to video content, news, public education materials, or policy-related discussions.

Data were collected using platform-provided APIs rather than direct scraping of private or restricted pages. Naver blog and cafe posts were retrieved using the Naver Open API search end points for blogs and cafe articles, and YouTube videos and comments were retrieved using the YouTube Data API. Searches were conducted using Korean keywords related to life-sustaining treatment and end-of-life decision-making, including terms related to withdrawal or withholding of treatment, hospice and palliative care, death with dignity, euthanasia, and end-of-life decision-making (Table S1 in Multimedia Appendix 1). The platform-specific data collection procedures are summarized in Table S2 in Multimedia Appendix 1.

The representativeness of Naver-based data should be interpreted in light of changes in the Korean online environment during the study period. From 2014 to 2025, online communication in South Korea became increasingly mobile and platform diversified, with greater use of video-based and global social media platforms. Accordingly, Naver blog and cafe data should not be interpreted as representing the Korean general population. Rather, they capture an important segment of Korean-language online discourse, particularly longer-form narrative and community-based discussions on life-sustaining treatment.

For Naver, searches were conducted separately for each platform, keyword, and year. The search query combined each base keyword with the target year, and retrieved records were

XSL•FO
RenderX

retained only if the publication date matched the target year. Duplicate Naver records were removed based on URL. For YouTube, keyword-year searches were first used to identify relevant videos through the YouTube Data API. Video metadata, including title, description, publication date, channel title, view count, like count, and comment count, were retrieved when available. Videos were retained only if the upload year matched the target year, and duplicates were removed using video IDs. Top-level comments were then collected from the included videos using the comment thread end point, with a maximum of 30 comments per video ordered by relevance. Videos with disabled comments were skipped, and duplicate comments were removed using comment IDs. Only publicly accessible posts and comments were collected. Usernames and other direct personal identifiers were not analyzed, and results are reported only in aggregate form.

Data Preprocessing

Prior to analysis, several preprocessing steps were undertaken to prepare the textual data. Duplicate documents and entries not directly related to life-sustaining treatment were removed to ensure the relevance of the dataset. The text was then cleaned by removing URLs, punctuation, special characters, and other nonlinguistic elements.

Given the morphological complexity of the Korean language, we performed morphological analysis using the Kiwi tokenizer to segment sentences into meaningful linguistic units [16]. Tokens corresponding to nouns, verbs, and adjectives were retained, whereas stop words and other function words with limited semantic value were excluded. The resulting tokens were used to construct a document-term matrix, which served as the basis for subsequent topic modeling and sentiment analysis.

Topic Modeling

To identify dominant themes in online discourse, we applied latent Dirichlet allocation (LDA) topic modeling to the preprocessed corpus [17]. LDA is an unsupervised machine learning approach that identifies latent topics within large text collections by estimating the distribution of words across topics and documents. Candidate LDA models with 1 to 20 topics were evaluated using the coherence metric [18]. Although the 4-topic model showed the highest coherence score (0.4789), the 6-topic model showed a comparable coherence score (0.4754) and provided greater substantive granularity for distinguishing clinically and ethically adjacent themes. Therefore, the 6-topic solution was selected based on both coherence and interpretability. The final model was estimated using the Gensim LdaModel with a random state of 42 and 20 passes. Default alpha and eta settings were used. Sensitivity analyses using alternative random seeds indicated that the major thematic structure of the 6-topic solution was broadly retained. Core themes, including hospice and palliative care, end-of-life decision-making, euthanasia debate, and aging and social care, showed consistent high-probability keywords across runs. Some keyword overlap was observed among conceptually adjacent topics, particularly withdrawal of treatment, end-of-life decision-making, and euthanasia debate, reflecting the close substantive relationship among these themes in end-of-life discourse (Tables S3 and S4 in Multimedia Appendix 1).

Topic labels were assigned through an iterative interpretive review process. The first author reviewed the highest-probability keywords and representative documents for each topic and proposed preliminary labels. The labels were then reviewed and refined through author discussion, with reference to the original Korean text and the substantive meaning of representative documents. The final labels were intended to summarize the dominant thematic meaning of each topic rather than defining mutually exclusive conceptual categories.

Temporal Analysis

To examine how discourse evolved over time, we analyzed topic prevalence on an annual basis across the study period. Annual topic proportions were calculated by aggregating document-level topic probabilities within each year, allowing us to trace temporal patterns in discourse. Notably, a larger volume of data was collected in more recent years, reflecting the growth of online platform use and increased user activity over time. To account for this imbalance, all temporal analyses were conducted using normalized topic proportions rather than raw document counts.

Policy Period Comparison

To assess changes in discourse associated with the implementation of the Life-Sustaining Treatment Decision Act, we divided the dataset into 2 periods: before implementation (2014 - 2017) and after implementation (2018 - 2025). Topic distributions were compared between these periods to examine shifts in the thematic focus of online discussions following the introduction of the life-sustaining treatment decision system.

Sentiment Analysis and Change Assessment

We adopted a categorical lexicon-based sentiment approach, classifying the overall emotional tone of each document into positive, neutral, or negative categories. This approach was selected because it provides a transparent and reproducible framework for comparing aggregate sentiment patterns across topics and periods in a large Korean-language social media corpus. A range of lexicon-based sentiment resources has been developed for different languages and affective representations, including English-language polarity lexicons such as the Semantic Orientation Calculator and AFINN; dimensional affective lexicons such as the Affective Norms for English Words and its extended version, the National Research Council Canada Valence, Arousal, and Dominance Lexicon; and language-specific resources such as the Chinese EmoBank [19-24]. However, these resources are not directly applicable to Korean-language social media text because sentiment-bearing expressions are language and culture dependent. Applying non-Korean lexicons would require translation or cross-lingual mapping, which may alter affective meaning, particularly in culturally sensitive discourse on death, dignity, family caregiving, and end-of-life decision-making. Therefore, we used the KNU Korean sentiment lexicon, a Korean-language sentiment resource, to classify sentiment in a manner consistent with the linguistic characteristics of the corpus [25].

XSL•FO
RenderX

The KNU Korean sentiment lexicon is a general-purpose Korean sentiment resource and has not been specifically developed or validated for end-of-life care discourse. Accordingly, sentiment classification in this study was interpreted as a descriptive indicator of aggregate emotional tone rather than as a direct measure of individual affective states.

To summarize topic-specific sentiment changes between the pre- and postimplementation periods, we calculated a study-defined descriptive net sentiment change index. For each topic, the index was defined as (positivepost – positivepre) – (negativepost – negativepre), where "positivepost" and "positivepre" indicate the percentage of documents classified as positive in the post- and preimplementation periods, respectively, and "negativepost" and "negativepre" indicate the corresponding percentages of documents classified as negative in the post- and preimplementation periods, respectively. Positive values, therefore, indicate an increase in positive sentiment, a decrease in negative sentiment, or both. Negative values indicate a decrease in positive sentiment, an increase in negative sentiment, or both. This index was used as a descriptive summary of direction and magnitude rather than as an externally validated psychometric measure.

All analyses, including data preprocessing, topic modeling, temporal analysis, and sentiment analysis, were conducted using Python (version 3.14.3; Python Software Foundation).

Ethical Considerations

This study used publicly available, deidentified data and did not involve human participants as defined by institutional guidelines. Therefore, institutional review board approval was not required.

Results

Overview of the Dataset

A total of 16,191 social media documents related to life-sustaining treatment were collected from Naver and YouTube between 2014 and 2025 after preprocessing. The volume of discourse increased markedly over time. The number of posts rose from 178 in 2014 to 3160 in 2025, representing a more than 17-fold increase. A substantial rise was observed after 2020, with annual counts exceeding 500 posts from 2021 onward and peaking in 2025. The dataset was heavily concentrated in the postpolicy period. Of the 16,191 total documents, 15,618 (96.5%) were posted after the implementation of the Life-Sustaining Treatment Decision Act (2018 - 2025), whereas only 573 (3.5%) were from the prepolicy period (2014 - 2017). In terms of platform distribution, YouTube comments accounted for 52.6% (n=8513) of the documents, whereas Naver blogs and cafes constituted 47.4% (n=7678) of the documents. Notably, prepolicy data were drawn exclusively from Naver blogs, whereas postpolicy data included both platforms, reflecting the increasing role of video-based platforms in public discourse over time.

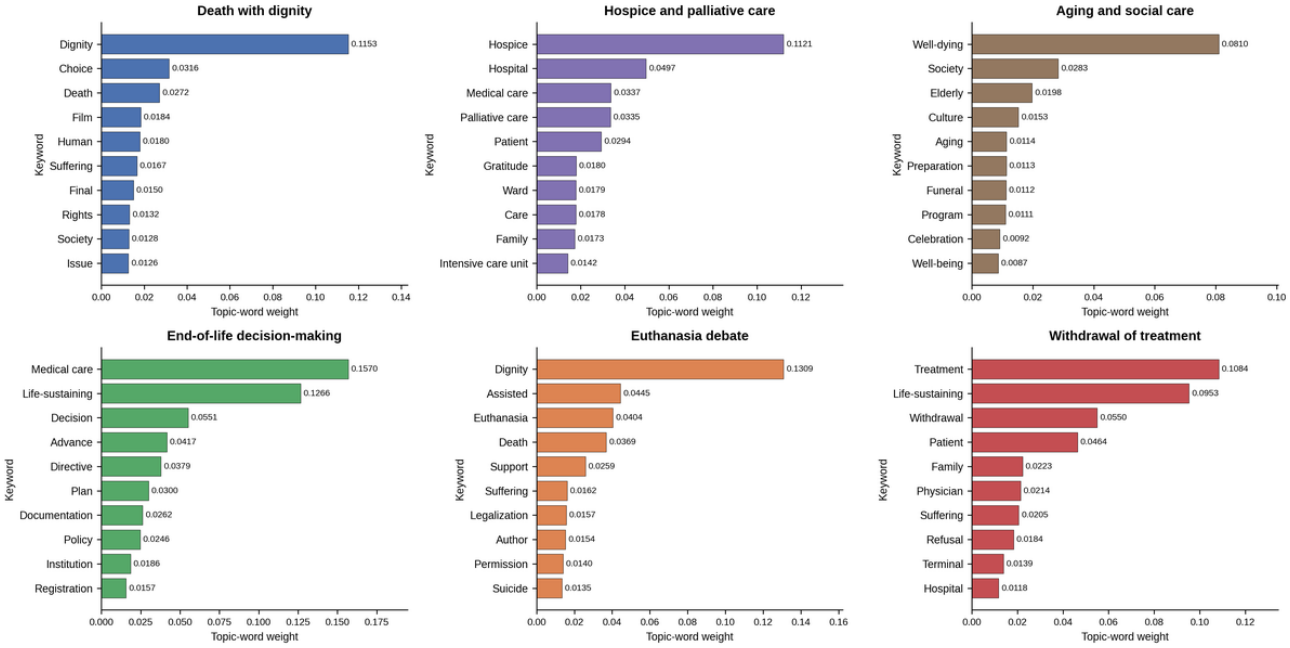
Topic Characteristics

LDA identified 6 distinct topics representing major dimensions of public discourse: death with dignity, hospice and palliative care, aging and social care, end-of-life decision-making, euthanasia debate, and withdrawal of treatment (Figure 1). Each topic was characterized by a coherent set of frequently co-occurring keywords. Some keywords were shared across topics, reflecting their use in multiple contexts within end-of-life discourse.

The death with dignity topic was primarily associated with high-weight terms such as "dignity," "choice," and "death," reflecting discussions centered on autonomy and the right to self-determined end-of-life decisions. The hospice and palliative care topic was dominated by terms such as "hospice," "hospital," and "palliative care," indicating a focus on care provision in clinical settings. The aging and social care topic included high-weight terms such as "well-dying," "society," and "elderly," suggesting a broader societal perspective on aging and end-of-life preparation. The end-of-life decision-making topic was characterized by terms such as "medical care," "life-sustaining," and "decision," reflecting institutional and procedural aspects of treatment decision-making. The euthanasia debate topic was defined by terms such as "dignity," "assisted," and "euthanasia," indicating ongoing ethical and legal debates surrounding assisted death. Finally, the withdrawal of treatment topic was associated with terms such as "treatment," "life-sustaining," and "withdrawal," highlighting discussions related to the discontinuation of medical interventions. As the analysis was conducted using Korean-language data, the extracted keywords were translated into English to facilitate interpretation and presentation. The full list of Korean-English keyword mappings for each topic is provided in Table S2 in Multimedia Appendix 1.

XSL•FO
RenderX

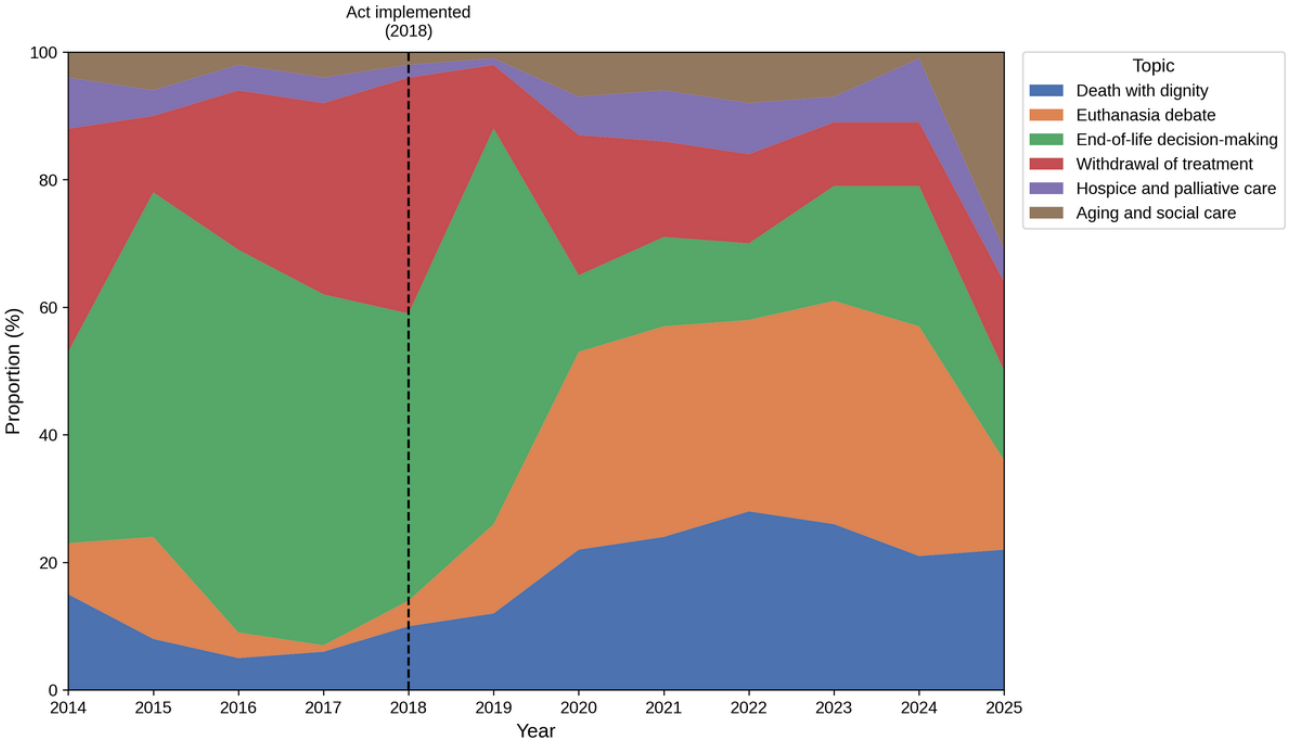
Figure 1. Top 10 keywords and their weights for each of the 6 identified topics in online discourse on life-sustaining treatment in South Korea (2014 - 2025).



Temporal Trends in Topic Prevalence

Temporal analysis revealed substantial shifts in topic prevalence over the study period (Figure 2). Prior to 2018, discourse was dominated by the end-of-life decision-making and withdrawal of treatment topics, which together accounted for most discussions, often exceeding 60% to 70% of total topic proportions. Around 2018, a marked transition was observed. The relative dominance of the end-of-life decision-making topic declined, whereas discussions on topics related to death with dignity and euthanasia debate increased, indicating a shift toward ethical and normative discussions during the policy implementation period. From 2020 onward, the discourse became more diversified. The proportion of hospice and palliative care and aging and social care topics increased steadily, whereas the previously dominant clinical topics declined. By the later years, discourse was more evenly distributed across topics, with care-oriented and social themes occupying a larger share. These patterns indicate a transition from clinically focused discussions toward broader ethical debates and, subsequently, toward care-oriented and social perspectives.

Figure 2. Temporal changes in topic proportions. The dashed line indicates the implementation of the Life-Sustaining Treatment Decision Act in 2018.

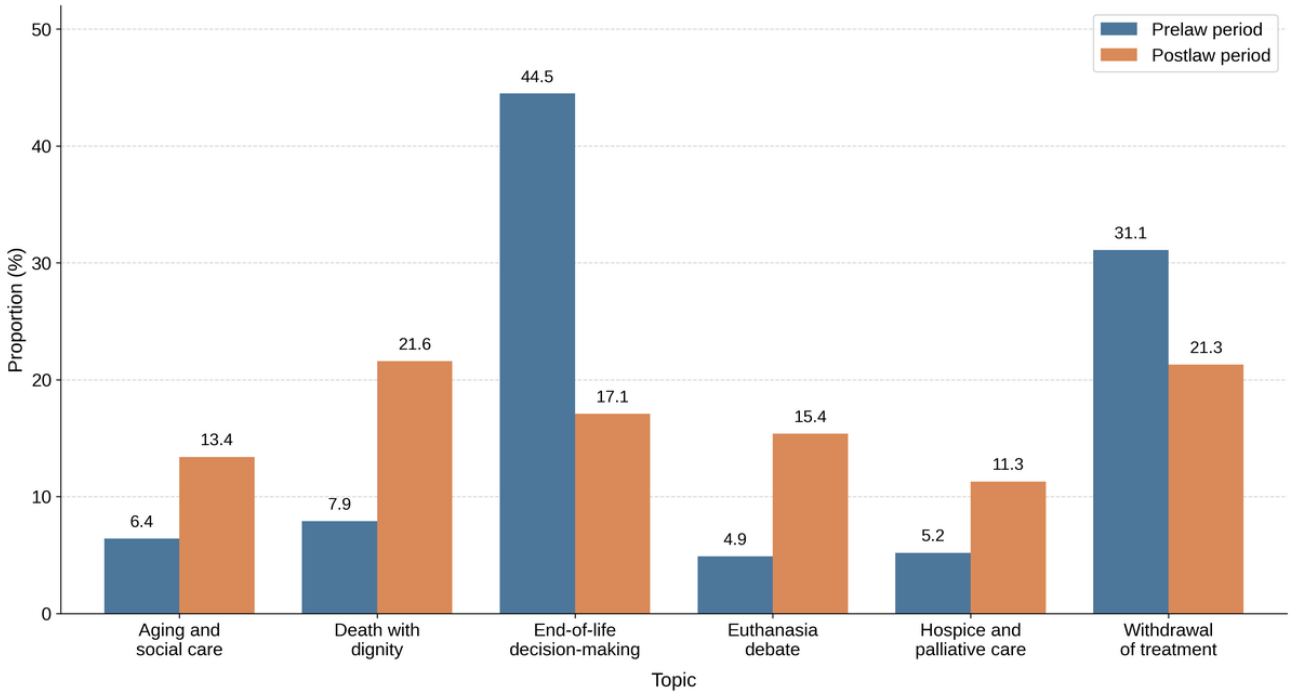


XSL•FO
RenderX

Changes in Topic Distribution Before and After Policy Implementation

Comparisons between the prelaw (2014 - 2017) and postlaw (2018 - 2025) periods further illustrate these shifts (Figure 3). In the prelaw period, end-of-life decision-making was the most prominent topic, accounting for the largest share of discourse, followed by withdrawal of treatment. In contrast, the postlaw period showed a clear redistribution of topic prevalence. The share of end-of-life decision-making decreased substantially (255/573, 44.5% to 2671/15,618, 17.1%), whereas death with dignity showed the largest relative increase (45/573, 7.9% to 3373/15,618, 21.6%), followed by euthanasia debate (28/573, 4.9% to 2405/15,618, 15.4%) and hospice and palliative care (30/573, 5.2% to 1765/15,618, 11.3%). These changes indicate temporal shifts in online discourse around and after the implementation period of the act from a narrower focus on clinical decision-making toward broader discussions encompassing ethical considerations, care practices, and social contexts.

Figure 3. Topic distribution before and after implementation of the Life-Sustaining Treatment Decision Act. Proportions are compared between the prelaw (2014 - 2017) and postlaw (2018 - 2025) periods.

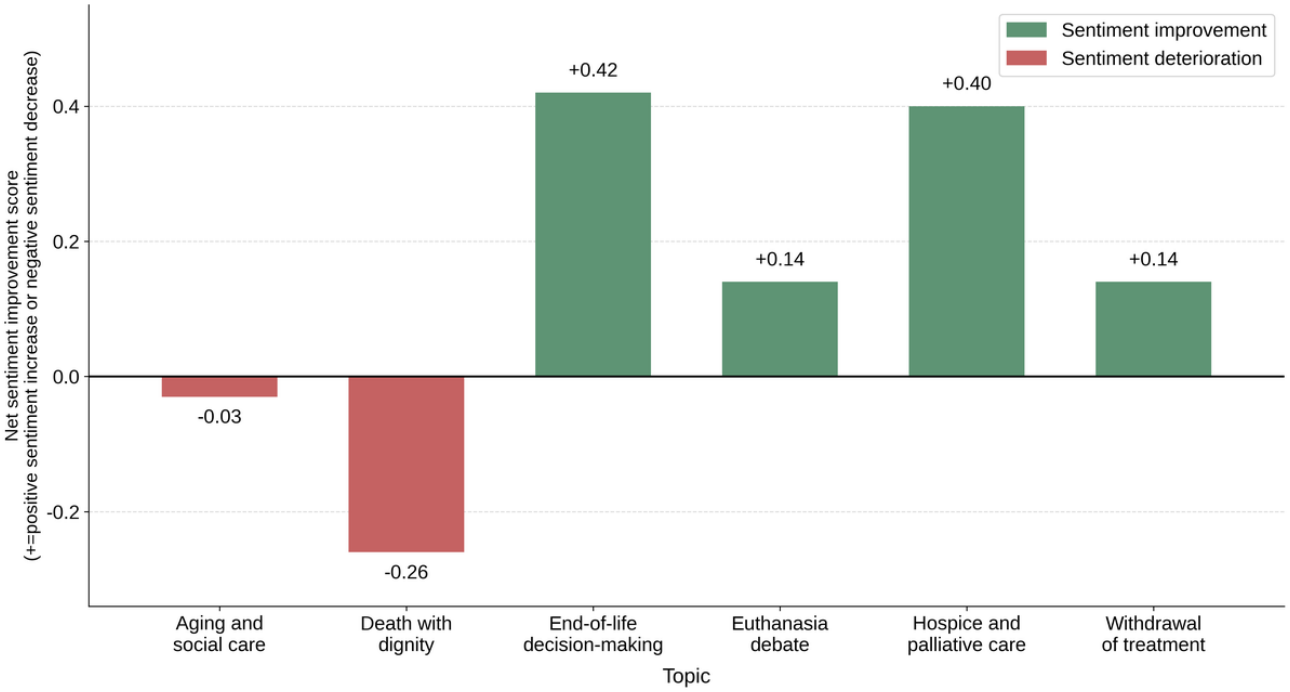


Sentiment Distribution and Changes After Policy Implementation

To examine sentiment changes following the implementation of the Life-Sustaining Treatment Decision Act, the descriptive net sentiment change index was calculated for each topic (Figure 4). The magnitude and direction of sentiment change varied across topics. The largest positive shifts were observed in end-of-life decision-making and hospice and palliative care, both of which showed increases in positive sentiment combined with reductions in negative sentiment. More moderate positive shifts were observed in euthanasia debate and withdrawal of treatment, whereas aging and social care showed minimal change in sentiment, indicating limited shifts in the overall emotional tone of these discussions. In contrast, death with dignity was the only topic that exhibited a negative shift in the index, which appears to be primarily attributable to an increase in negative sentiment, as illustrated in Figure S1 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 4. Descriptive net sentiment change index by topic. Positive values indicate an increase in positive sentiment or a decrease in negative sentiment following the act.



Discussion

Principal Findings

This longitudinal social media analysis found substantial changes in the volume, thematic composition, and aggregate sentiment of online discourse on life-sustaining treatment in South Korea between 2014 and 2025. These changes were most apparent around and after the implementation period of the Life-Sustaining Treatment Decision Act, but they should be interpreted as temporal associations rather than causal effects of the legislation. Prior to 2018, discussions were largely centered on clinical aspects of care, particularly end-of-life decision-making and withdrawal of treatment. Around the time of implementation of the Life-Sustaining Treatment Decision Act, ethical topics such as death with dignity and euthanasia debate became more prominent. In subsequent years, discourse increasingly reflected care-oriented and broader social perspectives, including hospice and palliative care and aging and social care. These findings suggest that the implementation of the Life-Sustaining Treatment Decision Act was associated not only with changes in the scope of medical decision-making but also with broader shifts in the structure of public discourse.

Previous studies reporting an increase in the actual implementation of life-sustaining treatment withdrawal in clinical practice following the act provide empirical context for the observed increase in discourse volume in this study [6]. In addition, this study found that sentiment toward end-of-life decision-making and hospice and palliative care became more positive over time. This pattern may reflect increasing social acceptance as these domains became more clearly defined within a legal and institutional framework. Park et al [8] suggested that the concept of well-dying in South Korea extends beyond individual preparation for death and is closely linked to family burden and social care. Consistent with this, the increasing prominence of "family" as a high-weight keyword in both the withdrawal of treatment and hospice and palliative care topics indicates that life-sustaining treatment is increasingly framed not only as a medical decision but also as a broader social care issue. However, the minimal change in sentiment observed for aging and social care suggests that these issues continue to be perceived as structural burdens and that policy implementation alone may be insufficient to substantially shift public perceptions in this domain.

The results also showed an increase in discourse related to death with dignity and euthanasia debate following policy implementation, suggesting that the introduction of the legal framework contributed to expanding these topics into the public sphere. The observed reduction in negative sentiment toward euthanasia debate is consistent with prior survey-based findings indicating relatively high public support for euthanasia legalization in South Korea [14]. In contrast, death with dignity showed an increase in negative sentiment. This may be explained by the conceptual ambiguity of the term "death with dignity," which is not a strictly defined legal concept and is often used in multiple ways in public discourse. As a result, it may be interpreted differently from the legally defined concept of life-sustaining treatment withdrawal, leading to increased debate and heightened emotional tension.

The trajectory of discourse on life-sustaining treatment in South Korea appears to differ from that observed in many Western countries. In Western contexts, discussions on end-of-life care have developed gradually over time, driven by bioethics, patient rights, and the expansion of palliative care, and may be described as a discourse-driven model [3,26,27]. In contrast, the Korean case may be better understood as a policy-associated discourse shift in which legislative implementation and institutional change coincided with increased public engagement with a previously marginalized topic [28]. This framing

XSL•FO
RenderX

emphasizes the temporal and contextual relationship between policy implementation and online discourse while avoiding causal interpretation.

Differences are also evident in the structure of decision-making reflected in the discourse. Whereas end-of-life decision-making in Western contexts is often centered on individual autonomy, the discourse observed in this study reflects a stronger emphasis on family involvement [29]. Previous research has shown that a substantial proportion of life-sustaining treatment decisions involve family members in South Korea, and in this study, family-related keywords were frequently identified in both the withdrawal of treatment and hospice and palliative care topics [6]. This pattern reflects the influence of family-centered cultural norms in Asian contexts, where major life decisions, including those at the end of life, are often made collectively [30].

This study has several strengths. First, it analyzed a large-scale dataset of social media content spanning more than a decade (2014 - 2025), allowing for a comprehensive examination of longitudinal changes in public discourse on life-sustaining treatment. Second, by integrating topic modeling, temporal analysis, and sentiment analysis, this study was able to capture both the thematic structure and emotional orientation of discourse, providing a multidimensional understanding of how public discussions evolved over time. Third, the use of multiple platforms, including Naver blogs, Naver cafes, and YouTube comments, enabled the analysis to reflect diverse forms of user-generated content, ranging from personal narratives and community discussions to public reactions to policy issues. Fourth, this study offers a conceptual contribution by characterizing the Korean experience as a policy-triggered discourse model, in contrast to the discourse-driven trajectories observed in Western contexts. This distinction provides a framework that may be applicable to other Asian societies that have undergone similar legislative reforms in end-of-life care—including Taiwan's Patient Right to Autonomy Act and Japan's ongoing policy deliberations—and invites comparative inquiry into whether legislative enactment similarly catalyzes public discourse in these contexts [31,32].

However, several limitations should be considered when interpreting the findings. First, the data were derived from selected online platforms and may not be representative of the general population as social media users differ in demographic and behavioral characteristics. Platform-specific differences should also be considered. Blogs, cafes, and YouTube comments differ in text length, user motivation, community structure, and degree of reactivity to media content. Therefore, the combined corpus should not be interpreted as a representative measure of public opinion in the general Korean population. Rather, it reflects discourse patterns within selected online environments where end-of-life issues were publicly discussed. Second, the volume of data was substantially higher in the postimplementation period than in the preimplementation period, which may have influenced the stability of topic estimation and sentiment comparisons despite the use of normalized proportions in temporal analyses. Third, although sensitivity analyses suggested that the major thematic structure of the 6-topic LDA solution was broadly retained across alternative random seeds, some keyword overlap was observed among conceptually adjacent themes. Topic labels should therefore be interpreted as dominant thematic summaries rather than mutually exclusive conceptual categories. Fourth, sentiment analysis was conducted using a general-purpose lexicon-based approach, which may not fully capture the contextual nuances of Korean end-of-life discourse. End-of-life discussions often contain emotionally ambiguous and context-dependent expressions. For example, narratives about withdrawal of treatment, death with dignity, or hospice care may include negative lexical items while conveying relief, acceptance, respect for patient autonomy, or support for dignified dying. Conversely, apparently neutral statements may reflect moral concern, anxiety, or distrust. Therefore, the sentiment estimates in this study should be interpreted as aggregate lexicon-based indicators of emotional tone rather than definitive classifications of subjective emotion. Future research should validate sentiment classification using manually annotated Korean end-of-life discourse and develop domain-adapted or context-aware sentiment models, including dimensional or aspect-based approaches such as multidimensional relation models and recent shared-task initiatives in dimensional aspect-based sentiment analysis [33,34]. Fifth, this study was observational and descriptive in nature; therefore, the findings cannot establish a causal effect of the Life-Sustaining Treatment Decision Act on public discourse. Although thematic and sentiment changes were observed around the implementation period of the act, other social and historical factors may also have contributed to these trends. These include population aging, increasing public awareness of hospice and palliative care, media coverage of end-of-life issues, changes in platform use, COVID-19–related experiences with serious illness and death, and broader public discussion of advance directives. Accordingly, the results should be interpreted as policy-associated temporal shifts in online discourse rather than evidence of a direct policy effect.

Conclusions

Public discourse on life-sustaining treatment in South Korea underwent substantial structural changes following the implementation of the Life-Sustaining Treatment Decision Act in 2018, shifting from a clinical focus to ethical debates and, subsequently, to care-oriented and social perspectives. Sentiment toward end-of-life decision-making and hospice and palliative care became more positive over time, whereas aging and social care showed minimal change in sentiment, indicating that structural challenges in this domain remain insufficiently addressed by policy alone. These findings should be interpreted as policy-associated temporal changes rather than causal effects of the act. They highlight the importance of integrating social and familial dimensions into public health communication and advance care planning and underscore the value of social media data in monitoring policy-associated changes in public discourse.

XSL•FO
RenderX

Funding

This research was supported by a Korea Basic Science Institute (National Research Facilities and Equipment Center) grant funded by the Korean Ministry of Education (RS-2026-25533557), by the ANCHOR program through the Jeollanamdo ANCHOR center, funded by the Korean Ministry of Education (2026-ANCHOR-14-003), by a National Research Foundation of Korea grant funded by the Korean Ministry of Science and ICT (RS-2024-00345981 and RS-2026-25591616), and by the Korean Ministry of Food and Drug Safety (21153MFDS601).

Data Availability

The data used in this study were collected from publicly accessible online platforms (Naver blogs, Naver cafes, and YouTube) and are subject to platform policies and terms of service. Therefore, the raw data cannot be shared. The code used for data processing and analysis is available from the corresponding author on reasonable request.

Authors' Contributions

MK conceived and designed the study, collected and analyzed the data, interpreted the results, and drafted the manuscript. HSS reviewed the study design and analytical methods and critically reviewed the manuscript. Both authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials comprising data collection keywords, platform-specific data collection procedure, topic number selection using c_v coherence scores, latent Dirichlet allocation random-seed sensitivity analysis, Korean–English translation of topic keywords, and change in positive and negative sentiment proportions after the act, by topic.
[DOCX File, 93 KB - infodemiology_v6i1e95918_app1.docx]

References

1. Bone AE, Gomes B, Etkind SN, et al. What is the impact of population ageing on the future provision of end-of-life care? Population-based projections of place of death. Palliat Med 2018 Feb;32(2):329-336. [doi: 10.1177/0269216317734435] [Medline: 29017018]
2. Sleeman KE, de Brito M, Etkind S, et al. The escalating global burden of serious health-related suffering: projections to 2060 by world regions, age groups, and health conditions. Lancet Glob Health 2019 Jul;7(7):e883-e892. [doi: 10.1016/S2214-109X(19)30172-X] [Medline: 31129125]
3. Borgstrom E, Walter T. Choice and compassion at the end of life: a critical analysis of recent English policy discourse. Soc Sci Med 2015 Jul;136-137:99-105. [doi: 10.1016/j.socscimed.2015.05.013] [Medline: 25989003]
4. Heo DS, Yoo SH, Keam B, Yoo SH, Koh Y. Problems related to the act on decisions on life-sustaining treatment and directions for improvement. J Hosp Palliat Care 2022 Mar 1;25(1):1-11. [doi: 10.14475/jhpc.2022.25.1.1] [Medline: 37674892]
5. Kim HJ, Kim YJ, Kwon JH, et al. Current status and cardinal features of patient autonomy after enactment of the life-sustaining treatment decisions act in Korea. Cancer Res Treat 2021 Oct;53(4):917-925. [doi: 10.4143/crt.2021.324] [Medline: 34082494]
6. Kim JS, Yoo SH, Choi W, et al. Implication of the life-sustaining treatment decisions act on end-of-life care for Korean terminal patients. Cancer Res Treat 2020 Jul;52(3):917-924. [doi: 10.4143/crt.2019.740] [Medline: 32204581]
7. Lee YJ, Ahn S, Cho JY, et al. Change in perception of the quality of death in the intensive care unit by healthcare workers associated with the implementation of the “well-dying law”. Intensive Care Med 2022 Mar;48(3):281-289. [doi: 10.1007/s00134-021-06597-7] [Medline: 34973069]
8. Park SY, Lee B, Seon JY, Oh IH. A national study of life-sustaining treatments in South Korea: what factors affect decision-making? Cancer Res Treat 2021 Apr;53(2):593-600. [doi: 10.4143/crt.2020.803] [Medline: 33227190]
9. Wang Y, Koffman J, Gao W, Zhou Y, Chukwusa E, Curcin V. Social media for palliative and end-of-life care research: a systematic review. BMJ Support Palliat Care 2024 May 17;14(2):149-162. [doi: 10.1136/spcare-2023-004579] [Medline: 38594059]
10. Wang Y, Chukwusa E, Koffman J, Curcin V. Public opinions about palliative and end-of-life care during the COVID-19 pandemic: Twitter-based content analysis. JMIR Form Res 2023 Aug 7;7:e44774. [doi: 10.2196/44774] [Medline: 37368840]
11. Jang H, Rempel E, Roth D, Carenini G, Janjua NZ. Tracking COVID-19 discourse on Twitter in North America: infodemiology study using topic modeling and aspect-based sentiment analysis. J Med Internet Res 2021 Feb 10;23(2):e25431. [doi: 10.2196/25431] [Medline: 33497352]

XSL•FO
RenderX

12. Kim A, Woo K. Public perspectives on palliative and hospice care: social media content analysis using topic modeling and multiclass sentiment analysis. J Med Internet Res 2025 Sep 16;27:e70836. [doi: 10.2196/70836] [Medline: 40957008]
13. Hong M, Hong S, Adamek ME, Kim MH. Death attitudes among middle-aged Koreans: role of end-of-life care planning and death experiences. Int J Aging Hum Dev 2018 Jan;86(1):51-68. [doi: 10.1177/0091415016689473] [Medline: 28105867]
14. Yun YH, Sim JA, Choi Y, Yoon H. Attitudes toward the legalization of euthanasia or physician-assisted suicide in South Korea: a cross-sectional survey. Int J Environ Res Public Health 2022 Apr 24;19(9):5183. [doi: 10.3390/ijerph19095183] [Medline: 35564575]
15. Fu J, Li C, Zhou C, et al. Methods for analyzing the contents of social media for health care: scoping review. J Med Internet Res 2023 Jun 26;25:e43349. [doi: 10.2196/43349] [Medline: 37358900]
16. Lee MC. Kiwi: developing a Korean morphological analyzer based on statistical language models and skip-bigram [Article in Korean]. Korean J Digit Humanit 2024;1(1):109-136. [doi: 10.23287/KJDH.2024.1.1.6]
17. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022. [doi: 10.5555/944919.944937]
18. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. In: WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining: Association for Computing Machinery; 2015. [doi: 10.1145/2684822.2685324]
19. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Comput Linguist 2011 Jun;37(2):267-307. [doi: 10.1162/COLI_a_00049]
20. Nielsen F. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In: Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big Things Come in Small Packages: CEUR-WS; 2011:93-98.
21. Bradley MM, Lang PJ. Affective Norms for English Words (ANEW): instruction manual and affective ratings. Technical report c-1. : The Center for Research in Psychophysiology, University of Florida; 1999.
22. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav Res Methods 2013 Dec;45(4):1191-1207. [doi: 10.3758/s13428-012-0314-x] [Medline: 23404613]
23. Mohammad S. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Association for Computational Linguistics; 2018:174-184. [doi: 10.18653/v1/P18-1017]
24. Lee LH, Li JH, Yu LC. Chinese EmoBank: building valence-arousal resources for dimensional sentiment analysis. Trans Asian Low Resour Lang Inf Process 2022;21(4):1-18. [doi: 10.1145/3489141]
25. Park SM, Na CW, Choi MS, Lee DH, On BW. KNU Korean sentiment lexicon: Bi-LSTM-based method for building a Korean sentiment lexicon [Article in Korean]. J Intell Inf Syst 2018;24(4):219-240. [doi: 10.13088/jiis.2018.24.4.219]
26. Wolf SM, Berlinger N, Jennings B. Forty years of work on end-of-life care--from patients' rights to systemic reform. N Engl J Med 2015 Feb 12;372(7):678-682. [doi: 10.1056/NEJMms1410321] [Medline: 25671263]
27. Singer PA, Bowman KW. Quality end-of-life care: a global perspective. BMC Palliat Care 2002 Jul 25;1(1):4. [doi: 10.1186/1472-684x-1-4] [Medline: 12139768]
28. Kingdon JW. Agendas, Alternatives, and Public Policies, 2nd edition: HarperCollins College Publishers; 1995.
29. Ruhnke GW, Wilson SR, Akamatsu T, et al. Ethical decision making and patient autonomy: a comparison of physicians and patients in Japan and the United States. Chest 2000 Oct;118(4):1172-1182. [doi: 10.1378/chest.118.4.1172] [Medline: 11035693]
30. Mori M, Morita T. End-of-life decision-making in Asia: a need for in-depth cultural consideration. Palliat Med 2020 Feb;34(2):NP4-NP5. [doi: 10.1177/0269216319896932] [Medline: 31959084]
31. Tanaka M, Kodama S, Lee I, Huxtable R, Chung Y. Forgoing life-sustaining treatment - a comparative analysis of regulations in Japan, Korea, Taiwan, and England. BMC Med Ethics 2020 Oct 16;21(1):99. [doi: 10.1186/s12910-020-00535-w] [Medline: 33066771]
32. Tu Y, Young Y, O'Connor M. Implications for end-of-life care: comparative analysis of advance directives laws in Taiwan and the United States. Am J Hosp Palliat Care 2026 Mar;43(3):310-322. [doi: 10.1177/10499091251328007] [Medline: 40127462]
33. Xie H, Lin W, Lin S, Wang J, Yu LC. A multi-dimensional relation model for dimensional sentiment analysis. Inf Sci 2021 Nov;579:832-844. [doi: 10.1016/j.ins.2021.08.052]
34. Yu LC, Becker J, Muhammad SH, et al. SemEval-2026 task 3: dimensional aspect-based sentiment analysis (DimABSA). In: Proceedings of the 20th International Workshop on Semantic Evaluation: Association for Computational Linguistics; 2026:3753-3778. [doi: 10.18653/v1/2026.semeval-1.452]

Abbreviations

LDA: latent Dirichlet allocation

XSL•FO
RenderX

Edited by I Brooks; submitted 23.Mar.2026; peer-reviewed by AA Mir, LC Yu; revised version received 13.Jul.2026; accepted 24.Aug.2026; published 23.Sep.2026.

Please cite as:
Kim M, Suh HS
Shifting Public Discourse on Life-Sustaining Treatment in South Korea From 2014 to 2025: Longitudinal Social Media Analysis
JMIR Infodemiology 2026;6:e95918
URL: https://infodemiology.jmir.org/2026/1/e95918
doi:10.2196/95918

© Miryoung Kim, Hae Sun Suh. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 23.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

How Social Media Analysis Offers an Opportunity to Understand the Reality of People Living With Multiple Sclerosis: Descriptive French Study

Emmanuelle Leray[1], PhD; Stéphane Schück[2], MD; Pamela Voillot[2], MSc, MPH; Nathalie Texier[2], PharmD

[1]Univ Rennes, EHESP, CNRS, Inserm, ARENES UMR 6051, RSMS U 1309, EHESP, Avenue du Pr Léon Bernard, Rennes, France

[2]Kapcode, Paris, France

Corresponding Author:
Emmanuelle Leray, PhD
Univ Rennes, EHESP, CNRS, Inserm, ARENES UMR 6051, RSMS U 1309, EHESP, Avenue du Pr Léon Bernard, Rennes, France

Abstract

Background: Multiple sclerosis (MS) is a chronic neurological disease that starts in young adulthood and can significantly affect quality of life (QoL) due to various symptoms, and the risk of disability. MS directly affects people living with the disease and indirectly affects their relatives and family caregivers.

Objective: The objective of this social media analysis was to identify the main topics of discussion among people affected by MS and their perceptions of the impact of MS on their QoL.

Methods: Publicly available French messages, posted between January 2017 and October 2022, were retrieved using an extraction query that contained keywords related to MS. The effects on QoL were detected using a machine learning algorithm specifically trained on social media data. Five specific models covered the following health-related QoL dimensions: physical well-being, psychological well-being, daily activities (including professional and academic activities), social or relational well-being, and material well-being. Descriptive statistics were provided and illustrated with quotes from social media.

Results: The analysis corpus for the 2017 to 2022 period included 3225 messages corresponding to 2034 different social media users, either people living with MS (654/3225, 20%) messages or family caregivers (2571/3225, 80%) messages, identified from 32 sources. Women represented 42.5% (864/2034) and men represented 28.2% (574/2034) of social media users (gender was unknown for 596/2034, 29.3%), and their mean age was 35 (SD 6.6) years. The 2 main themes of posts were “Caregivers and family members” (1032/3225, 32%) and “Disability” (774/3225, 24%). Overall, 847 messages described at least one impact of MS on QoL: relational or social (n=431, 50.9%), physical (n=284, 33.5%), psychological (n=76, 9.0%), financial or material well-being (n=32, 3.8%), and daily activities (n=24, 2.8%).

Conclusions: Our findings confirm the high impact of MS on everyday life and QoL for both patients and family caregivers. Caregivers were the most numerous to express themselves and post messages on social media. The most affected QoL dimension was relational or social well-being, which is probably linked to the fact that social networks and digital patient communities are places for discussion, sharing experiences, and looking for support. These findings confirm that social media is a way for people affected by MS to express themselves and look for support and understanding. They also show that social media provides an opportunity to discover the fears, questions, needs, and thoughts of those affected by the disease, particularly caregivers, who are rarely considered in research studies.

(JMIR Infodemiology 2026;6:e88763) doi:10.2196/88763

KEYWORDS

multiple sclerosis; social media; quality of life; burden; testimonies

Introduction

Multiple sclerosis (MS) is a chronic neurological disease that starts in young adulthood (mean age at disease onset: 30 y), with a female-to-male ratio of 2.5 [1]. According to the most recent estimate (2020), 2.8 million people live with MS worldwide [2].

MS symptoms are variable and unpredictable, and in each patient, they can change or fluctuate over time. The most common symptoms are fatigue, walking difficulties, and vision and bladder problems [1]. MS is characterized by physical disability, cognitive impairment, and other symptoms that affect the quality of life (QoL). In addition, as the disease usually begins at a time when people are building their personal lives

XSL•FO
RenderX

and careers, it can have a significant impact on their lives and QoL.

Social media has become very popular since the 2000s. Indeed, in 2017, it was estimated that there were over 2.3 billion active social media users worldwide, and this number was growing by approximately 1 million new users every day [3]. In 2025, it was estimated that there would be 5.4 billion social media users worldwide, corresponding to two-thirds of the total global population [4].

The term "social media" refers to internet-based tools that facilitate the gathering of individuals and communities to communicate and share information, ideas, and experiences in real time [3]. Social media have broadly affected medicine, perhaps most publicly by enabling increased communication with and among patients [5]. They allow more rapid, spontaneous, and wider communication and data sharing directly to, from, and between patients. People can anonymously express their feelings and experiences.

To our knowledge, despite the opportunity to access direct, free, and spontaneous testimonies from people living with and/or affected by a disease, few studies in the field of MS have been based on social media analysis. Indeed, we found several analyses of discussions and/or experiences related to specific topics, such as cannabinoid medical use [6], modafinil (a stimulant used in cases of sleep disorders) [7], and pain [8], that included but were not limited to people with MS. Therefore, the aim of the present study was to identify the main themes of discussion by people affected by MS (patients and family/caregivers) on social media and their perceptions of the impact of the disease on their QoL.

Methods

Study Design and Population

This noninterventional, retrospective, real-world study included data gathered from social media posts (eg, X, YouTube, Doctissimo, web forums, and blogs, such as la-sclerose-en-plaques.fr and aufeminin.com), written in French by people affected by MS. The term "people affected by MS" refers to people living with the disease, as well as any individual whose life is affected by the disease, such as family members and caregivers. Publicly available messages posted between January 2017 and October 2022 were considered. Private-account posts or data-restricted posts were not accessible and, therefore, were not included.

Data Extraction

For the present study, an extraction query was constructed (available in Multimedia Appendix 1) with keywords related to MS. All publicly available posts on social media or the web that contained one of the relevant keywords (identified before the extraction) and were posted during the study period were targeted and extracted using the Brandwatch extractor [9]. This tool is based on logical queries using specific keywords evocative of the topic of interest combined with Boolean operators to target specific verbatims. Using the query, the Brandwatch extractor searched through the available sources and posts to find words that matched those in the query. Then, the identified posts were downloaded with metadata (eg, date and web user pseudonym). This constituted the study dataset.

The cleaning process started by removing posts from irrelevant sources, such as potential advertising sites, or messages not fully written in French. The resulting dataset underwent further cleaning to exclusively obtain testimonies by people affected by MS. This was performed using an extreme gradient boosting classifier, a machine learning algorithm that identified posts related to patients' and caregivers' experiences with MS. This algorithm was previously trained on a social media dataset constructed using various pathologies and data sources [10,11]. The algorithm used three principal features (syntax, lexical field, and semantic features) to perform predictions. Its recorded performance was as follows: sensitivity of 78% (ie, the proportion of identified true positives) and positive predictive value of 69% (ie, the proportion of true positives among the detected positives). Only posts predicted to have been written by patients or family caregivers were kept, and the others were discarded. Then, a manual review of the predicted posts or authors was performed to exclude false positives. This review was conducted by an expert in semantic analysis of social networks.

The resulting posts constituted the analysis dataset, and the author type (ie, patient or caregiver) was added as a label to each post.

Age and gender were retrieved from information contained in the messages when disclosed by the user (eg, "I'm a 27 yo woman"); otherwise, they were considered to be missing.

In order to highlight the main themes of discussion, the analysis was based on a natural language processing technique (Biterm Topic Models for short text) to automatically identify the topics of discussion addressed in a set of messages. It provides an initial overview of the main topics of discussion. All messages were classified into a single, most representative category based on their content relative to the other messages. A manual interpretation (2 persons) was carried out based on the terms most representative of the category to provide a meaningful title for the category. There is no single metric that can guarantee the best readability or the best clustering of messages. Consequently, several models were tested using different numbers of topics (from 8 to 15), and the most relevant model was then selected based on the maximum value of the log-likelihood, a statistical indicator measuring the extent to which the model correctly explains the observed data. Then, a qualitative manual interpretation phase was carried out to merge certain topics exhibiting a high degree of semantic similarity and to group topics of relatively low importance into an "Other" category.

Impact of Multiple Sclerosis on Health-Related Quality of Life

The effects of MS on QoL were assessed using a machine learning algorithm specifically trained on social media data [12]. Its operating mode allowed the detection of QoL alterations in verbatims. Then, the effects were classified using 5 specific machine learning models that corresponded to 5 health-related QoL dimensions: physical, psychological, daily activities

XSL•FO
RenderX

(including professional and academic activities), social, and material or financial [12]. Then, the output of the models was validated by a health care professional (medical doctor with specific training in data science) to ensure the correctness of the results. This algorithm was previously used in other contexts (atopic dermatitis [13] and oncology treatment [11]).

Ethical Considerations

By design, this study only included open data publicly available on the web or social media. Private groups, closed online forums, profile information, and personal conversations were not accessed or included. Because of the social platforms' terms and conditions of use, users give consent for data reuse purposes. This study reused publicly available secondary data without any link to the actual authors of the messages. Furthermore, gathering consent from all users for the retrospective reuse of their data was not technically feasible given the large number of authors.

Data Analysis

The whole corpus was analyzed using descriptive statistics, and each post corresponded to a statistical unit. The following characteristics were considered: number of posts (overall and by month over the study period), number of authors, discussion themes, occurrence of QoL-related posts and QoL dimensions, data source, and users' characteristics (gender and age). Examples of posts (verbatim translated from French) are provided to illustrate the findings.

Results

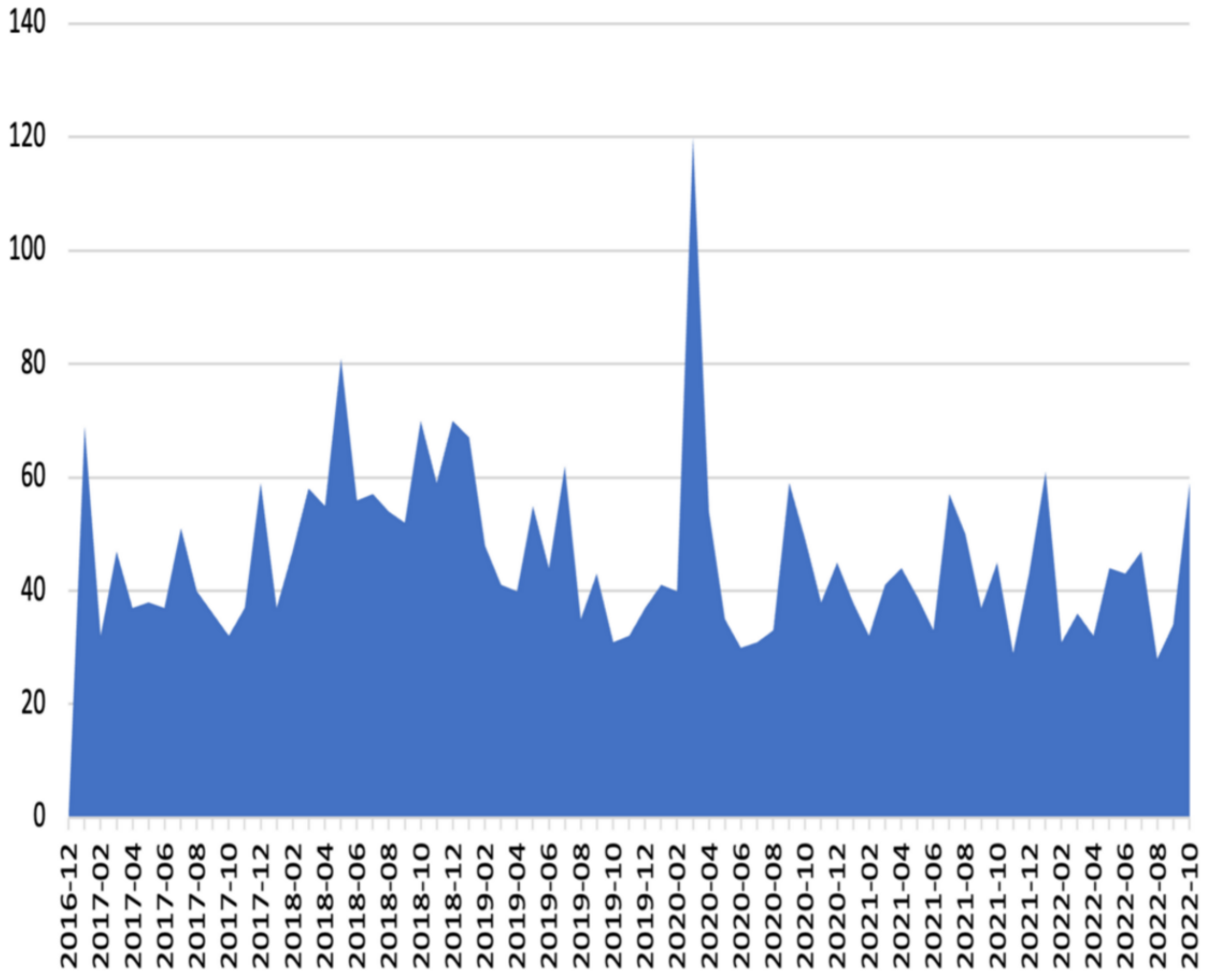
Study Population

Over the period 2017 to 2022, the analysis corpus included a total of 3225 messages written in French by 2034 different web or social media users: either people with MS (n=654, 20% messages) or family caregivers (n=2571, 80% messages). They came from 32 different sources.

The respondents' gender could not be determined for 29.3% (596/2034) of users and was categorized as woman for 42.5% (864/2034) and man for 28.2% (574/2034). Information on age could not be found for 6% (123/2034) of users. The mean age was 35 (SD 6.6) years, and users were predominantly aged 30 to 40 years (57%, 1840/3225 posts).

Figure 1 shows the monthly number of posts over time. There was no specific trend except some peaks from time to time, notably during World MS Day (May 30 each year) and during the COVID-19 lockdown period that started in March 2020.

Figure 1. Number of multiple sclerosis–related public posts during the 2017 - 2022 period (by month).



XSL•FO
RenderX

Themes of Discussion

The main themes of the posts are listed in Table 1, overall and according to the patient's or caregiver's perspective. The model initially identified an optimal number of 12 themes, determined using the log-likelihood criterion. Subsequently, the manual interpretation phase enabled certain topics with a high degree of semantic similarity to be merged and topics of relatively low importance to be grouped together under an "Other" category, resulting in a total of 7 categories. All the topics related to the disease were covered. People with MS were more likely to talk about their symptoms and the nature of their illness, while caregivers were more likely to talk about themselves and their family members. The 2 most frequent topics are explored below.

Table . Main themes of discussion identified in the 3225 posts written in French by people affected by multiple sclerosis during the 2017 - 2022 period.

Theme	Posts, n (%)	Written by	
		People with MS[a], %	Caregivers, %
Caregivers and family members	1032 (32)	20	80
Disability	774 (24)	44	56
Symptoms, pain, and fatigue	419 (13)	67	33
Treatments and relapses	323 (10)	69	31
MS impact and difficulties	193 (6)	75	25
Diagnosis, care, and MRI[b]	129 (4)	86	14
Other	355 (11)	—[c]	—

[a]MS: multiple sclerosis.

[b]MRI: magnetic resonance imaging.

[c]Not applicable.

In theme 1, "Caregivers and family members," which represented 32% (1032/3225) of the corpus, many messages were short and not really informative. Many (2571/3225, 80%) came from family caregivers, and the others (654/3225, 20%) came from people with MS. For instance, children talked about the diagnosis, symptom or disability onset or progression in their mother or father living with MS, or how everyday life was affected by the disease. The way such messages were written suggests the need to share, talk about their experiences, and discuss with people who are going through the same thing.

Examples of posts are as follows:

> *At 24, a few months after the birth of my daughter, I found myself with vision loss in one eye caused by a relapse, a diagnosis of #multiple sclerosis and a beautiful baby less than a year old.*
>
> *I received the diagnosis of MS 10 years ago and for the past 6 months I have been feeling really bad (since my second delivery). Headaches, dizziness, neck pain, leg pain and fatigue. I can't tolerate any treatment, and my morale is starting to take a serious hit!!! The doctors aren't too concerned, but I am. I really need some support!!*
>
> *I truly understood that my father was ill the day my mother came to pick me up from school and told me that dad could no longer get out of bed because a relapse of #multiple sclerosis had caused him to lose the use of his legs.*
>
> *My mother has multiple sclerosis! It's been three years now, and it's been difficult at first, but you have to be strong! Stay strong*
>
> *But my mum has this incredible mental strength. She has multiple sclerosis, there's no cure, and the only treatment available makes you depressed, but she's there, dealing with my dad's moods, her job, my brother...*

Others shared a specific event in their medical history or care pathway to provide feedback to other people and to animate the discussion within the patient community. Few messages evoked parenthood when living with MS or the risk of MS in the family, as exemplified by the following post:

> *I spoke with my mother earlier about the fact that the doctors want to check that I don't have multiple sclerosis (we hadn't talked about it together yet) and it hurts me because she doesn't say anything but she's really worried and she's not doing well at all, it makes me sad.*

Lastly, a high proportion of messages concerned the risk of COVID-19 and/or vaccination for family members with MS.

Examples of posts are as follows:

> *I have multiple sclerosis with immunosuppressive treatment, and I'm one of the most vulnerable people in relation to COVID-19. The start of the new school year on September 1 has me terrified...*
>
> *My mother has multiple sclerosis and is on immunosuppressive treatment. If she catches the virus, well, too bad. I'm starting to get angry that nothing is being done, when it's clear that it's spreading.*

Regarding theme 2, "Disability related to MS," which represented 24% (774/3225) of the posts, many discussions were around "invisible disability."

XSL•FO
RenderX

Patients and caregivers posted messages saying that the disability associated with MS is not well known or recognized by the general population and also by the administration or social security.

Examples of posts are as follows:

> *Thank you very much for this video and for talking about invisible disabilities. My sister has multiple sclerosis and received her diagnosis four months ago, but her illness and above all, the prejudices of her boss, who refuses to make accommodation for her, have forced her to quit. Her boss does not understand the irregular, unpredictable, yet exhausting nature of this condition on a daily basis. I sincerely hope that your video will encourage people who get in contact with individuals with disabilities to reconsider their preconceptions and give them a chance to be more than just a disease.*

> *No real handicap! A bit like the big idiots who park in disabled spaces and who don't think it's a big deal to take my father's space because he walks (with a limp). Except that he has multiple sclerosis and it's not written on his face!*

Some messages reported that disability was the consequence of symptoms in a very personal and qualitative way. These testimonies indicated the important part of pain and fatigue in the disease course. Likewise, many social media users discussed the potential impacts of such symptoms and of disability (eg, activities and work).

Examples of posts are as follows:

> *I have an appointment with my neurologist on June 6... I have tingling sensations on my face, hands and legs! It feels like electric shocks in my legs, cramps... I feel like I'm walking drunk, I lose my balance... I won't even mention the fatigue and pain throughout my body.*

> *I have a lot of pain with my multiple sclerosis. I'm in category 1 [Disability pension], but I can't do my job anymore. Continuous pain and my tiredness mean that I have to force myself to work. I have a lot of pain with my multiple sclerosis, and I have relapses. I'd like to move up to category 2 and be able to stop working, but I don't know if I need to make an appointment with the social security doctors. And will they understand what I put up with every day? I'm afraid they won't understand me.*

Some posts were short, and the disability issue could be identified because the word "wheelchair" was used in association with MS.

Impact of Multiple Sclerosis on Quality of Life

Overall, 847 verbatims (ie, 26% of the total corpus) contained at least one impact of MS on QoL (Figure 2). There was no direct correlation between the thematic analysis and QoL analysis, but the results were consistent and logical (main themes: caregivers and disability; main impacts on QoL: relational and physical).

The relational or social impact was the most frequently mentioned (50.9%, 431/847), especially by caregivers. It included effects on social and family life, such as family proximity, social isolation, and sexual relationships. The family aspect was recurrent in the testimonies (73% of posts). For instance, many posts said that the diagnosis of MS extends to the whole family. Some messages talked about unconditional help and support from family members, but also about cases of suffering, neglect, and divorce linked to the disease.

Examples of posts are as follows:

> *My sister was great, she helped me to walk again and to get around without rushing me. She made me think about other things. She mentally supported me. I don't know how to thank her. #carer #multiple sclerosis.*

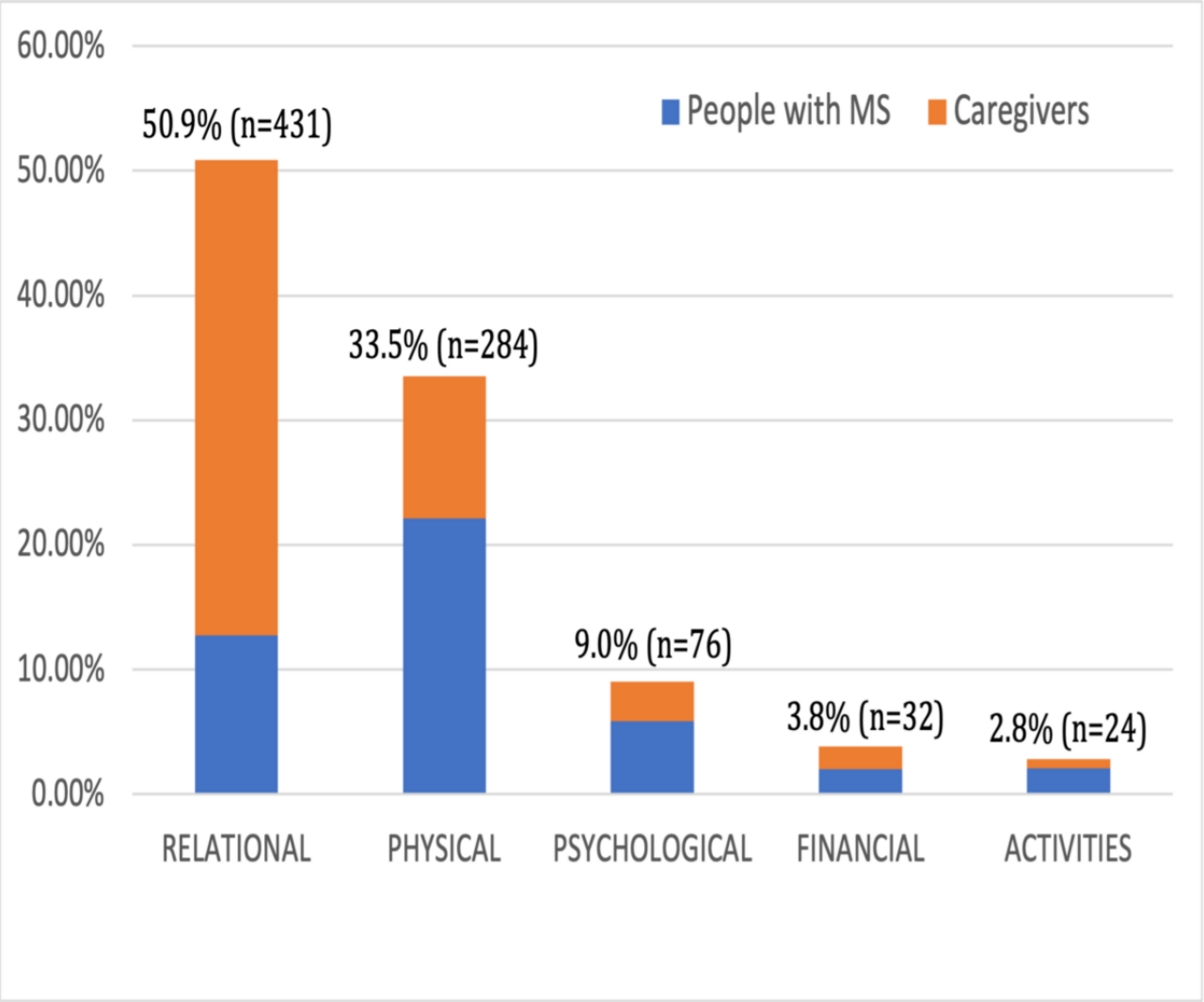
> *I look after my friend all the time and I'm always on the alert, it's exhausting.*

> *I have multiple sclerosis, and I have little doubt about what I will experience in the future: solitude, loneliness, isolation, patience, willingness and so much more.*

> *I've known from the start that my wife has multiple sclerosis #multiple sclerosis, but that has never stopped me from looking to the future with her, including building a family.*

XSL•FO
RenderX

Figure 2. Percentage of public posts concerning the different health-related quality-of-life dimensions (N=847), overall and according to the writers' status (patient with MS or family caregiver). MS: multiple sclerosis.



Patients (4%) who no longer wanted to be a burden on people around them also mentioned isolation.

Examples of posts are as follows:

> *My first reaction after the diagnosis #multiple sclerosis (6 years ago) was to divorce to avoid this ordeal for my husband... It was stupid... The two of us are stronger and my husband supports me in an exceptional manner.*
>
> *Simple example: my cousin has multiple sclerosis; she risks to become paralyzed later on, so she doesn't want to have a child if she's going to be a burden because of her future disability.*

The physical impact was the second QoL-related topic (284/847, 33.5%) and was most frequently reported by people with MS. It was defined as the impact on the physical condition, such as physical fatigue, motor limitations, pain, and weight gain or loss. In 48% of posts, the authors described various types of pain in their daily lives, including migraine and diffuse pain at the time of relapses.

Examples of posts are as follows:

> *I have an appointment with my neurologist on June 6... I have tingling on my face, hands and legs! A sort of electric shock in my legs, cramps... The impression of walking as if I were drunk, loss of balance... I'm not even talking about the fatigue, or the pain all over my body.*
>
> *My father was "placed" in a long-term care facility at the age of 38; his multiple sclerosis had become too much for his wife and mother of his two young children to handle.*

Disability and its progression were mentioned in several testimonies (22% of posts), evoking different types of plegia and/or loss of sensation (visual symptoms, for instance).

An example of a post is as follows:

> *I was stubborn, I didn't want to use crutches or walking sticks. As a result, I fell a lot.*

Lastly, 2% of posts concerned the severe fatigue experienced by people living with MS, and its impact on their everyday life.

The psychological impact was the third QoL-related topic (76/847, 9.0%) and was also frequently reported by people with MS. It referred to the effects on the psychological condition,

XSL•FO
RenderX

such as moral fatigue, self-image and self-esteem, loneliness, depression, and hope. In 25% of the posts, patients and their family caregivers referred to MS by using the expression "daily struggle." They underlined the difficulties related to the disease and its symptoms, but also the difficulties related to the management and care of this chronic lifelong disease.

Examples of posts are as follows:

> *Because when I received the diagnosis of MS at the age of 40, I had the impression that everything was falling apart, even though days later I was starting the fight.*
>
> *No, I'm not being treated at a university hospital, but at a regular hospital. I didn't have a hard time accepting my MS, but having 6 relapses in 9 months is just very hard on the body and the mind. At the beginning of my MS, I had two relapses. And before my pregnancy, I had none, although I've had MS for 13 years and have taken all the treatments.*
>
> *I have multiple sclerosis, and I have a pretty good idea of what I need to learn in this life: loneliness, isolation, patience, willpower... and many other things besides.*
>
> *I watch over my friend all the time and am always on the alert, it's exhausting. #caregiver #multiple sclerosis*

Panic, anxiety, and terror toward the future, disease progression, the occurrence of new relapses, or health events could be identified in 15% of the posts:

> *Why are you crying, Mum? Each wave of pain reminds me that my multiple sclerosis will probably prevent me from seeing you grow, see you grow up and that I won't be able to feed myself.*

Lastly, in 6% of the messages, loneliness was discussed by people living with MS, with the feeling of being sidelined by their disease and by society.

The material or financial impact was the fourth QoL topic (32/847, 3.8%). This category included the effect on the financial situation that can lead to debts, bank loans, and difficulties in care management. In this category, 47% of the posts referred to the social security system, including obtaining long-term disease status for MS ("Affection de Longue Durée" in French). In the French health insurance system, individuals with this status have full health care coverage. Messages that mentioned this status reported either satisfaction, requests for information, or discussions on the risk of losing their rights. In several testimonies (34%), patients and their caregivers expressed their need and desire to receive the "Allocation Adulte Handicapé" (allowance for disabled adults) but regretted that the administration refused their application:

> *The MDPH [French organism that helps people with disabilities] refused my boyfriend's #AAH [application]. Saying that his disease is fluctuating. Well, yes, yes. It's true that his #MS doesn't tire him out at all, doesn't make his legs ache, doesn't give him migraines, and doesn't prevent him from finding work.*

Lastly, in 22% of the messages, the authors mentioned financial difficulties related to their chronic condition, through budget calculations that underlined the high cost of living. Some shared an online fundraising pot.

An example of a post is as follows:

> *I have multiple sclerosis, and I can't work. So, can you explain to me how, on €860 a month, I can change my car? I can't take out a loan.*

The impact on daily activities was the least mentioned by people affected by MS and consequently was the fifth and last QoL topic (24/847, 2.8%). In 42% of the posts, people expressed their difficulty in keeping their jobs, particularly when symptoms were increasing.

Examples of posts are as follows:

> *Because of multiple sclerosis, I could no longer do my job. I tried to retrain but I couldn't follow the courses.*
>
> *I've got multiple sclerosis and frankly it's hard to find a job.*
>
> *I lost my old job because of multiple sclerosis, which I've had since I was 16...*
>
> *If I'm struggling to feel well today, it's not because of my MS (...), but because of my employers...*

In 37% of the posts, difficulties in walking, running, and other leisure activities, such as horse riding, were mentioned. Several people also reported the use of a wheelchair, which was seen as a marker of disability.

Examples of posts are as follows:

> *Since I've had multiple sclerosis, I've stopped running. My dog helps me to keep my balance.*
>
> *My daughter is 4. She can see that Mum can't run behind her like she used to, it's impossible to take her bike to the park and some days I need crutches. I say that my legs hurt, that the doctor is looking for how to treat me #multiple sclerosis.*

In 8% of the posts, urinary symptoms, such as dysuria, and their consequences were evoked.

Discussion

Principal Results

The aim of this infodemiology retrospective study [14] was to describe how French-speaking people affected by MS communicated on social media and web forums about their disease in general and about their QoL in the last 5 years. To this aim, we included 3225 posts from more than 30 sources during the study period. The 2 main themes of posts were "Caregivers and family members" (32%) and "Disability" (24%), which confirms that social media is a way to express the reality and burden of people affected by MS. Most of the messages were written by caregivers (2571/3225, 80%), which probably explains the high frequency of messages relating to

XSL•FO
RenderX

relationships and family members. Overall, 847 posts described at least one impact of MS on QoL: relational or social (n=431, 50.9%), physical (n=284, 33.5%), psychological (n=76, 9.0%), and material or financial well-being (n=32, 3.8%), as well as daily activities (n=24, 2.8%). Those results were coherent with the thematic analysis.

Most of the social media users whose messages were analyzed were women aged 30 to 40 years. This corresponds to the age when many people receive a diagnosis of MS and to the age of people who routinely use social media. Among the authors of the retained posts, 42% were women and 28% were men, and 80% (2571/3225) were family caregivers. The large proportion of caregivers in the sample highlights the burden that the disease places on families. We used examples of posts to illustrate our analysis because we think that this is the best way to describe what people are saying on social media.

In the present study, we assessed the impact of MS on QoL through 5 dimensions: physical well-being, social or relational well-being, material or financial well-being, emotional or psychological well-being, and development or activity. The relational impact of the disease was the most frequently mentioned on social media (431/847, 50.9%). This does not mean that the impact of MS on QoL is mainly or only relational. In addition to the fact that many posts were written by caregivers, we think that this is linked to the fact that social networks and digital patient communities are places for discussion, sharing experiences, and looking for support and, therefore, are particularly conducive to capturing this information. Indeed, it is likely that people affected by chronic diseases (patients and family caregivers) experience painful emotions and specific situations in their everyday lives. Caregivers have widely reported the impact on relationships, while people with MS have also reported an impact on their physical and psychological QoL. Posts on social media illustrate their need to share these emotions and experiences and to discuss them with people experiencing the same thing. This is consistent with a recent study on data on the reach and engagement of Facebook posts [15]. This analysis of 5881 comments demonstrated that people with chronic health conditions want to engage on social media and find value in supporting and sharing their experiences with others. This study included people living with MS, migraine, irritable bowel syndrome, rheumatoid arthritis, lung cancer, and prostate cancer.

Comparison With Prior Work

In the literature, we found 2 studies on posts related to MS, 1 on Twitter [16] and 1 on Facebook [17]. The first one [16] analyzed 74,076 original tweets on MS written in English from February to June 2019 and identified 4 main topics: related chronic conditions, disease burden, disease-modifying drugs, and awareness raising. Its findings were consistent with those of the other study based on 7029 posts from 2 MS Facebook groups [17]. The authors reported 8 information categories: information and awareness (70.0%), event advertising and petitions (5.2%), fundraising (5.0%), patient support (3.1%), drug discussion (2.1%), clinical trials and research studies (0.8%), product and drug advertising (0.7%), and other (13.1%). Giunti et al [16] also found that tweets on disease burden and related chronic conditions had the most negative feelings, probably linked to their emotional burden, which seems to be in accordance with our results on QoL. In addition, Della Rosa and Sen [17] reported a high engagement level (eg, in terms of views, likes, and comments) for patient support and information or awareness. This confirms our opinion about social media as a means to look for and/or provide support among people affected by MS.

Some posts classified in the “relational or social impact” QoL were not far from the “psychological impact” QoL dimension. Our classification rules may have underestimated this category. Few posts were categorized in the “financial impact” QoL dimension, which may reflect a taboo surrounding this topic when it comes to discussing it publicly. We also found several posts related to the risk of losing a job or not being able to pursue a career. Such fear can affect several impacts of the QoL and reflect the consequences of MS in the everyday lives of young adults.

Limitations

We recognize some limitations related to our study. First, the analysis concerned only French posts during 5 years, but this does not mean that all people originated from France. Like for all infodemiology studies, only people with internet access who were capable, knowledgeable, and willing to post messages on social media were included. Therefore, this sample might not be representative of the whole population affected by MS [18]. Knowing that patient associations can federate online private communities, it would be of interest to include them in order to better grasp the patients’ reality. Indeed, what patients tell publicly on social media can be different from what they express in private, in patient associations, or with their physicians. Due to ethical and data privacy reasons, our study included only openly accessible online networks and, as a result, could lack potentially relevant information. Nevertheless, social media offers the opportunity to collect data from relatives who share information on the experiences of patients with MS who may not be active on social media.

An extraction bias is also possible because we only considered posts that contained predefined keywords related to our subject. If users expressed themselves by using other words, for instance, without mentioning MS, their posts were not included in the final corpus. To mitigate this bias, the set of keywords was as comprehensive as possible, based on our previous experience in the infodemiology field [8-11].

Another limitation of this study is that multiple posts may originate from a single user. Consequently, certain profiles that are particularly active on social media may be overrepresented in the analyzed corpus, which may influence the apparent frequency of certain themes or reported experiences. Consequently, the results should be interpreted as reflecting the conversations observed online rather than as a representation of the actual prevalence of such experiences within the entire population concerned.

XSL•FO
RenderX

Conclusion

The present study confirms the high burden of MS and the difficulties in everyday life among people affected by MS. It highlights the burden on family caregivers of people with MS, who were most active on social media in our dataset. Our findings confirm that social media is a way for people affected by MS to express themselves and look for support and understanding. This study also shows that social media provides an opportunity to discover the fears, questions, needs, and thoughts of those affected by the disease, particularly caregivers, who are rarely taken into account in research studies.

Acknowledgments

The authors thank Tom Marty, PharmD, for his time and work on this study when he was employed by KapCode.

Funding

This study was funded and carried out as part of the INSPIRE research chair, cosponsored by the EDMUS Foundation, Lyon, France, and EHESP, Rennes, France [19], and directed by Prof Emmanuelle Leray. This chair aims to understand the determinants of inequalities in multiple sclerosis and their impact on the lives of people affected by the disease.

Conflicts of Interest

EL received personal compensation for consulting, serving on a scientific advisory board, speaking, or other activities with Alexion, Biogen, Merck, Novartis, Roche, and Sanofi-Genzyme; none were related to the content of the present study.
SS, PV, and NT are employees of Kap Code, a contract research organization, and are contracted with EHESP to carry out this study.

Multimedia Appendix 1
Extraction query in social media.
[DOCX File, 14 KB - infodemiology_v6i1e88763_app1.docx]

References

1. McGinley MP, Goldschmidt CH, Rae-Grant AD. Diagnosis and treatment of multiple sclerosis: a review. JAMA 2021 Feb 23;325(8):765-779. [doi: 10.1001/jama.2020.26858] [Medline: 33620411]
2. Walton C, King R, Rechtman L, et al. Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS, third edition. Mult Scler 2020 Dec;26(14):1816-1821. [doi: 10.1177/1352458520970841] [Medline: 33174475]
3. Donegan K, Ovelgonne H, Flores G, Fuglerud P, Georgescu A. Social media and m-health data subgroup report. : European Medicines Agency (EMA); 2019 URL: https://www.ema.europa.eu/en/documents/report/social-media-m-health-data-subgroup-report_en.pdf [accessed 2023-08-17]
4. Global social media statistics. Data reportal. URL: https://datareportal.com/social-media-users [accessed 2026-05-31]
5. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. J Am Med Inform Assoc 2012;19(5):777-781. [doi: 10.1136/amiajnl-2012-000990] [Medline: 22759618]
6. Litvinova O, Baral B, Wochele-Thoma T, et al. Efficiency and safety of cannabinoid medical use: an analysis of discussions and observed trends on Instagram. Front Public Health 2024;12:1494018. [doi: 10.3389/fpubh.2024.1494018] [Medline: 39697283]
7. Walsh J, Cave J, Griffiths F. Combining topic modeling, sentiment analysis, and corpus linguistics to analyze unstructured web-based patient experience data: case study of modafinil experiences. J Med Internet Res 2024 Dec 11;26:e54321. [doi: 10.2196/54321] [Medline: 39662896]
8. Valades MT, Montero-Torres M, Lara-Abelenda FJ, et al. Understanding public perceptions and discussions on diseases involving chronic pain through social media: cross-sectional infodemiology study. BMC Musculoskelet Disord 2024 Jul 22;25(1):569. [doi: 10.1186/s12891-024-07687-5] [Medline: 39034398]
9. Brandwatch. URL: https://www.brandwatch.com/ [accessed 2026-05-31]
10. Schück S, Roustamal A, Gedik A, et al. Assessing patient perceptions and experiences of paracetamol in France: infodemiology study using social media data mining. J Med Internet Res 2021 Jul 12;23(7):e25049. [doi: 10.2196/25049] [Medline: 34255645]
11. Cotté FE, Voillot P, Bennett B, et al. Exploring the health-related quality of life of patients treated with immune checkpoint inhibitors: social media study. J Med Internet Res 2020 Sep 11;22(9):e19694. [doi: 10.2196/19694] [Medline: 32915159]
12. Renner S, Marty T, Khadhar M, et al. A new method to extract health-related quality of life data from social media testimonies: algorithm development and validation. J Med Internet Res 2022 Jan 28;24(1):e31528. [doi: 10.2196/31528] [Medline: 35089152]
13. Voillot P, Riche B, Portafax M, et al. Social media platforms listening study on atopic dermatitis: quantitative and qualitative findings. J Med Internet Res 2022 Jan 28;24(1):e31140. [doi: 10.2196/31140] [Medline: 35089160]

XSL•FO
RenderX

14. Eysenbach G. Infodemiology: the epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
15. Green BM, Van Horn KT, Gupte K, Evans M, Hayes S, Bhowmick A. Assessment of adaptive engagement and support model for people with chronic health conditions in online health communities: combined content analysis. J Med Internet Res 2020 Jul 7;22(7):e17338. [doi: 10.2196/17338] [Medline: 32492651]
16. Giunti G, Claes M, Dorronzoro Zubiete E, Rivera-Romero O, Gabarron E. Analysing sentiment and topics related to multiple sclerosis on Twitter. Stud Health Technol Inform 2020 Jun 16;270:911-915. [doi: 10.3233/SHTI200294] [Medline: 32570514]
17. Della Rosa S, Sen F. Health topics on Facebook groups: content analysis of posts in multiple sclerosis communities. Interact J Med Res 2019 Feb 11;8(1):e10146. [doi: 10.2196/10146] [Medline: 30741640]
18. McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? BMC Med 2019 Jan 16;17(1):11. [doi: 10.1186/s12916-018-1247-8] [Medline: 30646913]
19. Chaire INSPIRE Une recherche inédite et innovante centrée sur les besoins de patients [Article in French]. Fondation EDMUS. URL: https://www.fondation-edmus.org/fr/la-chaire-inspire/presentation [accessed 2026-08-05]

Abbreviations

MS: multiple sclerosis
QoL: quality of life

Edited by E Lee; submitted 01.Dec.2025; peer-reviewed by M Lteif, S Gupta; revised version received 16.Jun.2026; accepted 22.Jun.2026; published 19.Aug.2026.

Please cite as:
Leray E, Schück S, Voillot P, Texier N
How Social Media Analysis Offers an Opportunity to Understand the Reality of People Living With Multiple Sclerosis: Descriptive French Study
JMIR Infodemiology 2026;6:e88763
URL: https://infodemiology.jmir.org/2026/1/e88763
doi:10.2196/88763

© Emmanuelle Leray, Stéphane Schück, Pamela Voillot, Nathalie Texier. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 19.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Investigating Online Discussions About Cancer Screening on Twitter (Subsequently Rebranded as X): Corpus Analysis

Martin-Pieter Jansen[1,2], PhD; Hanneke Hendriks[3], PhD; Suzan Verberne[4], PhD; Gert-Jan de Bruijn[5], PhD; Enny Das[1], PhD

[1]Center for Language Studies, Radboud University Nijmegen, Nijmegen, Gelderland, The Netherlands

[2]Amsterdam School of Communication Research, University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam, North Holland, The Netherlands

[3]Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, Gelderland, The Netherlands

[4]Leiden Institute of Advanced Computer Science, Leiden University, Leiden, South Holland, The Netherlands

[5]Department of Communication Studies, University of Antwerp, Antwerp, Flanders, Belgium

Corresponding Author:
Martin-Pieter Jansen, PhD
Center for Language Studies, Radboud University Nijmegen, Nijmegen, Gelderland, The Netherlands

Abstract

Background: While cancer screening is proven to be effective in the early detection of the disease and early detection enables better treatment options, screening uptake has been declining. Research shows that online health information helps people to make health-related decisions. However, not all online health information is credible, and misinformation might play a role in people's choice to take part in screening.

Objective: This study aimed to analyze online discussions about cancer screening programs using corpus analysis. Specifically, we aimed to investigate the full dataset through corpus analysis and misinformation in a manually coded subset. This enabled us to study naturalistic discussions about cancer screening over time, what information people share, and how prevalent misinformation is in these discussions. We differentiated tweets on Twitter (subsequently rebranded as X) for cervical, breast, colorectal, and general screening.

Methods: We extracted a corpus of 55,403 tweets from 2011 to 2023, tweeted by 22,493 users from a database containing over 5.9 billion tweets. We used specific search strings corresponding to different types of screening to gather our corpus. The corpus consisted of tweets, timestamps, hashtags, and shared URLs. We used a machine learning classifier trained on another dataset of tweets about cancer screening to automatically code whether a tweet fell within the scope of the study. We manually coded a randomly drawn stratified subset of 1200 tweets representative of the full corpus regarding year and screening program for the presence of misinformation.

Results: Tweets were not uniformly distributed across different screening programs and over time (χ^2_{36}=4045.99, n=55,403; P<.001). Most tweets discussed population screening in general (n=35,199), and the volume of tweets increased around real-world events. Hashtags in the tweets predominantly focused on the screening programs that were discussed in those tweets. In our corpus, most shared URLs linked to other tweets (n=10,569) or news websites (n=2807). In our coded subset, information was shared in 679 tweets. Overall, 23 tweets contained misinformation. Topics in those tweets showed criticism toward the programs and policies, suspicions about conflicts of interest, and antivaccination beliefs regarding human papillomavirus (HPV). Most users used rhetorical questions, sarcasm, fearmongering, or expressed anger.

Conclusions: Our findings reveal that cancer screening programs are actively debated across social media platforms. We observed that conversations tend to spike in response to real-world events, suggesting social media can serve as a valuable lens into public reactions to health policy changes. Link-sharing behavior was common, though we noted a tendency for sources to reference back to the same platform where discussions originated. Despite finding limited instances of misinformation, we caution that even modest amounts of inaccurate information may have meaningful consequences for public health messaging and screening uptake.

(JMIR Infodemiology 2026;6:e90916) doi:10.2196/90916

KEYWORDS

cancer; cancer screening; misinformation; social media; health communication

XSL•FO
RenderX

Introduction

Background

Cancer screening is a crucial aspect of cancer prevention and early detection. Detection provides better treatment options for cancer and early-stage cancer, thereby decreasing mortality rates [1]. The World Health Organization (WHO) recognizes three cancer screening programs as the most promising: breast, colorectal, and cervical programs [2]. Participation rates of at least 70% are recommended by the WHO for these programs to be effective [3]. However, an increasing number of people do not take part in screening programs, and some levels of participation are already below this threshold [4]. This makes it important to understand factors that may contribute to this nonparticipation. Earlier research already investigated determinants of this nonparticipation regarding demographics and psychological factors, but less is known about (user-generated) content on social media. Besides official sources, the internet has become one of the most prevalent sources of health information [5]. For instance, in the European Union, 85% of adults state that they use the internet for medical or health information [6].

On social media platforms, users share posts about their experiences, discuss opinions on health-related topics, or ask questions about procedures [7-9]. One of the issues with social media is that anyone can post anything, and that most information that is shared contains personal stories [10-12]. These personal stories are often not based on scientific evidence, but research shows that they are often presented as equally valid [10], and while social media has been found to increase access to health information and can function as a decision aid [13], concerns can be raised about the accuracy of information and the prevalence of misinformation in these discussions [14].

The fact that people frequently use social media to talk about health, and that such discussions are often based on personal narratives instead of scientific evidence, makes it vital to understand how cancer screening is talked about in online discussions and how misinformation may play a part in this, as this may influence people's decision to take part in screening [15,16]. We investigate this by quantitatively analyzing a large corpus of 6 billion Dutch tweets on Twitter (subsequently rebranded as X). The study has 3 objectives. First, to gather insights into the volume of tweets and retweets about cancer screening. Second, to determine which types of resources are shared in the tweets. Research shows that this is one of the ways that people share news on Twitter (subsequently rebranded as X), and that the design of the platform makes it easy to do so [17,18]. Investigating which websites are shared can provide descriptive insights into the sharing of different types of external information on the platform. The third and final objective is to explore the amount of misinformation in tweets. By scrutinizing how prevalent misinformation is within cancer screening discussions in the Dutch Twitter (subsequently rebranded as X)-sphere, we provide insight into a part of health-related information that individuals might use to inform themselves about cancer screening programs.

Online Health and Social Media

The internet has emerged as a dominant source of health information, with worldwide usage growing from 2.5 billion in 2013 to 5.2 billion in 2023 [19]. By 2024, an estimated 85% of European adults were using the internet to search for health or medical information [6]. Research indicates that multiple demographic groups in Europe show an increased likelihood of seeking health information online [20-22]. The European Union reported that in 2020, 76% of Dutch adults used the internet to look for health or medical information [23].

Not all online health information comes from medical or governmental websites. One of the types of online health information that people consume is discussions on social media [9,24,25]. Twitter (subsequently rebranded as X) is broadly used for the communication of health-related information, making it a very fitting platform in the context of the current work [26]. The platform had 3.1 million Dutch users in 2023. One of the main reasons that Twitter (subsequently rebranded as X) is an important platform to study is that it is embedded in everyday social and communicative interactions. Moreover, research using actual tweets allows us to study natural conversations in an unobtrusive manner [27]. Moreover, the platform's role as a very public and real-time communication channel highlights the fact that it provides a window on contemporary society [28].

There are several actions that improve the spread or diffusion of messages (ie, tweets). Retweets may indicate the popularity of a message within a network and can be seen as a form of approval [29], especially when there is no comment that disagrees attached to the retweet. Another action that improves the spread of tweets on the platform is the use of hashtags. These word-labels help readers contextualize the message or provide emotional information that was not yet in the tweet [30,31]. Also, "hashtagged" words can result in promoted topics, or "trending topics" as they are called on the platform, gaining extra exposure and showing up in lists with trending words or topics [32].

Another way of sharing information on Twitter (subsequently rebranded as X) is via the attachment of a link to your tweet. This way of sharing news, or other information, means that one adds a URL to a tweet that leads to an external source. While research shows that more and more people rely on social media for their news consumption [33], Twitter (subsequently rebranded as X) often acts as a diffusing factor for news and information, where more traditional news is channeled differently compared to the presocial media age (eg, a tweet containing a link to an online news article) [34]. However, while sharing a URL to strengthen an argument might improve face validity to other users, the websites that are linked do not necessarily contain accurate information. This makes it important to investigate what types of websites the URLs in tweets lead to, as inaccurate information can misinform people, which can lead to lower intentions to take part in screening.

Cancer and Social Media

Earlier work shows that online discussions about cancer screening contribute to people's health-related decision-making,

XSL•FO
RenderX

for example, whether or not to take part in screening [35]. However, it should be noted that differences between cancer screening programs exist in the role of internet information for a citizen's decision to take part in screening. Specifically, research has shown that internet use has been positively associated with taking part in cervical cancer screening but that this was not the same for taking part in breast cancer screening [36].

While health information on the internet means that it may be more accessible for users, information sources that are shared on social media tend to be community-driven and grassroots-oriented (eg, a blog) and not news, from governmental sources, or from an established health institution [37]. Moreover, research shows that official health and news agencies, which typically provide reputable, scientific evidence-based information, are largely underrepresented in discussions on social media [38]. As a consequence, accurate information that would counter misinformation is less prevalent compared to false information. There are other concerns regarding online health information. Information can be one-sided or outdated, as has been reported in information about thyroid cancer [39]. But also, anyone with internet access, with or without expertise, can post information, which may open the door for potential misinformation. This raises concerns about the accuracy of online health information [14].

Misinformation

A major issue associated with online health information is the growing spread of misinformation. Studies indicate that health-related misinformation has become increasingly common across digital platforms [40-42]. These worries are also underlined by the WHO, which speaks of an infodemic [43]. Research argues that features of social media, like the extensive use and relative ease of sharing information quickly, facilitate the spread of misinformation [44]. This is problematic, as studies show that users often encounter difficulties in recognizing misinformation [45,46]. Research in cancer-specific contexts reveals significant misinformation prevalence, with one analysis showing that 44% of highly engaged cancer-related social media posts contained false information [47]. Other work shows that tweets containing misinformation contain significantly more attractive language, fewer science-based facts, use more personal pronouns, and discuss trust issues and deception more than tweets containing correct information [48]. The occurrence of messages containing misinformation is especially pronounced regarding breast cancer screening on social platforms [16,49], where misleading content spans from exaggerated claims about foods' anticancer properties to baseless fears about mammography risks. In this study, we do not assume that we know whether users share false information about cancer screening programs with the intention to harm others, which is often described as disinformation. Therefore, we use the definition of Wardle and Derakhshan [50] for misinformation as "when false information is shared, but no harm is meant." Regarding "false information" in that definition, we follow Tan et al [51] and define this as: "what is considered to be correct or incorrect by expert consensus contemporaneous with the time period of this study."

Research Questions (RQs)

To systematically examine the nature of tweets and content about cancer screening in Twitter (subsequently rebranded as X) discussions, we pose the following research questions (RQs):

- RQ 1. What is the volume of tweets and retweets about cancer screening?
  - RQ 1.1. What are the differences in the volume of tweets and retweets, between types of cancer screening, and cancer screening in general?
  - RQ 1.2. What are the differences in the volume of tweets and retweets between types of cancer screening over time (per year/month)?
  - RQ 1.3. Which #hashtags are used most often, and does this differ between types of cancer screening, and over time?
- RQ 2. How is information from different types of resources (based on URLs) shared in the tweets?
  - RQ 2.1. Are there differences between cancer screening programs regarding what resources or types of resources are mentioned?
  - RQ 2.2. Is there a difference in the use of resources over time? And could these differences be linked to real-world events?
- RQ 3. What is the relative amount of misinformation regarding cancer screening programs in the tweets?
  - RQ 3.1. Are there differences regarding the amount of misinformation (1) over time, (2) between different cancer screening programs, or (3) per resource?

Methods

Study Design

This preregistered [52] content analysis study investigated how the Dutch cancer screening programs that are offered as population screening were discussed on Twitter (subsequently rebranded as X) between January 2011 and March 2023.

Data Collection and Preprocessing

In this study, we used a corpus of tweets that we retrieved from a larger dataset that was developed in the TwiXL project [53]. This project started as the TwiNL collection and contains 50% of all Dutch-language tweets from January 2011 to March 2023. Please note that, while the TwiXL project is still aiming to retrieve tweets, this has become more complex and expensive after March 2023, when the platform decided to limit the API access for researchers [54]. At the time that we used the database to retrieve tweets, 5.9 billion tweets were accessible. We accessed the database via the TwiXL API in Python (version 3.11.1; Python Software Foundation) [55].

To ensure that we did not collect tweets that are outside of the scope of our project, we only downloaded tweets from the database based on four specific strings of keywords related to different types of cancer screening (ie, one for breast cancer screening, one for cervical cancer screening, one for colorectal cancer, and one for population-level cancer screening in

XSL•FO
RenderX

general). For example, “mammogram” or “uitstrijkje” [swab] (Please see Multimedia Appendix 1 for the full search strings).

Our first step in data cleaning was deleting duplicate tweets based on the tweet ID. To ensure we only included Dutch-language tweets in the dataset, we used the langdetect Python library [56] to filter out non-Dutch content. Further manual investigation of the data revealed some noise in the langdetect output, caused by shortened URLs, hashtags, and usernames. Therefore, we decided to only exclude the tweets that were labeled as English, as manually assessing those tweets showed that the package was able to correctly label those tweets. Our final exported data consisted of 100,384 tweets. To ensure that tweets fell within the scope of this project, we used machine learning (a word-based Random Forest classifier) [57] to code whether tweets fell into the scope of the study (binary classification) [58]. The classifier was trained on a manually labeled dataset consisting of 1629 tweets.

We evaluated the classifier on a held-out test set of the labeled data. Precision at a threshold of .50 was .73 for inclusion, with a recall of .95. For exclusion, the precision was .67 with a recall of .24. As we wanted a higher precision (ie, less irrelevant tweets), we set the classification threshold for the ‘inclusion’ label at .70. At this threshold, the precision of the model was .81 for inclusion, with a recall of .77; for exclusion, the precision was .54 and the recall was .60. We focus more on precision than recall, as for this project we deem the correctness of positive predictions more important than the ability to find all relevant instances. Moreover, there are more inclusion tweets compared to exclusion tweets, meaning that precision is a more meaningful metric than accuracy. Our final corpus consisted of 55,403 tweets, tweeted by a total of 22,493 users.

We first investigated whether tweets were replies to other tweets by looking if a tweet started with an ‘@,’ we added a binary variable in our data indicating if this was true. We then did the same to see if tweets were retweets by looking if a tweet started with “RT.” Second, we used the rapidfuzz Python library [59] to calculate a similarity score (0 - 1) based on the Levenshtein distance. Manual assessment of the tweets shows that retweets had a similarity ratio of >.85. We added a binary variable to indicate whether a tweet was a retweet if any of the conditions were true.

We cleaned the scraped URLs in the corpus. Twitter (subsequently rebranded as X) uses shortened URLs if a link is added to a tweet (eg, bit.ly). We used the requests Python library [60] to expand the URLs, and extracted the top-level domains via the tldextract Python library [61] for further analysis of the mentioned source domains.

Coding Procedure

We manually coded a stratified random sample of 1200 tweets for the presence of misinformation (RQ 3). The sample was representative of the full corpus with respect to year and screening program. Tweets were double-coded by 4 coders making use of a codebook which is available on OSF in Dutch and as a version that is translated into English [52].

Before coding the full subset of 1200 tweets, the coders test-coded 125 tweets from these 1200. All tweets were double-coded in a rotational format, and all coders coded 50 tweets in total. This means that tweet 1 - 50 were coded by coder 1, coder 2 coded tweets 25 - 75, coder 3 coded tweets 51 - 100, and finally coder 4 coded tweets 76 - 100 and tweets 1 - 24. The same rationale was applied to the coding of the other 550 tweets for each coder, resulting in all tweets being double coded.

First, the tweet was coded for inclusion (with three labels: inclusion, exclusion, and doubt). While we already used the specific search string and the machine learning classifier to make sure that tweets fell into the scope of the project, we still coded our subset for inclusion so that we were sure the correct tweets were included and excluded. Second, the tweets were coded for the cancer screening program (general, cervical cancer, breast cancer, and colorectal cancer) that was discussed in the tweet. General was coded if the tweet discussed population screening in general, or if more than one cancer screening program was discussed in the tweet. Third, tweets were coded for information sharing in general (yes-no). To be able to better distinguish misinformation in the tweets, it was coded whether information was shared in the tweet, regardless of whether this was correct information. Instructions clearly stated that this meant that information was presented as general, nonpersonal facts, with or without a source, and that it did not concern purely personal experiences. Finally, the tweets that were coded as having information shared in them were also coded for a suspicion of potential misinformation (yes-no). We did this because the coders could run into tweets that they were unsure of whether they contained misinformation (as our coders were no medical experts). The tweets that were coded for a suspicion of misinformation were exported and later manually assessed for the presence of misinformation by the main researcher (MJ) and other members of the project team and authors of this manuscript. To ensure we had as much data as possible, we included all tweets from the coded subsample that were coded for inclusion or doubt about inclusion at least once.

Coding and Reliability

As indicators of agreement between the coders, we analyzed Cohen Kappa (κ) and the agreement percentage. For Cohen Kappa, we use the classification of Landis and Koch [62]. While we are reporting Cohen Kappa, we would like to underline that there are certain issues with the statistic. For example, in cases of class imbalance, a higher level of agreement can lead to lower test statistics, and the value can be inflated due to multiple answer options and/or coders [63,64]. This is why we choose to also report the plain percent agreement (PA) between our coders [65].

As described earlier, a briefing and test-coding meeting was organized by the main researcher before the coders started with test-coding of the first part of the subset of tweets. The tweets that were coded in the meeting were exported from the corpus and did not overlap with the tweets in the eventual subset of tweets. After this training, the coders started with the first subsample of 125 tweets. After this initial coding stage, the Kappa statistic and PA for the variables: Cancer Screening Program (κ=0.79, PA=85%), inclusion (κ=0.55, PA=77%), information sharing (κ=0.60, PA=75%), and suspicion of

XSL•FO
RenderX

misinformation (κ=0.45, PA=68%) were calculated. Overall, the agreement according to the classifications of the Kappas was moderate to substantial, and the PA showed overlap in 76% of all cases.

After the coding of the first subset, another meeting was organized to discuss any disagreements, inconveniences, or things that were unclear that the coders came across while coding. After that, the other 550 tweets per coder were released. The final Kappas for the variables were fair, but far from desired: Cancer Screening Program (κ=0.95, PA=97%), inclusion (κ=0.45, PA=71%), information sharing (κ=0.38, PA=69%), and suspicion of misinformation (κ=0.36, PA=62%). After assessing the data, we calculated what would happen with the reliability indicators if each coder was left out of the analyses. We found that for one coder, the average Kappa increased by 0.1 if they were left out, which is what we did, and we finally ended up with the following reliability indicators: Cancer Screening Program (κ=0.95, PA=97%), Inclusion (κ=0.48, PA=71%), Information sharing (κ=0.51, PA=74%), and Suspicion of Misinformation (κ=0.61, PA=81%). These reliability indicators were deemed acceptable (and are moderate or better) and were much better than the indicators for all coders combined.

Ethical Considerations

We received ethics approval from the Ethics Assessment Committee Humanities of Radboud University under approval number 2024 - 1889 on June 27, 2024. While we did not collect usernames (these are pseudonymized), the combination of the following data in the dataset may lead to identification: tweet content, URLs (if shared in the tweet) (eg, hyperlinks to a news website), tweet ID, date of publication. If the tweet ID is used to find the tweet online, a user account could be identified. However, the usernames of the profiles that posted the tweets were pseudonymized and not anonymized. This also meant that if the tweet was a reply, retweet, or mention, the tweet included the username of the person that was mentioned, as it was part of the text of the tweet. It depends on what the user has made available whether this information can be used for the identification of an individual. The author ID or username will not be made publicly available. TwiXL does not include private tweets. Therefore, all information is publicly available. However, as the tweets in the database were scraped at different moments, it could be that tweets that are in the dataset have since been deleted by the user.

As we have collected tweets related to cancer screening, tweets that discuss the screening program itself were not expected to contain personal health information (ie, stating that they are invited to the screening, which is based on age groups mostly). However, the possibility existed that individuals did disclose personal health information. For instance, if a user tweeted that they received the results of their Pap smear and that these are positive and that further medical examination is needed. In that case, personal health information is collected.

Finally, the usernames of the profiles that posted the tweets were pseudonymized and not anonymized. This allowed us to see the structure of the conversations and how people interact on the platform and take part in (or leave) discussions. Moreover, if the tweet mentioned someone's name or username in the text of the tweet (string), we also processed that information as it was part of the text of the tweet (eg, two friends who know each other and mention each other by their first name in a discussion and not their account names).

Analysis

RQ 1: Volume of Tweets on Cancer Screening

For our first set of RQ's, we counted the number of tweets and retweets per year to make crosstabulations to provide a descriptive overview. We did this for all tweets taken together, but also for each cancer screening program separately. We used chi-square tests to see if the number of tweets or retweets for each respective program and each year was higher or lower than the expected values and used a P value of $P<.001$ for these differences to be deemed significant. We did that because the sample size was very large and this may inflate the P value. We also calculated standardized residuals to examine which cells differ from the expected values in those cells. For the final part of the first RQ, we counted the number of hashtags in total, per cancer screening program and per year.

RQ 2: Shared Resources in Tweets on Cancer Screening

For the second RQ, we investigated information sharing in the tweets via the URLs that were added to tweets as attachments. We extracted the top-level domains as website names. Subsequently, we counted the times each website was included as an attachment and made rankings in total and per cancer screening program. Finally, we looked at the top 10 most used URLs over time.

RQ 3: Misinformation in Tweets on Cancer Screening

For the final RQ, we investigated the relative amount of misinformation in a stratified subset of 1200 tweets that were manually double-coded. We counted the relative amount of misinformation in the subset as well as the number of tweets potentially containing misinformation per cancer screening program.

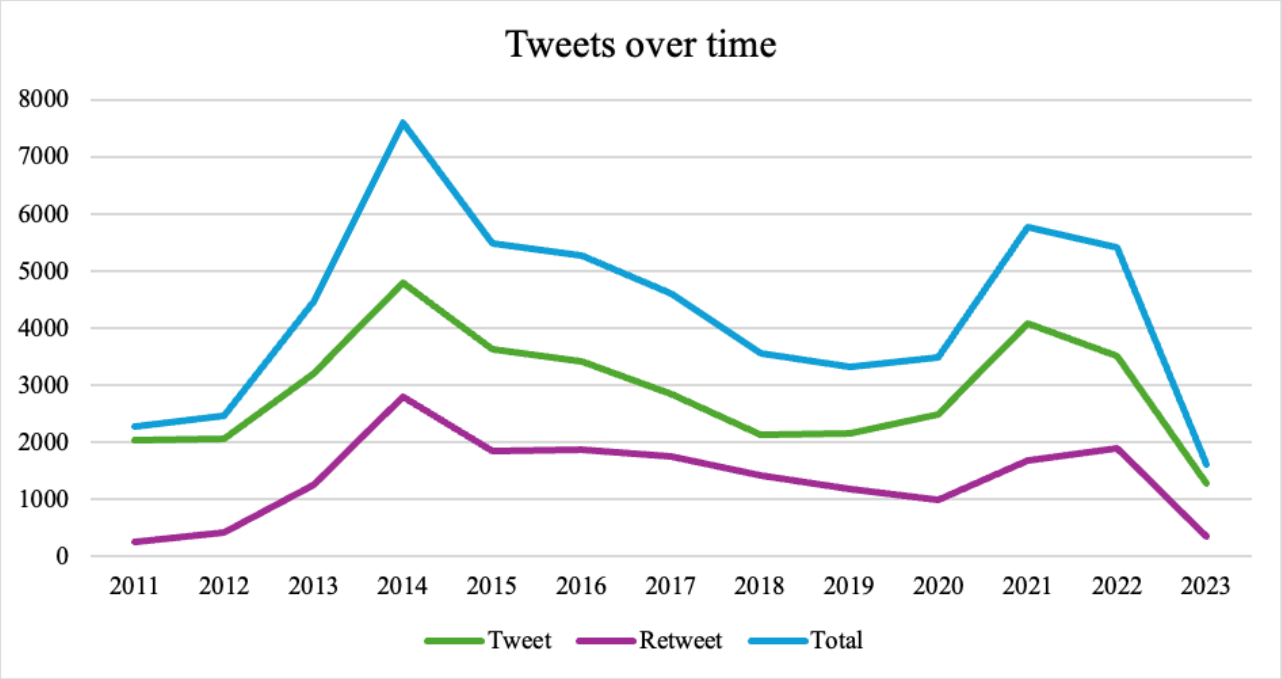
Results

RQ 1: Volume of Tweets on Cancer Screening

To address our first RQ, which asked: "What is the volume of tweets and retweets about cancer screening?," this section examines the overall dataset. The dataset contained 55,403 tweets. In our sample, 10,312 out of 55,403 tweets (18.6%) started with a "@," indicating that the tweet is a reply. We also investigated how many tweets were retweets in our dataset. In total, out of the 55,403 tweets in our dataset, 17,733 (32%) were retweets. To answer RQ 1, we investigated the differences in volume of tweets and retweets over time. Figure 1 shows the volume of tweets and retweets over time. Multimedia Appendix 2 shows an overview of this including standardized residuals.

XSL•FO
RenderX

Figure 1. Volume of tweets, retweets, and regular tweets over time.

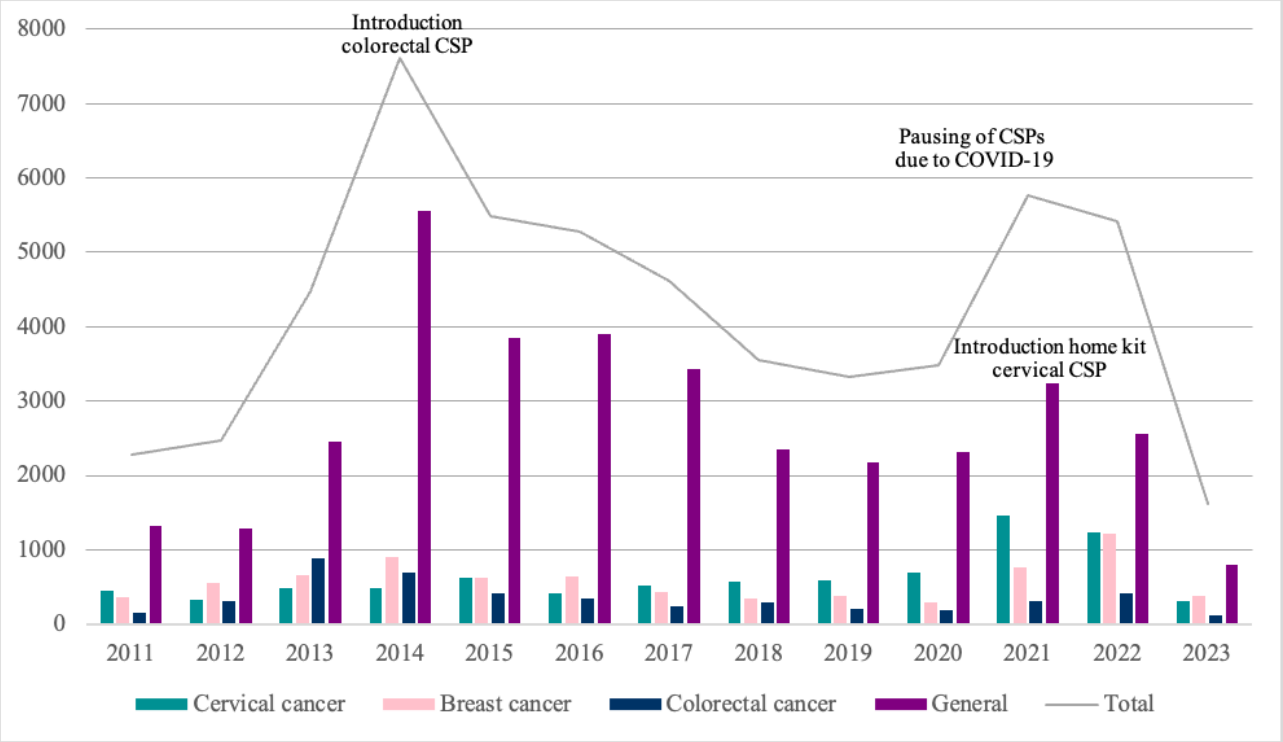


Chi-squared tests showed a significant relation between year and whether a tweet is a retweet or not (χ^2_{22}=1720.88, n=55,403; *P*<.001). Examining the standardized residuals, we found that all cells besides the tweets in 2015 are overrepresented (SR>2) or underrepresented (SR<–2) in the volume of retweets from 2011 to 2023. The biggest deviations are the retweets in 2011, 2012, and 2023, where the volume was much lower than expected. Conversely, the number of tweets in 2011 and 2012, and retweets in 2018, were much higher than expected. We also found that the tweets were not distributed uniformly across the different cancer screening programs over time (χ^2_{36}=4045.99, n=55,403; *P*<.001).

Multimedia Appendix 3 shows the volume of tweets per cancer screening program over time, including standardized residuals. Standardized residuals show that the colorectal cancer screening program in 2013 was the biggest positive deviation from the expected number of tweets. This deviation co-occurred with the introduction of the population-based screening program for colorectal cancer, which started with a pilot in 2013 and was introduced a year later [66]. The second biggest deviation was the cervical cancer screening program in 2021; the third and fourth were the breast and cervical cancer screening programs in 2022. These increases in tweets co-occurred with the pausing of all population-based screening programs due to the allocation of health care resources to battle the COVID-19 pandemic [67], and the release of advice for the home kit for cervical cancer [68]. Figure 2 presents a visual overview of the tweets per cancer screening program per year, including those real-world events.

XSL•FO
RenderX

Figure 2. Number of tweets per cancer screening program from 2011 - 2023; the line is total tweets about CSPs. CSP: cancer screening program.



In total, 20,161 tweets (36.4%) of the 55,403 tweets in our dataset contained at least one hashtag. Most hashtags were used multiple times: 4685 hashtags were used a total of 68,003 times. A total of 20,161 tweets contained at least one hashtag; 8679 tweets contained at least two hashtags, and 3050 contained at least 3 hashtags. The top 5 most used hashtags in the dataset were (translated): breastcancer, populationscreening, cancerscreening, colorectalcancer, and breastcancerscreening. Table 1 shows the top 10 hashtags per cancer screening program (translated to English). The use of the top 10 hashtags overall, and how much they were used over time, can be found in Multimedia Appendix 4.

XSL•FO
RenderX

Table . Top 10 hashtags per cancer screening program.

	Cervical cancer		Breast cancer		Colorectal cancer		General	
	Hashtag	Frequency, n	Hashtag	Frequency, n	Hashtag	Frequency, n	Hashtag	Frequency, n
1	swab	291	mammography	377	coloscopy	218	populationscreening	2139
2	cervicalcancer	164	breastcancer	248	colorectalcancer	145	breastcancer	1589
3	populationscreening	71	Corona	174	populationscreening	53	Populationscreening	894
4	HPV	43	breastexam	92	stooltest	44	vacancy	892
5	coronavirus	41	mammogram	76	colonoscopy	37	colorectalcancer	815
6	health	37	health	56	health	36	news	689
7	hpv	37	populationscreening	44	knowledgesharing	36	cancerscreening	385
8	pauliencornelisse	35	news	42	colorectalexam	34	breastcancerscreening	370
9	GGD	31	thermography	39	onpc[a]	32	colorectalcancerscreening	304
10	NZa[b]	29	screening	30	complications	29	health	243

[a]onpc is a French talk show, manual assessment of these tweets showed that they were French, which could be because we only excluded English language tweets, please see Methods.

[b]NZa is the Dutch Healthcare Authority

RQ 2: Shared Resources in Tweets on Cancer Screening

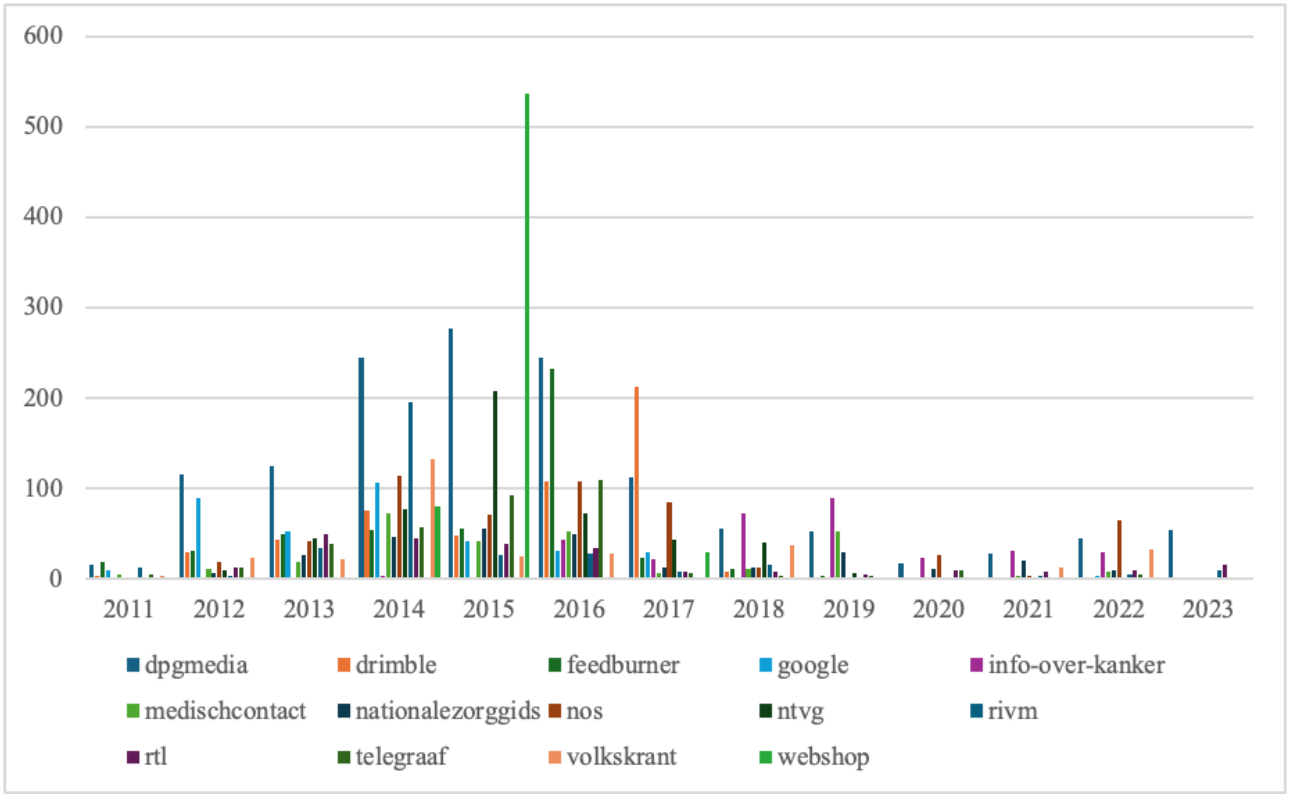
This section addresses our second RQ, which asked: "How is information from different types of resources (based on URLs) shared in the tweets?" We discuss the 10 most shared websites in more detail. In total, 23,579 (42.6%) tweets out of the 55,403 tweets in our dataset had a URL attached to them. The most shared website in the dataset was Twitter (subsequently rebranded as X) (n=10,569). If Twitter (subsequently rebranded as X) as a website is attached to a tweet, this means that people shared a link to another tweet attached to their original tweet. Please note that this is not the same as a reply or a retweet.

The second most shared website in the dataset is dpgmedia (n=1384). This means that news or an online article was shared in this tweet. DPG Media Group ("De Pers Groep" [The Press Group]) is one of the largest media publishers in the Netherlands. It owns some of the country's largest national newspapers (eg, het Algemeen Dagblad, de Volkskrant, Trouw), and regional newspapers (eg, Brabants Dagblad, De Gelderlander, Tubantia). The main reason that DPG is the most shared website is that the publisher's cookie pop-up is on that top-level domain (ie, dpg.nl), which leads to a website of the newspaper after agreeing to or rejecting the cookie pop-up.

The third most shared domain was "webshop" (n=650), which upon further inspection corresponded to affiliate redirect links from the marketing platform TradeTracker [69]. All tweets containing this domain were sharing job vacancies associated with the cancer screening programs, presumably distributed through a recruiting agency that used TradeTracker to track clicks before redirecting users to the actual vacancy page. Since the tweets in our dataset are pseudonymized; however, we cannot confirm this interpretation with certainty. As reflected in Figure 3, these tweets were concentrated almost exclusively in 2015, suggesting the agency discontinued use of this platform shortly thereafter. Because the URLs were no longer operational at the time of data collection, they resolved to TradeTracker's general homepage. The fourth most shared website was NOS (n=547), the Dutch public broadcaster, meaning that news or an online article was shared in the tweet. The fifth most shared website was Drimble (n=530). Drimble is a so-called news aggregation website. The website collects and shares news and information, with a special focus on local Dutch news.

XSL•FO
RenderX

Figure 3. Shared sources over time.



The sixth most shared website was ntvg, the website of the Dutch Journal of Healthcare (Nederlands Tijdschrift voor Geneeskunde; n=504). The seventh most shared website was feedburner, a platform for managing RSS feeds (n=480). The eighth most shared URL was Google (n=366). The ninth most shared was Telegraaf, a Dutch newspaper (n=346). Finally, the tenth most shared website was RIVM, the website of the Dutch National Institute for Public Health and the Environment (n=343). Figure 2 shows the number of shares per website over time. In this figure, we excluded "twitter" (subsequently rebranded as X) as a shared resource, as the volume was so big that the other sources were not distinguishable anymore. We see one large peak for the Dutch Journal of Healthcare in 2015. Manual assessment of those tweets shows that there were tweets that were critical of colorectal cancer screening, and that those tweets shared a medical paper published in 2015 [70]. Other users were also critical of colorectal screening but shared a different article [71]. Finally, the journal posted an interview about the benefits and risks of colorectal cancer screening [72]. Those tweets resonated within the community and were retweeted relatively often. For more information, the crosstabs and figure including Twitter (subsequently rebranded as X) as a source, please see the appendix file on OSF [52].

In conclusion, our results show that most tweets that share a URL share one back to Twitter (subsequently rebranded as X). Besides that, 2807 tweets share links to news websites or sites belonging to a newspaper, and 504 tweets share a link to a scientific journal. Finally, the website of the body that organizes the cancer screening programs is the tenth most shared website but is still only shared 343 times in 23,579 tweets.

RQ 3: Misinformation in Tweets on Cancer Screening

This section addresses our third RQ, which asked: "What is the relative amount of misinformation regarding cancer screening programs in the tweets?" The definition of misinformation we used can be found in the introduction. It should be noted that we do not equate criticism of cancer screening programs with misinformation. There is ongoing scientific debate about the effectiveness and harms of cancer screening, particularly for breast cancer, including substantive disagreement about overdiagnosis, cost-benefit ratios, and mortality benefits [73-75]. What we coded as misinformation were tweets making specific factual claims without any evidence base within the tweet itself, that is, claims not supported by a source, reference, or verifiable reasoning, and that contradict the scientific consensus at the time of the study.

To answer this RQ, we investigated a representative subsample of 1200 tweets (see Methods). The tweets were first coded for inclusion in the study, meaning that they contained references to the Dutch cancer screening programs. While we classified the tweets for inclusion in the project, we also manually coded the subset of 1200 tweets for inclusion and found that 924 tweets were discussing the Dutch cancer screening programs. In total, 679 tweets out of the 924 (73.5%) shared information about the cancer screening programs. From that subset, 266 tweets (39.2% of the subset) were flagged as potentially containing misinformation. Of this subset, 23 tweets (see OSF), which is 3.4% of the tweets in which information was shared, contained misinformation. Eight tweets pertained to cancer screening programs in general, and the other 8 tweets pertained to breast cancer screening. Four tweets pertained to the cervical screening program, and 3 tweets pertained to the colorectal cancer screening program. Due to the low number of tweets containing

XSL•FO
RenderX

misinformation, other than preregistered, we do not differentiate beyond the different cancer screening programs or over time.

While a low number of tweets contained misinformation, we aimed to provide insight into the contents of those tweets. The tweets about cancer screening in general that contained misinformation included claims without an evidence base about the harms of population-level screening, unsupported assertions that the government tracks citizens who opt out of screening, and unsubstantiated suspicions about conflicts of interest (eg, pharmaceutical industry involvement). Tweets containing misinformation about breast cancer screening included claims without an evidence base in the tweet: assertions of ineffectiveness stated as established fact (eg, that screening does not save lives and is a colossal failure) without reference to the scientific literature, policy criticism framed as factual implication (eg, suggestively implying that pausing screening during COVID-19 caused large numbers of deaths), and medical confusion (eg, claiming that more mammograms directly cause more mastectomies and breast cancers). It should be noted that questions about the mortality benefit of mammography are genuinely contested in the scientific literature. We were therefore careful to code as misinformation only those tweets that made unqualified factual claims misrepresenting or distorting this debate, rather than engaging with it.

Tweets describing cervical screening with misinformation discussed unsupported distrust of HPV vaccination and unsubstantiated policy criticism. In this case, the policy criticism is misinformation that would be in favor of cancer screening rather than opposed to it: one user claimed that more frequent screening always equals fewer deaths, which oversimplifies the evidence. Tweets containing misinformation about colorectal screening made unsupported claims about harms caused by screening (eg, deaths due to screening attributed without evidence to the procedure) and expressed financial cynicism without factual grounding. Looking at the emotional and rhetorical styles in the tweets containing misinformation, most users used rhetorical questions, sarcasm, fearmongering, and expressions of anger. In sum, we did not find much misinformation in the subset of tweets. The misinformation that was present appeared to be primarily rooted in claims without an evidence base, nested within criticism and concerns about cancer screening, as well as unfounded assumptions about pharmaceutical industry involvement. Please note that these results may be anecdotal in the broader context of misinformation and cancer screening.

Discussion

Principal Findings

In this study, we investigated how the Dutch cancer screening programs were discussed on Twitter (subsequently rebranded as X) between 2011 and 2023. We did this by conducting a corpus analysis of an existing corpus of Dutch tweets. We had three overarching RQs. First, we looked at the volume of tweets about cancer screening. Second, we investigated how information from different sources is shared, based on attached URLs. Third, we analyzed a manually coded subset of tweets to scrutinize the relative amount of misinformation in the tweets.

Online discussions have the potential to contribute to people‘s thoughts about cancer screening programs, and these thoughts may predict their eventual choice whether to take part in screening. This study shows that cancer screening programs are discussed on social media, and that most tweets discuss cancer screening in general. Discussions increase around real-world events, underlining that social media discussions provide insights into events that happen in society. We find that people share news on social media by attaching links, but that these links also often circle back to the same platform. While we did not find much misinformation, we emphasize that these small amounts might still have large effects.

In the following, we discuss our 3 main findings. First, we found that of the total number of tweets, most tweets discussed cancer screening in general, and that this was true for all separate years as well. Further, we found that regarding the volume of tweets over time, tweets were not uniformly distributed across cancer screening programs and over time. More specifically, 2 peaks occurred. The first peak (2014, general and colorectal cancer) in the number of tweets in our data was in 2014, which co-occurred with the introduction of the colorectal cancer screening program. We see that compared to other years, the volume of tweets about colorectal cancer screening was higher in those years, as were tweets discussing cancer screening in general in 2014, which we treat as a combined, single peak. We tentatively conclude that the introduction of a new screening program (the colorectal program) also started discussions about screening in general.

The second peak in our data occurred in 2021 (2021, cervical and general cancer). This peak co-occurred with 2 real-world events. The first event was the pausing of the screening program in the Netherlands due to the COVID-19 pandemic in 2020. Research shows that only during the first peak of the COVID-19 pandemic, search volume for terms related to cancer screening programs on search engines decreased, and that the volume went back to the levels before the pandemic in a few months [76]. This might indicate that people only started discussing the pausing of cancer screening programs after a while, as they first heavily discussed the pandemic itself, as it was very newsworthy at that specific time and social media consumption increased [77]. The second real-world event that co-occurred around the second peak was the introduction of the home kit for the cervical program, which was introduced in 2021. In line with this, we saw an increase in the volume of tweets about cervical cancer screening, particularly in that year. This underlines that social media, besides traditional media, has an active role as a real-time means of information sharing [78], which has recently been confirmed in the case of cervical cancer on TikTok [79]. Moreover, it confirms earlier findings that social media and news media can complement each other regarding setting the public agenda [80], and other research showing that in regular (news) media, peaks in portrayals exist around real-world events in cancer screening [68], COVID-19 [77], and H1N1 [81]. This emphasizes that social media is an important channel to keep investigating, as it not only offers a window on society, but it is also both an important source and an important outlet for (health) information.

XSL•FO
RenderX

Our second main finding regarded information sharing in tweets by means of attaching links to other websites. We found that almost half of the tweets in our dataset included a URL. This way of sharing information on Twitter (subsequently rebranded as X) leads users to external sources, or other tweets on the platform, although it is not the same as retweeting or replying. For more than 40% of all tweets that had a URL attached, the link led to another tweet. In addition to referring to another tweet, around 12% of all tweets sharing a link contained a URL to a news website, or websites belonging to a newspaper. This brings us back to the previous argument that social media is not only an important source but also an outlet for news. In this case, the news that is published by publishers is shared by Twitter (subsequently rebranded as X) users, which may amplify the reach of news articles [17,18]. Further, in two percent of tweets accompanied by a URL, a link to an academic journal was shared, and in around one-and-a-half percent of tweets, a link to the governmental body that organizes the cancer screening programs was shared.

While we were unable to investigate the content of the linked tweets that were attached via the URLs, as they were not scraped in the original dataset, our results show that the attached links mostly led to reliable websites. This contrasts with other research showing that official health and news agencies were largely underrepresented in health-related discussions on Twitter (subsequently rebranded as X) [38]. However, we would like to emphasize that the percentages mentioned are relative to the tweets that referred to another website, which was around half of the original corpus, which means that the percentages would be half as large compared to the full corpus. This means that, looking at all tweets, most content is still user-generated and without an attached link to an external source, which confirms earlier research on cancer screening discussions on the platform [37].

Surprisingly, while there are rightful concerns about the sharing of incorrect health-related information on Twitter (subsequently rebranded as X) [40], we did not find websites with “alternative” or incorrect information in the attached URLs. However, as many URLs led back to tweets, they might be present indirectly, as the referred tweet may contain links to misinformation. In the scraped dataset, we could not access those tweets anymore, as they were not scraped in the original dataset. Moreover, many tweets do not exist anymore as people deleted older tweets or their accounts in general. This means that these findings should be considered in light of the dataset.

The third, and final, main finding of this work regards misinformation in tweets about cancer screening. Contrary to earlier research [47], which showed that 44% of the top 100 most popular cancer-related tweets contained misinformation, we found that in tweets where information was shared, less than 4% contained misinformation. Misinformation regarding cancer screening in general and the breast program was found eight times each, while four tweets about cervical screening and three tweets about colorectal screening contained misinformation. That we found more misinformation regarding breast cancer and general screening is partially in line with earlier work showing that breast cancer screening misinformation is most pronounced on social media [16], and Twitter (subsequently rebranded as X) specifically [49].

Before discussing the content of these tweets, it should be noted that criticism of cancer screening programs is not the same as misinformation. There is ongoing and substantive scientific debate about the effectiveness and harms of screening, particularly for breast cancer, including disagreement about overdiagnosis, mortality benefits, and cost-benefit ratios [73-75]. Policy criticism informed by this debate does not constitute misinformation as we define it. What we coded as misinformation were tweets making specific factual claims without an evidence base in the tweet itself, that is, claims not supported by a source, reference, or verifiable reasoning, and that contradict the scientific consensus at the time of the study.

The content of the tweets containing misinformation showed various subjects of discussion. Tweets with misinformation discussing population-based cancer screening in general mainly contained claims without an evidence base about the programs, their effectiveness, and the cost-benefit ratio of screening. Whereas tweets regarding the breast cancer screening program discussed the ineffectiveness of screening and policy failures without grounding, tweets regarding cervical and colorectal screening leaned more into existing conspiracy beliefs and showed beliefs of pharmaceutical industry involvement and financial cynicism (ie, the assumption that a few people get very rich over screening programs). This is in line with earlier research in general health information showing that different semantic features are found for different health topics [82]. However, we can only speculate about these ratios as we consider the volume of tweets containing misinformation in our subsample too small to draw conclusions. Moreover, as prior research found higher levels of misinformation in online health communication [40,41], we encourage future research to investigate misinformation across different domains of health communication and potentially compare these domains regarding the misinformation that appears. We also want to underline that several tweets that we classified as misinformation are indeed incorrect information but may be classified as personal narratives. These narratives are ungrounded, or based on memories of experiences (ie, mental models), or misperceptions that may not be in line with scientific consensus. They might not be tweeted with harmful intentions but may rather reflect a lack of knowledge [83,84].

As a final point, we want to emphasize that we do not mean to downplay the presence and risks of misinformation in online health communication. While we did not find many instances of misinformation in our data, we simultaneously do not want to underestimate the effects that even small amounts of misinformation can have. Research shows that cancer-related misinformation is designed to resonate through attractive language, personal pronouns, and deception [48] and that platforms facilitate the spread of misinformation through their respective designs [44]. These factors can strengthen one another, meaning that even small amounts of misinformation may have considerable effects.

XSL•FO
RenderX

Limitations and Future Work

In this work, we made use of an existing corpus of tweets. While the dataset consisted of 50% of all Dutch-language tweets from 2011 to March 2023, we recognize that there is still a chance that we missed certain tweets by chance. Unfortunately, the dataset lacks metadata. This means that we could not investigate the popularity of tweets in the sample. It could be, for instance, that the tweets containing misinformation were retweeted more often or liked, leading to a bigger reach and larger diffusion effects compared to other tweets. Previous research showed that articles containing rumors are found to be shared over three times more often compared to scientific information [80]. Future research may investigate this through other studies that work with actual Twitter (subsequently rebranded as X) data. While we recognize that the company now wants enormous amounts of money for academic access, and that researchers are still unsure whether the datasets represent the whole platform, we encourage other platforms and projects to keep investigating Twitter (subsequently rebranded as X), as this is a great way to investigate natural conversations. Moreover, we encourage researchers to investigate other platforms in a similar manner, as cross-platform validations will strengthen our understanding of social media discussions. This is very needed in times of questionable transparency by platforms.

We were unable to distinguish between disinformation and misinformation because we did not know whether the users who wrote the tweets intended harm. It is possible that the inaccurate information shared stemmed from a lack of knowledge rather than a malicious intent. While both disinformation and misinformation can have significant effects, the distinction between them remains important, as deliberate choices can amplify harmful messages and expand their reach.

Regarding the authors of the tweets containing misinformation, it would be interesting to investigate who is sending these tweets. However, because the data were pseudonymized, we could not identify the types of accounts behind the tweets containing misinformation. As a result, we cannot say with certainty whether misinformation came from individual users, organized accounts, or alternative news platforms. Based on the writing style of the tweets, we suspect these were mostly individual users, but we cannot verify this. We encourage future research to investigate this further.

Further, we would like to encourage researchers to extend our results by investigating if they can be applied in different contexts. While at first it might seem obvious that our results could be applicable to other WEIRD (Western, educated, industrialized, rich, and democratic)–countries, they might even replicate in other geographical contexts. Moreover, it is also possible that other health communication-related subjects may replicate these results, and we encourage new studies combining the approach of working with existing data to investigate natural conversations about online health information.

Finally, results of this study do not predict how people make choices, nor are we able to make causal claims. While we did provide insights into factors that, supported by earlier work, might play a role in individuals' health-related decision-making (ie, whether to take part in cancer screening programs), we were not able to establish causal claims and recognize that decision-making is a complex process. Future work may be able to investigate this process in a more causal way, or at least through retrospectively asking individuals what information they consider when making a choice to participate (or not), either qualitatively or quantitatively.

Conclusions

Online discussions have the potential to contribute to people's thoughts about cancer screening programs, and these thoughts may predict their eventual choice whether to take part in screening. The current work presents an overview of discussions about cancer screening in the Dutch Twitter (subsequently rebranded as X)-sphere. This study shows that cancer screening programs are discussed on social media, and that most tweets discuss cancer screening in general. Discussions increase around real-world events, underlining that social media discussions provide insights into events that happen in society. We find that people share news on social media by attaching links, but that these links also often circle back to the same platform. While we did not find much misinformation, we emphasize that these small amounts might still have large effects.

Acknowledgments

MJ was affiliated with the Center for Language Studies at Radboud University at the time of conducting this study and is currently affiliated with the Amsterdam School of Communication Research at the University of Amsterdam.

The authors thank the people working on the TwiXL project, Dr. Olga Eisele and Dr. Anne Kroon, for their help in providing us access.

The authors thank Emma Michielsens for developing the machine learning classifier.

The authors thank Madelon Bruggink, Naomi Lansink, Emma Thijssen, and Roos van de Wiel for their help in the coding of the manual dataset.

Funding

This research is part of the SENTENCES (Social Media Analysis To Promote Cancer Screening) project and is funded by ZonMw under grant project number 555004205. Project leader for this project is ED, who is also an author of this paper.

XSL•FO
RenderX

Data Availability

Data that was used in this study originated in a different project; therefore, we are not able to publicly share the dataset.

Authors' Contributions

Conceptualizations: MJ, HH, SV, GB, ED
Data curation: MJ
Formal analysis: MJ
Funding acquisition: GB, HH, ED, SV
Methodology: MJ
Project administration: ED
Supervision: MJ, ED
Validation: MJ
Writing – original draft: MJ
Writing – review and editing: MJ, ED, HH, SV, GB.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strings in Dutch and English.
[PDF File, 67 KB - infodemiology_v6i1e90916_app1.pdf]

Multimedia Appendix 2
Tweets and retweets over time.
[PDF File, 55 KB - infodemiology_v6i1e90916_app2.pdf]

Multimedia Appendix 3
Tweets per cancer screening program over time.
[PDF File, 63 KB - infodemiology_v6i1e90916_app3.pdf]

Multimedia Appendix 4
Top 10 hashtags over time.
[PDF File, 61 KB - infodemiology_v6i1e90916_app4.pdf]

References

1. Bongaerts THG, Büchner FL, Crone MR, et al. Perspectives on cancer screening participation in a highly urbanized region: a Q-methodology study in The Hague, the Netherlands. BMC Public Health 2022 Oct 15;22(1):1925. [doi: 10.1186/s12889-022-14312-4] [Medline: 36243684]
2. A short guide to cancer screening: increase effectiveness, maximize benefits and minimize harm. World Health Organization. 2022. URL: https://www.who.int/europe/publications/i/item/9789289057561 [accessed 2024-05-15]
3. National Cancer Control Programmes: Policies and Managerial Guidelines, 2nd edition: World Health Organization; 2002.
4. Cardoso R, Guo F, Heisser T, et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. Lancet Oncol 2021 Jul;22(7):1002-1013. [doi: 10.1016/S1470-2045(21)00199-6] [Medline: 34048685]
5. Battineni G, Baldoni S, Chintalapudi N, et al. Factors affecting the quality and reliability of online health information. Digit Health 2020;6:2055207620948996. [doi: 10.1177/2055207620948996] [Medline: 32944269]
6. Digital economy and society statistics - households and individuals. Eurostat. 2024. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digital_economy_and_society_statistics_-_households_and_individuals#Pattern_of_internet_use [accessed 2025-08-07]
7. Zhao YC, Zhao M, Song S. Online health information seeking among patients with chronic conditions: integrating the health belief model and social support theory. J Med Internet Res 2022 Nov 2;24(11):e42447. [doi: 10.2196/42447] [Medline: 36322124]
8. Chen YH, Keng CJ, Chen YL. How interaction experience enhances customer engagement in smart speaker devices? The moderation of gendered voice and product smartness. JRIM 2022 Jul 12;16(3):403-419. [doi: 10.1108/JRIM-03-2021-0064]
9. Link E, Baumann E, Linn A, Fahr A, Schulz P, Abuzahra ME. Influencing factors of online health information seeking in selected European countries. EJHC 2021 May 19;2(1):29-55. [doi: 10.47368/ejhc.2021.002]

XSL•FO
RenderX

10. Peterson JS, Swire-Thompson B, Johnson SB. What is the alternative? Responding strategically to cancer misinformation. Future Oncol 2020 Sep;16(25):1883-1888. [doi: 10.2217/fon-2020-0440] [Medline: 32564627]
11. Caulfield T, Marcon AR, Murdoch B, et al. Health misinformation and the power of narrative messaging in the public sphere. Bioethics 2019 Mar 20;2(2):52-60. [doi: 10.7202/1060911ar]
12. Kaňková J, Binder A, Matthes J. Health-related communication of social media influencers: a scoping review. Health Commun 2025 Jun;40(7):1300-1313. [doi: 10.1080/10410236.2024.2397268] [Medline: 39258728]
13. Johnson AR, Longfellow GA, Lee CN, et al. Social media as a platform for cancer care decision-making among women: internet survey-based study on trust, engagement, and preferences. JMIR Cancer 2025 Mar 5;11:e64724. [doi: 10.2196/64724] [Medline: 40053770]
14. Freeman JL, Caldwell PHY, Scott KM. How adolescents trust health information on social media: a systematic review. Acad Pediatr 2023;23(4):703-719. [doi: 10.1016/j.acap.2022.12.011] [Medline: 36581098]
15. Iacobucci G. Vaccination: “fake news” on social media may be harming UK uptake, report warns. BMJ 2019 Jan 23;364:l365. [doi: 10.1136/bmj.l365] [Medline: 30674463]
16. Wilner T, Holton A. Breast cancer prevention and treatment: misinformation on Pinterest, 2018. Am J Public Health 2020 Oct;110(S3):S300-S304. [doi: 10.2105/AJPH.2020.305812] [Medline: 33001732]
17. Bruns A, Burgess J. Researching news discussion on Twitter: new methodologies. J Stud 2012 Oct 1;13(5–6):801-814. [doi: 10.1080/1461670X.2012.664428]
18. Kümpel AS, Karnowski V, Keyling T. News sharing in social media: a review of current research on news sharing users, content, and networks. Social Media + Society 2015 Jul 1;1(2). [doi: 10.1177/2056305115610141]
19. Chereka AA, Butta FW, Demsash AW, et al. Examining internet use for health information seeking and influencing factors among undergraduate health science students in Southwest Ethiopia. Heliyon 2025 Jan 15;11(1):e41545. [doi: 10.1016/j.heliyon.2024.e41545] [Medline: 39845004]
20. Bach RL, Wenz A. Studying health-related internet and mobile device use using web logs and smartphone records. PLOS ONE 2020;15(6):e0234663. [doi: 10.1371/journal.pone.0234663] [Medline: 32530937]
21. Beck F, Richard JB, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128. [doi: 10.2196/jmir.2934] [Medline: 24824164]
22. Siliquini R, Ceruti M, Lovato E, et al. Surfing the internet for health information: an Italian survey on use and population choices. BMC Med Inform Decis Mak 2011 Apr 7;11(1):21. [doi: 10.1186/1472-6947-11-21] [Medline: 21470435]
23. One in two EU citizens look for health information online. Eurostat. 2021. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20210406-1 [accessed 2024-08-19]
24. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libraries J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192]
25. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917. [doi: 10.2196/17917] [Medline: 33978589]
26. Yeung AWK, Kletecka-Pulker M, Eibensteiner F, et al. Implications of Twitter in health-related research: a landscape analysis of the scientific literature. Front Public Health 2021;9:654481. [doi: 10.3389/fpubh.2021.654481] [Medline: 34307273]
27. 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]
28. Weller K, Bruns A, Burgess JE, Mahrt M, Puschmann C. Twitter and Society: Peter Lang; 2014.
29. Meier F, Elsweiler D, Wilson ML. More than liking and bookmarking? Towards understanding Twitter favouriting behaviour. Presented at: Eighth International AAAI Conference on Weblogs and Social Media; Jun 1-4, 2014; Ann Arbor, Michigan USA p. 346-355. [doi: 10.1609/icwsm.v8i1.14541]
30. Kunneman F, Liebrecht C, van den Bosch A. The (un)predictability of emotional hashtags in twitter. 2014 Presented at: Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM); Apr 26, 2014 to Apr 30, 2026; Gothenburg, Sweden p. 26-34. [doi: 10.3115/v1/W14-1304]
31. Huang J, Thornton KM, Efthimiadis EN. Conversational tagging in twitter. 2010 Jun 13 Presented at: HT ’10; Jun 13-16, 2010; Toronto Ontario Canada p. 173-178. [doi: 10.1145/1810617.1810647]
32. Rauschnabel PA, Sheldon P, Herzfeldt E. What motivates users to hashtag on social media? Psychology and Marketing 2019 May;36(5):473-488. [doi: 10.1002/mar.21191]
33. Gaol FL, Maulana A, Matsuo T. News consumption patterns on Twitter: fragmentation study on the online news media network. Heliyon 2020 Oct;6(10):e05169. [doi: 10.1016/j.heliyon.2020.e05169] [Medline: 33083617]
34. Rauchfleisch A, Vogler D, Eisenegger M. Transnational news sharing on social media: measuring and analysing twitter news media repertoires of domestic and foreign audience communities. Digital Journalism 2020 Oct 20;8(9):1206-1230. [doi: 10.1080/21670811.2020.1835511]
35. Qin L, Zhang X, Wu A, et al. Association between social media use and cancer screening awareness and behavior for people without a cancer diagnosis: matched cohort study. J Med Internet Res 2021 Aug 27;23(8):e26395. [doi: 10.2196/26395] [Medline: 34448708]

XSL•FO
RenderX

36. Park HG, Kim YI, Huh WK, Bae S. The association between social media use for health related information and compliance with breast and cervical cancer screenings. Res Rep (Montgomery, Ala) 2020;4:e1-e14. [Medline: 34278179]
37. Himelboim I, Han JY. Cancer talk on twitter: community structure and information sources in breast and prostate cancer social networks. J Health Commun 2014;19(2):210-225. [doi: 10.1080/10810730.2013.811321] [Medline: 24111482]
38. Ola O, Sedig K. Understanding discussions of health issues on Twitter: a visual analytic study. Online J Public Health Inform 2020;12(1):e2. [doi: 10.5210/ojphi.v12i1.10321] [Medline: 32577151]
39. Air M, Roman SA, Yeo H, et al. Outdated and incomplete: a review of thyroid cancer on the world wide web. Thyroid 2007 Mar;17(3):259-265. [doi: 10.1089/thy.2006.0300] [Medline: 17381360]
40. 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]
41. Borges do Nascimento IJ, Pizarro AB, Almeida JM, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022 Sep 1;100(9):544-561. [doi: 10.2471/BLT.21.287654] [Medline: 36062247]
42. Fridman I, Johnson S, Elston Lafata J. Health information and misinformation: a framework to guide research and practice. JMIR Med Educ 2023 Jun 7;9:e38687. [doi: 10.2196/38687] [Medline: 37285192]
43. Call for action: managing the infodemic. World Health Organization. 2020. URL: https://www.who.int/news/item/11-12-2020-call-for-action-managing-the-infodemic [accessed 2024-08-07]
44. Loeb S, Langford AT, Bragg MA, Sherman R, Chan JM. Cancer misinformation on social media. CA Cancer J Clin 2024;74(5):453-464. [doi: 10.3322/caac.21857] [Medline: 38896503]
45. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
46. Ognyanova K, Lazer D, Robertson RE, Wilson C. Misinformation in action: fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. HKS Misinfo Review 2020 Jun 2. [doi: 10.37016/mr-2020-024]
47. Kureyama N, Terada M, Kusudo M, et al. Fact-checking cancer information on social media in Japan: retrospective study using Twitter. JMIR Form Res 2023 Sep 6;7:e49452. [doi: 10.2196/49452] [Medline: 37672310]
48. Bal R, Sinha S, Dutta S, Joshi R, Ghosh S, Dutt R. Analysing the extent of misinformation in cancer related tweets. ICWSM 2020 May 26;14:924-928. [doi: 10.1609/icwsm.v14i1.7359]
49. Yussof I, Ab Muin NF, Mohd M, Hatah E, Mohd Tahir NA, Mohamed Shah N. Breast cancer prevention and treatment misinformation on Twitter: an analysis of two languages. Digit Health 2023;9:20552076231205742. [doi: 10.1177/20552076231205742] [Medline: 37808244]
50. Wardle C, Derakhshan H. Information disorder: toward an interdisciplinary framework for research and policy making. : Council of Europe; 2017 Report No.: DGI(2017)09 URL: https://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf [accessed 2025-10-31]
51. Tan ASL, Lee CJ, Chae J. Exposure to health (mis)information: lagged effects on young adults' health behaviors and potential pathways. J Commun 2015 Aug;65(4):674-698. [doi: 10.1111/jcom.12163]
52. Jansen MP, Hendriks H, Verberne S, Bruijn GJ, Das E. Exploring dialogs on cancer screening: an investigation of online discussions from 2012 until 2023. : OSF; 2025 Jan 6. [doi: 10.17605/OSF.IO/DA3KT]
53. About. Twi-XL. URL: https://twi-xl.humanities.uva.nl/?page_id=88 [accessed 2024-08-08]
54. Twitter's API access changes could mark 'end of an era' in academic research on the platform. Center for an Informed Public, University of Washington. 2023. URL: https://www.cip.uw.edu/2023/02/02/twitters-api-access-changes-academic-research/ [accessed 2024-08-28]
55. Python. Python Software Foundation. 2022. URL: https://www.python.org/ [accessed 2025-08-25]
56. Danilak MM. Langdetect. Python Package Index (PyPI). 2021. URL: https://pypi.org/project/langdetect/ [accessed 2025-02-26]
57. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
58. Emmamich/bachelor-thesis-emma-michielsens. GitHub. 2024. URL: https://github.com/Emmamich/Bachelor-Thesis-Emma-Michielsens [accessed 2026-06-10]
59. Bachmann M. RapidFuzz documentation. RapidFuzz: rapid fuzzy string matching. 2024. URL: rapidfuzz.github.io/RapidFuzz/ [accessed 2026-08-20]
60. Reitz RK. Requests. Python Package Index (PyPI). 2023. URL: https://pypi.org/project/requests/ [accessed 2025-08-21]
61. Kurkowski J. Tldextract: accurately separate a URL's subdomain, domain, and public suffix. Python Package Index (PyPI). 2025. URL: https://pypi.org/project/tldextract/ [accessed 2025-08-21]
62. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
63. Feinstein AR, Cicchetti DV. High agreement but low Kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [doi: 10.1016/0895-4356(90)90158-l] [Medline: 2348207]
64. Dettori JR, Norvell DC. Kappa and beyond: is there agreement? Global Spine J 2020 Jun;10(4):499-501. [doi: 10.1177/2192568220911648] [Medline: 32435572]

XSL•FO
RenderX

65. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276-282. [doi: 10.11613/BM.2012.031]
66. Achtergrond [Web page in Dutch]. Rijksinstituut voor Volksgezondheid en Milieu (RIVM). URL: https://www.rivm.nl/bevolkingsonderzoek-darmkanker/professionals/achtergrond [accessed 2025-10-29]
67. Eijkelboom AH, de Munck L, Lobbes MBI, et al. Impact of the suspension and restart of the Dutch breast cancer screening program on breast cancer incidence and stage during the COVID-19 pandemic. Prev Med 2021 Oct;151:106602. [doi: 10.1016/j.ypmed.2021.106602] [Medline: 34217417]
68. Jansen MP, Stortenbeker I, Hendriks H, Verberne S, de Bruijn GJ, Das E. Unpleasant but effective: newspaper coverage of cancer screening and cancer in the Netherlands from 2010 to 2022. PLOS ONE 2025;20(10):e0334121. [doi: 10.1371/journal.pone.0334121] [Medline: 41124173]
69. Affiliate marketing without borders. Trade Tracker. URL: https://tradetracker.com/ [accessed 2026-06-10]
70. Franssen SJ, van der Wal MF, Jansen P, van Eijsden M. Thinness and overweight in children from Amsterdam: a trend analysis and forecast [Article in Dutch]. Ned Tijdschr Geneeskd 2015;159:A8967. [Medline: 26395567]
71. Vermeer NCA, Liefers GJ, Hoop A, Peeters K. Bevolkingsonderzoek naar darmkanker: zucht of zegen? [Article in Dutch]. Ned Tijdschr Geneeskd 2015;159:a9059 [FREE Full text]
72. Mevius L, Eijsden P, Zaat J. Screening op darmkanker: wat levert het op en wat richt het aan [Article in Dutch].. Ned Tijdschr Geneeskd 2015;159:C2471 [FREE Full text]
73. Hofmann B, Reid L, Carter S, Rogers W. Overdiagnosis: one concept, three perspectives, and a model. Eur J Epidemiol 2021 Apr;36(4):361-366. [doi: 10.1007/s10654-020-00706-4] [Medline: 33428025]
74. Brodersen J, Schwartz LM, Woloshin S. Overdiagnosis: how cancer screening can turn indolent pathology into illness. APMIS 2014 Aug;122(8):683-689. [doi: 10.1111/apm.12278] [Medline: 24862511]
75. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian national breast screening study: randomised screening trial. BMJ 2014 Feb 11;348(feb11 9):g366. [doi: 10.1136/bmj.g366] [Medline: 24519768]
76. Snyder A, Jang S, Nazari IS, et al. Google search volume trends for cancer screening terms during the COVID-19 pandemic. J Med Screen 2021 Jun;28(2):210-212. [doi: 10.1177/0969141321999426] [Medline: 33663240]
77. Van Aelst P, Toth F, Castro L, et al. Does a crisis change news habits? A comparative study of the effects of COVID-19 on news media use in 17 European countries. Digital Journalism 2021 Oct 21;9(9):1208-1238. [doi: 10.1080/21670811.2021.1943481]
78. Becker H, Naaman M, Gravano L. Beyond trending topics: real-world event identification on Twitter. ICWSM 2011;5(1):438-441. [doi: 10.1609/icwsm.v5i1.14146]
79. Matthews MR, Abdulbaki H, Ryan WR, Hackman TG, Farzal Z. Oropharyngeal cancer and the HPV vaccine: analysis of social media content. Laryngoscope 2025 Aug;135(8):2770-2776. [doi: 10.1002/lary.32076] [Medline: 39960222]
80. Biancovilli P, Makszin L, Csongor A. Breast cancer on social media: a quali-quantitative study on the credibility and content type of the most shared news stories. BMC Womens Health 2021 May 15;21(1):202. [doi: 10.1186/s12905-021-01352-y] [Medline: 33992111]
81. Klemm C, Das E, Hartmann T. Swine flu and hype: a systematic review of media dramatization of the H1N1 influenza pandemic. J Risk Res 2016 Jan 2;19(1):1-20. [doi: 10.1080/13669877.2014.923029]
82. Zhang S, Ma F, Liu Y, Pian W. Identifying features of health misinformation on social media sites: an exploratory analysis. Library Hi Tech 2022 Nov 22;40(5):1384-1401. [doi: 10.1108/LHT-09-2020-0242]
83. Strömbäck J, Wikforss Å, Glüer K, Lindholm T, Oscarsson H. Uninformed or misinformed? a review of the conceptual–operational gap between (lack of) knowledge and (mis) perceptions. In: Knowledge Resistance in High-Choice Information Environments: Routledge; 2022:187-206. [doi: 10.4324/9781003111474-10]
84. Angelillo S, Miraglia Del Giudice G, Della Polla G, Angelillo IF. Cervical cancer and screening: knowledge, attitudes, and adherence among university students in Italy. Hum Vaccin Immunother 2026 Dec;22(1):2682535. [doi: 10.1080/21645515.2026.2682535] [Medline: 42246278]

Abbreviations

HPV: human papillomavirus
PA: percent agreement
RQ: research question
WEIRD: Western, educated, industrialized, rich, and democratic
WHO: World Health Organization

XSL•FO
RenderX

Edited by R Cuomo; submitted 06.Jan.2026; peer-reviewed by AWK Yeung, M Solbjor; revised version received 22.Jun.2026; accepted 03.Jul.2026; published 28.Aug.2026.

Please cite as:
Jansen MP, Hendriks H, Verberne S, de Bruijn GJ, Das E
Investigating Online Discussions About Cancer Screening on Twitter (Subsequently Rebranded as X): Corpus Analysis
JMIR Infodemiology 2026;6:e90916
URL: https://infodemiology.jmir.org/2026/1/e90916
doi:10.2196/90916

© Martin-Pieter Jansen, Hanneke Hendriks, Suzan Verberne, Gert-Jan de Bruijn, Enny Das. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 28.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Discourse Around Zyn Nicotine Pouches on Instagram and TikTok: Content Analysis

Arpita Tripathi[1,2], MA, PhD; Beth Hoffman[1], MPH, PhD; Christine Larkin[1], MPH; Piper Narendorf[1], BS; Chaim Kittredge[1], BS; Coltin Kunz[3], BS; Jaime E Sidani[1], MPH, PhD

[1]University of Pittsburgh, 130 De Soto Street, Pittsburgh, PA, United States

[2]Rutgers Institute for Nicotine and Tobacco Studies, Rutgers University, New Brunswick, NJ, United States

[3]VISIMO Inc, Pittsburgh, PA, United States

Corresponding Author:
Arpita Tripathi, MA, PhD
University of Pittsburgh, 130 De Soto Street, Pittsburgh, PA, United States

Abstract

Background: Oral nicotine pouches, such as Zyn, have rapidly grown in popularity in the United States, with sales increasing from 83.2 million cans in 2020 to 385 million cans in 2023. This growth has occurred alongside concerns about youth use. At the same time, Zyn's visibility on social media has also expanded, where youth-targeted content may shape perceptions and influence product uptake.

Objective: This content analysis of Instagram and TikTok Zyn-related posts aimed to (1) examine their sentiment and content, (2) assess youth appeal, (3) identify potential misinformation, and (4) report the most frequently used hashtags in the selected posts as indicators of platform-specific framing and audience targeting.

Methods: In March 2024, we collected 10,502 Instagram posts and 609 TikTok posts with #Zyn. We used a systematically developed codebook to guide the analysis of a random Instagram subsample (n=1200) and all TikTok posts (n=609). Interrater reliability was assessed using a Cohen κ of more than 0.80 and percent agreement.

Results: Many coded posts expressed positive sentiment (n=789/887, 88.9%) and normalized Zyn use through comedic content (n=308/887, 35.2%). Posts revealed youth-targeted themes, including appealing flavors (n=417/887, 87.6%); Zyn usage methods (n=159/887, 18.1%); and associations with sports, athletic, and gym settings (n=70/887, 8%). Both platforms contained posts with potential misinformation. While TikTok featured more influencer-generated content, Instagram showcased more business and commercial content.

Conclusions: Findings suggest that Zyn-related social media content may appeal to youth. Zyn and other oral nicotine pouches are often presented favorably with engaging content, underscoring the need for updated regulatory strategies to address potential misinformation and their appeal to youth.

(JMIR Infodemiology 2026;6:e88825) doi:10.2196/88825

KEYWORDS

oral nicotine pouches; social media; Zyn; nicotine; nicotine and tobacco product

Introduction

Oral nicotine pouches (ONPs) are prefilled, tobacco-free, flavored or nonflavored, dissolvable pouches containing nicotine salts that can be used sublingually [1]. They are one of the least expensive nicotine delivery products and are cheaper than flavored e-cigarettes and cigarettes [2]. Recently, public health experts have raised concerns about the potential health effects of ONPs, considering the known associations between nicotine and neurological and cardiovascular disorders [3], as well as the impact of nicotine pouches on oral health [4]. ONPs may also contain a higher dose of nicotine—an addictive substance—compared to other tobacco-free products and nicotine replacement therapy products [1]. As ONPs are relatively new, nonindustry funded research on their chemical composition and health effects is limited. A recent scoping review found that most studies on nicotine content, pharmacokinetics, and toxicity were industry-funded [1]. One of the few independent studies analyzed 37 ONPs, revealing wide variations in total and free nicotine levels. It suggested that increased pH in some products could lead to blood nicotine levels comparable to snuff or snus [1]. Another recent study found significant dependence among adult ONP users and noted that most users reported at least one adverse event, with mouth lesions and upset stomach being the most frequent [5].

ONPs are the only category of nicotine products for which youth use has not recently declined in the United States [6]. Results from the 2024 National Youth Tobacco Survey found that

XSL•FO
RenderX

approximately 1.8% of middle and high school students were current users of ONPs, an increase from 1.5% the previous year [6]. Preliminary analysis of the 2025 National Youth Tobacco Survey indicates that past-30-day ONP use was 1.7% among middle- and high-school students, which is approximately the same as in 2024 [7]. Another study, conducted in 2021, found that 6.4% of respondents aged 13 - 20 years reported current use of ONPs, and 11.1% reported ever using them [8]. Similarly, ONP use is rising globally, with approximately 5% of youth reporting ever using them in Canada and New Zealand, and 10% of youth reporting ever using them in England in 2024 [9].

Of current adolescent ONP users in the United States, 24.5% reported using Zyn, which was the most popular oral nicotine product [10]. Zyn—owned by Swedish Match, a subsidiary of the tobacco firm Philip Morris International—has the highest market share in the United States, with over 58.5% of all ONP unit sales in 2022 [11]. Zyn's sales increased from 83.2 million cans in 2020 to 385 million cans in 2023, indicating a consistent surge in popularity [12]. The product's popularity in the United States led Philip Morris International to acknowledge a shortage in June 2024. The following month, the company announced a US $600 million investment in a new Zyn production facility [13].

The regulatory status of ONPs is rapidly evolving. In 2022, the US Food and Drug Administration (FDA) announced its regulatory authority over tobacco-free products that contain synthetic nicotine—including ONPs—to ensure that the products are regulated under the same standards as traditional tobacco products [14]. In January 2025, the FDA authorized the marketing of 20 Zyn products [15]. More recently, the FDA launched a pilot program to streamline and increase the efficiency of ONP authorization applications [16], which includes Zyn.

Despite the recent authorizations, Zyn is not yet authorized as a modified-risk tobacco product (MRTP). An MRTP is a nicotine or tobacco product authorized by the FDA to be marketed as a reduced-risk or reduced-harm product when compared to the risks associated with traditional cigarettes. As of 2026, Zyn currently has a pending application with the FDA to be classified as an MRTP, and if granted MRTP approval, it would allow for different risk exposure messaging compared to traditional tobacco products [17,18].

Of particular concern are Zyn's marketing descriptors, including the usage of terms such as "flavor ban approved," "smooth," "chill," and "tobacco-free" [19,20], since such terms can attract new users and promote continued use [21]. Furthermore, 44.4% of all unique ONP advertisements were sponsored by Zyn, with "freedom" as the most common theme (30.7%) [17]. The perception of Zyn as being tobacco-free, as well as its various flavor options, may add to its appeal among youth and young adults; a 2021 survey of US young adults found that 17% of respondents incorrectly believed that ONPs contain neither tobacco nor nicotine [22].

The growing online presence of Zyn's marketing is concerning. A 2024 Ad Watch report indicated that Zyn's digital marketing spending surged from US $2 million in 2022 to US $28 million in 2023 [23]. Campaigns included banner advertisements on the Entertainment and Sports Programming Network, Google mobile apps, and news sites like The Washington Post. Although Zyn's official Facebook and Instagram pages claim no influencer partnerships, "unbranded" pro-Zyn content, often driven by youth culture, has amassed significant views on platforms like Instagram and TikTok [23]. Research also suggests that exposure to nicotine and tobacco product (NTP) content on video-based platforms, including Instagram and TikTok, has been associated with NTP initiation and a positive attitude toward NTPs [24].

Considering that 46% of adolescents are online almost constantly, and Instagram and TikTok are 2 of the most popular platforms used by youth [25], it is important to categorize and analyze Zyn and other ONP-related content to which they may be exposed in order to inform regulatory efforts to protect consumers [26]. Relatedly, recent literature has examined Zyn-related content on social media platforms, including TikTok and Instagram, and identified themes related to the normalization of Zyn use [26], primarily among male individuals [23,26,27]; the portrayal of Zyn as a relatively harmless product [23,26,27]; and its presence in digital pop culture [27,28]. While past studies have conducted content analyses focused on marketing, demographics, health effects, and product types, there has been little examination of misinformation, post sentiment, or the strategies individuals use, such as the use of hashtags, to share Zyn-related posts across social media platforms. Addressing these gaps is important because platform-specific content may shape youth and young adults' perceptions of ONPs [29] or contribute to the spread of misleading or unsupported claims [30]. Thus, this content analysis of both Instagram and TikTok Zyn-related posts aimed to (1) examine their sentiment and content, (2) assess youth appeal, (3) identify potential misinformation, and (4) report the most frequently used hashtags in the selected posts to characterize platform-specific content framing [31].

Methods

Data Collection

All posts from Instagram and TikTok with the hashtag Zyn (ie, #Zyn) were collected on March 3, 2024. The collected posts were originally published between January 29, 2012, and February 3, 2024, on Instagram, and between March 7, 2019, and February 3, 2024, on TikTok. We used #Zyn because Zyn was the primary focus of this study and is the leading ONP brand in the United States. However, we recognize that this single-hashtag strategy may have excluded relevant posts using other nicotine pouch hashtags, slang terms, alternative spellings, brand comparisons, or no hashtag.

TikTok data (609 posts) was gathered using the open-source package "*TiktokPy*," which retrieves popular posts based on TikTok's proprietary algorithm [32]. Instagram data (10,502 posts) was collected via Apify, an online data collection platform that uses its own data collection algorithm [33]. In both cases, the collected data represented content that was publicly available through #Zyn, including profile names and URLs, post content, captions, and engagement metrics such as likes. All additional hashtags used in the posts were extracted for analysis.

XSL•FO
RenderX

The sampling frames in the study differed by platforms. Apify's Instagram hashtag scraper is designed to retrieve public Instagram posts associated with one or more hashtags. In contrast, *TikTokPy* is a Python scraping package (Python Software Foundation) that extracts TikTok data without application programming interface keys. This implies that retrieved posts may reflect TikTok's platform-specific ranking, availability, and algorithmic display at the time of collection. Therefore, the Instagram and TikTok datasets should be interpreted as platform-specific samples of publicly available #Zyn content. These platform-specific differences are an important challenge for social media data collection and reproducibility [34]. Similar data collection approaches have been used in prior public health studies, including studies using NTP data [28,35].

Codebook Development and Coding Procedure

The codebook was developed using a hybrid deductive-inductive coding approach [36]. The lead authors developed an initial codebook for human coding based on prior NTP research [37-39] and a review of 25 Instagram and 25 TikTok posts from the original dataset. To validate the first draft of the codebook, the initial draft was used to code a random subsample of 100 posts (50 from Instagram and 50 from TikTok). All posts in the initial subsample were double coded by trained human coders. To refine the codebook, all disagreements were resolved through an iterative adjudication process conducted over 3 rounds. The final codebook was applied to the remaining posts, which were coded by trained human coders.

The 609 TikTok posts were double-coded by 2 independently working coders, and annotations and memos were written to highlight key themes identified within each post. Any differences in coding were adjudicated by the coders during weekly sessions with the lead author.

For Instagram, considering the large volume of Instagram posts and because manual coding is resource intensive, a random subsample of 1200 posts was selected to be coded by 4 coders. This approach has been recommended for qualitative analysis of large datasets of social media data [40]. The approach also helped balance the feasibility of this study while retaining a large number of posts for content analysis. During coder training and reliability assessment, Instagram posts were double-coded in sequential batches. During this period, interrater reliability was assessed using the Cohen κ and percentage agreement across main categories, including relevance, account type, content characteristics, post purpose, and potential misinformation (see the "Coding Categories" section for codebook category explanations). Once a Cohen κ of greater than 0.8 was achieved for the main categories mentioned above, the remaining Instagram posts were single-coded. In total, 1809 posts were coded between the 2 platforms.

Coding Categories

Each post was first coded for availability, indicating whether the post was accessible at the time, and for English language. Relevant posts, defined as those that focused on Zyn or ONPs, were categorized by account type, post purpose, post content, sentiment, appeal, and potential misinformation. Account type codes included commercial, influencer (≥10,000 followers) [41], athlete, or youth or young adult-appearing. For posts coded as youth or young adult appearing, coders' self-assessment was used to make the decision. Adolescent-appearing individuals were defined as those individuals who appeared to be between the ages of 11 and 17 years, and young adult-appearing individuals were defined as those individuals who appeared to be between the ages of 18 and 20 years. These codes were based on criteria including self-reported age, school or college references, profile biography, captions, and visual appearance. This assessment, based on observable characteristics in posts or user profiles, has been used in prior studies [42]. Furthermore, account type and appearance codes were not mutually exclusive. For example, a post authored by an influencer could also be coded as youth- or young adult–appearing if the individual featured in the post appeared to fall within those age categories.

Posts were also classified by their primary purpose: informational, comedic, or product-focused, with youth-appealing content identified. The purpose category was mutually exclusive. When a post appeared to serve more than 1 purpose, such as being both comedic and product-focused, coders assigned the category that best reflected the post's primary purpose. A post coded as informational included posts that discussed statistics, prevalence, and health effects, among other data, related to the use of Zyn or ONPs. A post coded as product-focused referred to those that emphasized showcasing, displaying, or promoting Zyn or ONPs. All posts that could not be categorized under any of the above categories were coded as "others."

Posts were coded for overall themes, including depicting or suggesting Zyn or ONP use, and sentiment, including pro-Zyn or pro-ONP, anti-Zyn or anti-ONP, or neutral toward these products. Content codes included sports or gym depictions, flavors, personal experiences, and product promotion.

Potential misinformation was flagged for unverifiable claims. Potential misinformation was defined as any explicit or factual claim about ONPs, nicotine, health effects, cessation, NTP safety, and so on, that was either contradicted or could not be verified in the peer-reviewed literature as of November 2024. It was coded in 2 stages. First, the coders applied a broad interpretation of potential misinformation. Later, these flagged posts by the coders were reviewed by senior researchers (redacted for review). Medline and Google Scholar were used to assess claims against peer-reviewed literature; content contradicting or lacking scientific support (as of November 2024) was deemed misinformation. Content codes were not mutually exclusive unless specified otherwise. See Table 1 for the overarching themes used in the codebook and Table 2 for the complete codebook.

XSL•FO
RenderX

Table . Definitions of main themes and their descriptions.

Overarching theme	Description
Relevant	Identifies whether Zyn or oral nicotine pouches was a main topic of the post and whether the post met inclusion criteria for analysis.
Account type	Captures characteristics of the account or individuals featured in the post, including whether the account was commercial, athlete-associated, influencer-associated, or featured adolescents or young adults.
Presence of NTPs[a]	Identifies whether the post displayed or referred to Zyn use, Zyn product imagery, alcohol, or other NTPs.
Sentiment	Captures the overall attitude toward Zyn expressed in the post, including positive or negative sentiment.
Content	Describes the substantive content of the post, including personal experiences, flavor references, athletics, bodybuilding or gym-related content, and unrelated content that used Zyn imagery.
Purpose	Captures the primary purpose or function of the post, such as informing viewers, entertaining through humor, or displaying or discussing the product.
Appeal	Identifies whether the post contained elements that appeared to appeal to specific population groups, including youth and racial, ethnic, sexual, or gender minority groups.
Misinformation and misleading health claims	Captures posts containing potentially false, unsupported, or misleading factual claims about Zyn, nicotine, health effects, safety, or product effects.

[a]NTP: nicotine and tobacco product.

XSL•FO
RenderX

Table . Definitions for all codebook themes and their frequencies.

Content category and code	Definition	Example	TikTok (n=453), n (%)	Instagram (n=434), n (%)
Account type				
Commercial	The post is authored by a supplier, distributor, or marketer of Zyn or ONP[a] products.	A local distributor shows their Zyn stock in the post.	84 (18.5)	207 (47.7)
Athlete	The post is authored by or features a school, college, or professional athlete.	A college athlete discusses their use of Zyn before big games.	5 (1.1)	4 (0.9)
Adolescent appearance	The post features any individual who appears to be between the ages of 11 and 18 years.	A boy who appears to be middle-school age uses Zyn in the video.	12 (2.7)	0 (0)
Young adult appearance	The post features any individual who appears to be between the ages of 18 and 20 years.	A man who appears to be college age uses Zyn in the video.	110 (24.2)	24 (5.5)
Influencer	The post is authored by an account with 10,000 or more followers.	An individual with 16,000 followers makes a post featuring Zyn.	128 (28.3)	58 (13.4)
Presence of nicotine and alcohol products				
Zyn or ONP use	The post displays or suggests the use of Zyn or ONPs.	The post shows an individual placing or keeping Zyn in their mouth.	155 (34.1)	4 (0.9)
Zyn or ONP image	The post displays a Zyn or ONP package, can, pod, or something resembling the product.	The post shows an individual holding a Zyn pack with the product taking up most of the screen.	27 (5.6)	262 (60.4)
Alcohol	The post displays or refers to alcoholic beverages.	The post contains an advertisement for margarita-flavored Zyn.	43 (9.5)	15 (3.5)
Other nicotine and tobacco products	The post displays or refers to another nicotine or tobacco product.	The post displays a pack of cigarettes or nicotine vape alongside Zyn.	23 (5.1)	23 (5.3)
Sentiment				
Pro-Zyn or pro-ONP	Positive sentiment toward Zyn or ONPs is expressed in the post.	A student mentions that using Zyn helps them focus on their work.	409 (90.3)	380 (87.6)
Anti-Zyn or anti-ONP	Negative sentiment toward Zyn or ONPs is expressed in the post.	An adult recommends against Zyn use in the post.	13 (2.9)	4 (0.9)
Content				
Personal or proximal experience	The post contains or describes an individual's personal or proximal experience with Zyn or ONPs.	An athlete reports feeling dizzy and lightheaded the first time they used Zyn.	34 (7.5)	11 (2.5)
Flavorings	The post shows or mentions a flavor of Zyn or ONPs.	A post advertises a Zyn pack that is visibly cool mint flavor.	118 (26.0)	299 (68.9)
Athletics	The post contains content related to sports or athletic activity.	A post references or depicts tennis, basketball, soccer, running, swimming, or similar sports.	31 (6.8)	25 (5.8)

XSL•FO
RenderX

Content category and code	Definition	Example	TikTok (n=453), n (%)	Instagram (n=434), n (%)
Bodybuilding or gym	The post contains content related to gym use, bodybuilding, muscle gain, or fitness routines.	An individual displays a bodybuilding workout and states that Zyn is part of their routine.	11 (2.4)	3 (0.7)
Unrelated content with Zyn	The post uses Zyn or ONP product imagery to promote or discuss unrelated content.	A retailer of nonnicotine products replaces the text on a Zyn tin to advertise their own product.	8 (1.8)	37 (8.5)
Purpose				
Informational	The post provides information about Zyn/ONPs, including statistics, product use, trends, or claims.	A user presents statistics on Zyn use over the past year.	17 (3.7)	19 (4.4)
Comedic	The post uses humor, jokes, memes, or entertainment-based framing related to Zyn or ONPs.	A user makes a joke about Zyn.	185 (40.9)	123 (28.3)
Product-focused	The post displays, discusses, promotes, or centers the Zyn or ONP product.	An individual shows the Zyn product with no further explanation.	35 (7.7)	234 (53.9)
Appeal				
Youth	The post contains elements that may appeal to children or adolescents.	A post displays a Zyn flavor that resembles the flavor of well-known candies.	116 (25.6)	231 (53.2)
Black	The post contains elements that appear to appeal to Black individuals or communities.	A post uses images or themes related to Black identity or empowerment.	0 (0)	0 (0)
Hispanic	The post contains elements that appear to appeal to Hispanic individuals or communities.	A post uses Spanish language or Hispanic cultural imagery.	0 (0)	0 (0)
Sexual or gender minorities	The post contains elements that appear to appeal to sexual or gender minority communities.	A post uses LGBTQ+[b] signifiers such as a pride flag.	0 (0)	0 (0)
Misinformation and misleading health claims				
Potential misinformation	The post contains factual claims about Zyn or ONPs that are contradicted by or could not be verified in peer-reviewed literature.	A user claims that Zyn use can improve eyesight.	21 (4.6)	24 (5.5)

[a]ONP: oral nicotine pouch.

[b]LGBTQ: lesbian, gay, bisexual, transgender/transsexual, queer.

Analysis

Content and Thematic Analysis

Frequencies and percentages were calculated for each code. Following this, the senior authors (redacted for review) independently analyzed a total of 300 posts, consisting of 150 TikTok posts and 150 Instagram posts. For this thematic analysis, each author began by individually reviewing 50 posts (25 from each platform), consulting the coders' notes, and taking their own detailed notes. The authors (redacted for review) then met to discuss and identify emerging themes. This process was repeated, with each author analyzing an additional 25 posts from each platform before reconvening to refine and expand the identified themes. The review process continued until thematic saturation was achieved. Additionally, due to the high number of posts coded for athletic content, youth-targeting, and potential misinformation, the lead authors (redacted for review) conducted a separate in-depth qualitative analysis of all posts within these categories.

Hashtag Analysis

All hashtags used in the captions of the Instagram and TikTok posts coded as relevant were collected and separately processed using R software (version 4.1.2; R Foundation for Statistical

XSL•FO
RenderX

Computing). Text-cleaning was performed on all hashtags on both platforms, including converting text to lowercase and removing duplicates. Later, the frequency of each hashtag was calculated on both platforms. To qualitatively compare hashtags, the lead authors (AT, BH, and JS) independently—and later as a group—reviewed the most frequently occurring hashtags across the 2 platforms and compared them for overlap, platform, and broader thematic patterns.

Ethical Considerations

This study was reviewed by the University of Pittsburgh Human Research Protection Office (STUDY22080079) and was determined not to constitute human participant research because it used publicly available social media data and involved no direct interaction or intervention with individuals. Therefore, informed consent was not required. To protect privacy and confidentiality, no usernames, profile information, or other directly identifying information is reported, and findings are presented in aggregate and/or deidentified form.

Results

Content Analysis

Among all coded posts, 453 out of 609 (74.4%) TikTok posts and 434 out of 1200 (36.2%) Instagram posts were coded as relevant to Zyn or ONPs. All subsequent percentages are reported among relevant coded posts unless otherwise stated.

Among relevant coded posts, 27.0% (n=122) of the TikTok posts and 5.5% (n=24) of the Instagram posts appeared to be posted by youth or young adults. While 48.0% (n=207) of the Instagram accounts were coded as commercial, only 18.5% (n=84) of TikTok accounts were coded as commercial. Additionally, 28.0% (n=128) of TikTok posts and 13.0% (n=58) of Instagram posts were made by influencer accounts. The majority of posts—90.3% (n=409) on TikTok and 87.6% (n=380) on Instagram—showcased pro-Zyn or pro-ONP content. The primary purpose of the posts was to show comedic content, with 40.9% (n=185) on TikTok and 28.3% (n=123) on Instagram, or to showcase products related to Zyn or ONPs, with 7.7% (n=35) on TikTok and 53.2% (n=231) on Instagram. Active Zyn or ONP use was more common on TikTok than Instagram, appearing in 34.1% (n=155) of TikTok posts compared with 1.0% (n=4) of Instagram posts. Close-up images of Zyn or ONP cans or pods were more common on Instagram, appearing in 60.3% (n=262) of Instagram posts compared with 5.6% (n=27) of TikTok posts. Alcohol-related content appeared in 9.5% (n=43) of TikTok posts and 3.5% (n=15) of Instagram posts, while other NTPs appeared in 5.1% (n=23) of TikTok posts and 5.3% (n=23) of Instagram posts. No posts on either platform were coded as appealing specifically to Black, Hispanic, or sexual or gender minority audiences. See Table 2 for descriptives of the main coding categories.

Thematic Analysis

Figure 1 presents illustrative examples of each of the themes discussed in this section.

XSL•FO
RenderX

Figure 1. Example images representing each of the main themes explored in the paper. (A) The image highlights the use of Zyn in gym and athletic settings. (B) The image depicts Zyn flavors associated with popular cartoon characters. (C) The image exemplifies misleading health claims, humorously equating Zyn and Diet Coke to selective serotonin reuptake inhibitors (SSRIs), which are commonly prescribed antidepressants. (D) The image portrays Zyn paired with energy drinks, suggesting their use as substitutes for food before workouts. POV: point of view.



Misleading Health Claims and Misinformation

Of the 24 out of 434 Instagram posts (5.5%) coded for potential misinformation, and on TikTok, 33 out of 453 posts (7.2%) were coded as potential misinformation. These posts were reviewed by senior authors, and it was found that 9 out of 24 (37.5%) Instagram posts and 21 out of 33 (63.6%) TikTok posts contained misinformation. Common misinformation themes promoted Zyn as a smoking or vaping quit aid or claimed unproven health benefits, such as cognitive enhancement or antidepressant effects. Some anti-Zyn posts also contained unverified claims, such as Zyn causing mouth cancer or enamel decay. Others falsely advertised nonexistent Zyn concentrations (eg, 150 mg).

Thematic analysis revealed that misleading health claims were also common, particularly in athletic and youth-targeted posts. These posts implied benefits such as improved performance or weight loss. One Instagram post showed a Zyn-sponsored hockey team outscoring a Gatorade-sponsored team, suggesting that Zyn enhances athletic performance. A TikTok video depicted a man taking Zyn and immediately doubling his bench press record. Weight loss–related posts often used imagery to imply that Zyn or ONPs could serve as a meal replacement, featuring individuals pairing Zyn or ONPs with coffee for breakfast or prioritizing Zyn or ONP purchases over groceries.

Youth Appeal

About 53% (n=231) of the posts on Instagram and 26% (n=116) of the posts on TikTok were coded as appealing to youth. On Instagram, these posts featured flavors of Zyn or ONPs (eg, mint, fruit, and coffee). Some posts also displayed images of Zyn with fake flavors inspired by food items such as Cheetos. Other youth-appealing posts included Zyn cans depicted in various art styles and memes using Zyn for comedic effect. On TikTok, most youth-appealing posts prominently featured popular pop-culture movies, television shows, and celebrities. For example, 1 post used popular animated characters to depict a fictional island with Zyns. Furthermore, posts on TikTok merged the term “Zyn” into other cultural references to create attention-grabbing terms such as Zynbawe, Zyndiana Jones, and Zynley Cup. Finally, a recurring youth-appealing theme involved meme-based posts that linked Zyn or ONPs to childhood nostalgia using candy-related imagery. For example, some posts showed a “same feeling—different age” format, displaying colorful candies captioned “same feeling,” followed by Zyn or ONP cans captioned “different age.” These posts

XSL•FO
RenderX

suggested that Zyn or ONPs evoked feelings similar to those associated with candies, framing Zyn or ONPs through youth-oriented nostalgic imagery.

Normalization of Zyn or ONP Use

A prominent theme that emerged across both platforms was the normalization of Zyn use through the depiction of its integration into everyday life activities. On TikTok, 7.5% (n=34) of posts shared everyday experiences with using Zyn or ONPs, including the use of Zyn or ONPs with meals, drinks, and workouts. One video, titled "boy dinner," featured the use of Zyn alongside beer, ground beef, and sauce, while another showcased Zyn cans as Christmas decorations. On Instagram, 2.5% (n=11) of posts shared everyday experiences with using Zyn or ONPs. For example, commercial accounts reinforced normalization by marketing Zyn-related accessories like can holders, branded t-shirts, and recycling bins. Some posts framed vaping as outdated, using slogans like "Vaping is so 2019, join the revolution."

Use of Zyn or ONP for Athletics and Gym

Zyn was frequently linked to sports like golf, hockey, football, and baseball, with golf being the most mentioned. On TikTok, 6.8% (n=31) of the content was related to sports or athletics, while 2.4% (n=11) was specifically related to the gym. For example, in a TikTok video, a golfer denied using Zyn despite visibly having it in their upper lip, subtly normalizing its use. Another video showed a person placing Zyn pouches in hockey players' mouths mid-game. On Instagram, 5.8% (n=25) of the content was related to sports or athletics, while less than 1% (n=3) was specifically related to the gym. For example, on Instagram, an edited image of a well-known golfer holding a Zyn can promoted the product, and a "Zynley Cup" trophy made from Zyn cans was widely shared. Gym-related posts often framed Zyn as a preworkout product. One Instagram post featured a gym scene with dumbbells, a sugary drink, and a can of Zyn.

Hashtag Analysis

In addition to the hashtag #Zyn, there were a total of 4247 and 6935 additional hashtags used on TikTok and Instagram, respectively. The most frequently used hashtags on Instagram included niche brand names of other ONPs such as Lyft, Nicopods, Swedish Snus, Whitefox, Skruf, Siberia Snus, and Ace SuperWhite. Conversely, the most frequently used hashtags on TikTok were fyp ("for you page"), fy ("for you"), and viral.

Both platforms used hashtags to describe characteristics of Zyn, such as nicotine pouches, snus, tobacco-free, smokeless, nicotine, Nic, or 6 mg (a term that describes a popular strength of Zyn pouches). Posts on both platforms also used informal terms to describe the action of using Zyn, such as "upper decky," "lip/gum pillow," and "upper decky lip cushy." Furthermore, posts on both platforms frequently used wordplays on the term Zyn to create terms such as ZynTok, ZynGang, Zynbabwe, ZynFam, ZynFlavor, ZynUSA, and Zynfluencer.

Finally, a trend on TikTok was observed where sports-related hashtags, such as golf, hockey, golftiktok, and barstoolsports, were frequently used. Notably, "ferda," a term popular among hockey players meaning "for the boys," was also prevalent. Additionally, hashtags reflecting masculine themes or stereotypes, including cheddy, the boys, and frat were identified. Beyond male-specific content, TikTok featured more lifestyle-related hashtags such as collegetok, college life, and beer. See Multimedia Appendix 1 for the full list of hashtags.

Discussion

Principal Findings

This content analysis of Instagram and TikTok posts that used the hashtag #Zyn found that most expressed pro-Zyn or pro-ONP sentiment. While few Instagram posts appeared to have been created by young people, approximately one-quarter of TikTok posts were created by them. The use of comedy or entertainment was common in Zyn or ONP-related posts. Specific themes emerged from this analysis, including the depiction of flavors, a connection to athletics and working out, and the normalization of Zyn or ONP use. The analysis of hashtags was consistent with the content and thematic findings.

Influencer marketing, a focus on Zyn or ONP-imagery, and overwhelmingly pro-Zyn or pro-ONP content were important findings of this study. Considering that influencer marketing significantly impacts consumer behavior among younger audiences [43], the prevalence of pro-Zyn or pro-ONP content shared by influencers may contribute to a favorable image of the product, making it appear more appealing and socially acceptable. While this study does not directly examine audience perception or behavior, it may inform future studies aimed at assessing the impact of influencers and Zyn or ONP imagery on product uptake. Furthermore, the integration of promotional content within influencers' posts can make it difficult for young audiences to recognize the commercial intent, thereby influencing product uptake [44]. Finally, exposure to Zyn imagery may increase the desire to consume ONPs, as indicated by previous research examining behavioral and neural responses to e-cigarette advertising [45].

Several findings from this study may warrant the attention of policymakers. First, although Zyn does not maintain an official presence on TikTok or Instagram [46], this study highlights the potential impact of user-generated content on young people. Over 50% (231/453) of Instagram posts and 26% (116/453) of TikTok posts in this study appealed to youth, with 27% (n=134) posted by young creators. Content often featured flavors, memes, animations, and pop culture trends. Although these findings do not directly measure youth exposure or behavioral effects, they align with research on how visually engaging and culturally relevant content attracts young audiences to novel nicotine products [8]. Likewise, flavors have been a key driver of e-cigarette use among youth [47], highlighting the need to monitor and restrict youth access to flavored Zyn or ONPs. In its authorization of Zyn marketing, the FDA states that it will *impose stringent marketing restrictions for digital, TV, and radio* to reduce youth exposure to marketing, and that the company planned to implement a variety of measures to restrict youth access [15]. However, this study found that much of the Zyn- or ONP-related content on TikTok and Instagram was user-generated, suggesting that it will be available on these

XSL•FO
RenderX

platforms despite mitigation efforts. Considering that exposure to e-cigarette content on social media is reaching youth and is associated with increased susceptibility to the use of these products [48], addressing both user and manufacturer-posted content is crucial.

This study found that the normalization of Zyn or ONP use through its integration into everyday life was commonly observed on both platforms. Posts often portrayed Zyn or ONP as a socially acceptable part of daily life, integrating it into meals, beverages, workouts, and holiday celebrations. Commercial accounts further reinforced this with Zyn or ONP-related merchandise. This relatability of content via normalization may enhance its appeal to viewers and, by extension, may influence the uptake and use of ONPs [8,49]. Furthermore, these findings align with prior research that highlights the role of social media in integrating nicotine products within cultural and recreational settings, thereby reducing harm perceptions and increasing their social acceptability among younger audiences [26,50]. Relatedly, the hashtag analysis found frequent use of terms like ZynTok, ZynGang, and Zynbabwe, reflecting Zyn's popularity and a unique and engaging collective identity for its users. This viral, community-driven culture broadens its reach and further normalizes Zyn use, especially among younger audiences [51].

Misinformation and misleading health claims about Zyn or ONPs were present on both platforms, potentially influencing harm perception. Previous research on e-cigarette content has identified common misinformation themes, including claims of safety, significantly reduced harm compared to cigarettes, unproven nicotine benefits, and exaggerated product risks [37,38]. The Zyn or ONP misinformation themes observed in this study were almost identical, including pro-Zyn misinformation about Zyn or ONPs as a proven quit aid and Zyn or ONPs having unproven health benefits, as well as anti-Zyn or anti-ONP posts about health harms that have yet to be confirmed by research. These findings highlight the need for a comprehensive approach to addressing NTP misinformation rather than focusing on individual products. Additionally, these findings suggest the need for further research examining the presence of misinformation on Instagram and TikTok, as most NTP misinformation research to date has focused on X, formerly known as Twitter.

Moreover, this study extends previous research examining e-cigarette misinformation to include posts that were misleading. In some posts, misinformation was not clearly stated; instead, a combination of text and visual imagery implied that using Zyn or ONPs can provide performance or health benefits, such as improving athletic performance. These findings may offer a springboard for developing machine learning or other automated techniques to categorize misleading NTP posts on visual platforms. Future research can also examine whether viewing misleading posts is associated with susceptibility to Zyn or ONP initiation.

Finally, this study identified key cross-platform differences in Zyn or ONP-related content. The content on TikTok primarily focused on describing the uses and experiences with Zyn or ONPs. This content was diverse and user-generated. In contrast, content on Instagram was mainly driven by commercial accounts that displayed visually attractive images of flavored Zyn or ONPs. Hashtag analysis reinforced these distinctions—Instagram hashtags often included other ONP brands, targeting a brand-focused audience, while TikTok featured broader hashtags like #fyp and #viral, enhancing content discoverability. However, the distribution of hashtags on TikTok was skewed, reflecting generic hashtags such as #fyp being more frequently used than hashtags related to the content of the posts, such as hashtags with sports or fitness. This may indicate that TikTok hashtag practices reflect algorithmic visibility strategies more than the substantive content of the posts. Overall, these findings highlight the need to consider platform-specific dynamics and adopt tailored regulatory strategies.

Limitations

This study has several limitations. First, the sampling strategy limits representativeness and reproducibility. Data collection was restricted to posts containing the hashtag #Zyn, which may have excluded relevant Zyn or ONP-related content that used other hashtags, alternative spellings, or no hashtag. In addition, data were collected on a single day, making this study a cross-sectional snapshot of publicly available #Zyn content rather than an assessment of trends over time.

Second, Instagram coding was limited to a random subsample of 1200 posts from the 10,502 collected Instagram posts because manual content coding of the full dataset was not feasible. Although the subsample was randomly selected, it was not statistically compared with the full Instagram dataset on engagement metrics, posting patterns, or other observable features. Therefore, Instagram findings should be interpreted as descriptive of the coded random subsample.

Third, this study analyzed social media content rather than audience exposure, perceptions, susceptibility, or product use. As a result, the findings cannot determine whether youth viewed the content, how they interpreted it, or whether exposure influenced attitudes or behavior. The results of the study must, therefore, be interpreted as descriptive findings that may aid in hypothesis generation for future studies.

Fourth, qualitative interpretation and manual coding of social media posts can be inherently subjective. To improve consistency and reproducibility, the team used a structured codebook, trained coders, double coding, reliability assessment, and adjudication procedures that have been used in previous studies [37-39]. However, some constructs, including apparent age, youth appeal, and misinformation, required judgment based on observable post or profile features and available scientific evidence.

Fifth, misinformation classification was limited by the evidence available at the time of the review. Posts were coded as misinformation or misleading health claims only when the claims were contradicted by, unsupported by, or could not be verified in peer-reviewed literature available as of November 2024. Some claims may have since been confirmed, refuted, or clarified by newer evidence.

XSL•FO
RenderX

Finally, only posts that were accessible at the time of coding were analyzed, so posts that became unavailable, were deleted, or were made private before coding might have been missed.

Public Health Implications and Conclusions

In the evolving NTP landscape, video platforms offer insights into discourse on emerging products like ONPs. This study provides a cross-sectional description of publicly available #Zyn content on TikTok and Instagram and highlights prevalent positive sentiment toward Zyn or ONPs, including content with youth-appealing features. These findings may inform future tobacco control surveillance and policy discussions and reveal parallels in promotional tactics with e-cigarette marketing [52]. However, because this study did not measure youth exposure and product uptake, the findings should be interpreted as descriptive rather than causal. Overall, the study also underscores the importance of online monitoring and suggests the need to strengthen US online advertising regulations to curb youth-targeted marketing. For example, Australia recently introduced a regulation where ONPs are treated as therapeutic and prescription-only products, and advertising them on any platform, including social media, is banned [53]. Furthermore, Canada recently introduced a regulation for emerging nicotine replacement products, including ONPs, that prohibits the use of advertisements, labels, or packaging that can be appealing to youth. Taken together, these examples suggest that regulations on ONP advertising should be considered not only for paid promotions and advertisements but also for influencer-based marketing and youth-appealing imagery that is prevalent on popular social media platforms [54].

Acknowledgments

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

Funding

This work was supported by a grant from the National Institute on Minority Health and Health Disparities (R01MD018543; principal investigator: JES).

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of hashtags.
[DOCX File, 25 KB - infodemiology_v6i1e88825_app1.docx]

References

1. Travis N, Warner KE, Goniewicz ML, et al. The potential impact of oral nicotine pouches on public health: a scoping review. Nicotine Tob Res 2025 Mar 24;27(4):598-610. [doi: 10.1093/ntr/ntae131] [Medline: 38880491]
2. Ganz O, LaVake M, Hrywna M, King Jensen JL, Delnevo CD. National trends in sales and price for commercial tobacco and nicotine products, 2018-2022. JAMA Netw Open 2024 Mar 4;7(3):e241384. [doi: 10.1001/jamanetworkopen.2024.1384] [Medline: 38451527]
3. Whitehead AK, Erwin AP, Yue X. Nicotine and vascular dysfunction. Acta Physiol (Oxf) 2021 Apr;231(4):e13631. [doi: 10.1111/apha.13631] [Medline: 33595878]
4. Jackson JM, Weke A, Holliday R. Nicotine pouches: a review for the dental team. Br Dent J 2023 Oct;235(8):643-646. [doi: 10.1038/s41415-023-6383-7] [Medline: 37891304]
5. Dowd AN, Thrul J, Czaplicki L, Kennedy RD, Moran MB, Spindle TR. A cross-sectional survey on oral nicotine pouches: characterizing use-motives, topography, dependence levels, and adverse events. Nicotine Tob Res 2024 Jan 22;26(2):245-249. [doi: 10.1093/ntr/ntad179] [Medline: 37712111]
6. Jamal A, Park-Lee E, Birdsey J, et al. Tobacco product use among middle and high school students – National Youth Tobacco Survey, United States, 2024. MMWR Morb Mortal Wkly Rep 2024 Oct 17;73(41):917-924. [doi: 10.15585/mmwr.mm7341a2] [Medline: 39418216]
7. Sun H, Jiang Y, Tattan-Birch H, Fan S, Cox S, Jackson SE. Nicotine and tobacco product use among US middle and high school students, 2024-25. medRXiv. Preprint posted online on Mar 24, 2026. [doi: 10.64898/2026.03.20.26348931]
8. Gaiha SM, Lin C, Lempert LK, Halpern-Felsher B. Use, marketing, and appeal of oral nicotine products among adolescents, young adults, and adults. Addict Behav 2023 May;140:107632. [doi: 10.1016/j.addbeh.2023.107632] [Medline: 36731224]
9. Reid JL, Burkhalter R, Ball J, et al. Awareness and use of oral nicotine pouches among youth and young adults, 2022-2024: repeat cross-sectional surveys in Canada, England, the USA and New Zealand. BMJ Public Health 2025;3(2):e003457. [doi: 10.1136/bmjph-2025-003457] [Medline: 41367545]
10. Liu J, Halpern-Felsher B. Popularity of Zyn and other nicotine pouch brands among U.S. adolescents, young adults, and adults in 2021. Prev Med 2024 Aug;185:108068. [doi: 10.1016/j.ypmed.2024.108068] [Medline: 39002807]

XSL•FO
RenderX

11. Majmundar A, Okitondo C, Xue A, Asare S, Bandi P, Nargis N. Nicotine pouch sales trends in the US by volume and nicotine concentration levels from 2019 to 2022. JAMA Netw Open 2022 Nov 1;5(11):e2242235. [doi: 10.1001/jamanetworkopen.2022.42235] [Medline: 36378312]
12. Big tobacco may have its biggest new hit since cigarettes. CNN. 2024 Jul 27. URL: https://www.cnn.com/2024/07/27/business/zyn-nicotine-philip-morris-tobacco/index.html [accessed 2026-06-04]
13. Philip Morris International to invest over $800 million to meet growing Zyn market. Truth Initiative. URL: https://truthinitiative.org/research-resources/emerging-tobacco-products/philip-morris-international-invest-over-800-million [accessed 2026-06-04]
14. FDA continues to implement law regarding non-tobacco nicotine products. US Food and Drug Administration. 2022. URL: https://www.fda.gov/tobacco-products/ctp-newsroom/fda-continues-implement-law-regarding-non-tobacco-nicotine-products [accessed 2026-06-04]
15. FDA authorizes marketing of 20 ZYN nicotine pouch products after extensive scientific review. US Food and Drug Administration. URL: https://www.fda.gov/news-events/press-announcements/fda-authorizes-marketing-20-zyn-nicotine-pouch-products-after-extensive-scientific-review [accessed 2026-06-04]
16. FDA launches program to more efficiently review nicotine pouch applications. US Food and Drug Administration. 2025 Sep 18. URL: https://www.fda.gov/tobacco-products/ctp-newsroom/fda-launches-program-more-efficiently-review-nicotine-pouch-applications [accessed 2026-06-04]
17. Duan Z, Henriksen L, Vallone D, et al. Nicotine pouch marketing strategies in the USA: an analysis of Zyn, On! and Velo. Tob Control 2024 Feb 20;33(2):154-163. [doi: 10.1136/tc-2022-057360] [Medline: 35817549]
18. Modified risk tobacco products. US Food and Drugs Administration. URL: https://tinyurl.com/3nsecasf [accessed 2026-06-04]
19. Tackett AP, Barrington-Trimis JL, Leventhal AM. "Flavour ban approved": new marketing strategies from tobacco-free nicotine pouch maker Zyn. Tob Control 2023 Apr;32(e1):e134-e135. [doi: 10.1136/tobaccocontrol-2021-057222] [Medline: 35459750]
20. Czaplicki L, Tfayli D, Spindle TR, et al. Content analysis of marketing features in US nicotine pouch ads from 2021 to 2023. Tob Control 2024 Dec 3. [doi: 10.1136/tc-2024-059010] [Medline: 39627024]
21. Choi WS, Ahluwalia JS, Harris KJ, Okuyemi K. Progression to established smoking: the influence of tobacco marketing. Am J Prev Med 2002 May;22(4):228-233. [doi: 10.1016/s0749-3797(02)00420-8] [Medline: 11988378]
22. Morean ME, Bold KW, Davis DR, Kong G, Krishnan-Sarin S, Camenga DR. Does it come from tobacco? Young adults' interpretations of the term "tobacco-free nicotine" in a cross-sectional national survey sample. PLoS One 2022;17(5):e0268464. [doi: 10.1371/journal.pone.0268464] [Medline: 35559988]
23. Ganz O, Barnwell PV, Hrywna M, Donaldson SI, Allem JP, Delnevo CD. Marketing ZYN: examining branded and unbranded content. Tob Control 2024 Jul. [doi: 10.1136/tc-2024-058766]
24. Vassey J, Vogel EA, Unger JB. US adolescents' response to nicotine warning labels in influencer e-cigarette marketing social media posts: a survey-based randomised between-subject experiment. Tob Control 2025 Oct 3;34(5):609-617. [doi: 10.1136/tc-2023-058534] [Medline: 38749676]
25. Teens, social media and technology 2023. : Pew Research Center; 2023 URL: https://www.pewresearch.org/wp-content/uploads/sites/20/2023/12/PI_2023.12.11-Teens-Social-Media-Tech_FINAL.pdf [accessed 2026-06-04]
26. Donaldson SI, La Capria K, DeJesus A, Ganz O, Delnevo CD, Allem JP. Describing ZYN-related content on TikTok: content analysis. Nicotine Tob Res 2026 Jan 22;28(2):277-281. [doi: 10.1093/ntr/ntaf016] [Medline: 39823221]
27. Mand A, Fonteyne K, Struik L. Examining how oral nicotine pouches are trending on TikTok: a qualitative descriptive study. JMIR Form Res 2025 Nov 14;9:e73032. [doi: 10.2196/73032] [Medline: 41236804]
28. Amin S, Buente W, Duque JD, et al. Exploring #Zyn-related video content posted by Zynfluencers on TikTok. Subst Use Misuse 2026 Feb 16:1-9. [doi: 10.1080/10826084.2026.2624783] [Medline: 41699783]
29. Donaldson SI, Dormanesh A, Perez C, Majmundar A, Allem JP. Association between exposure to tobacco content on social media and tobacco use: a systematic review and meta-analysis. JAMA Pediatr 2022 Sep 1;176(9):878-885. [doi: 10.1001/jamapediatrics.2022.2223] [Medline: 35816331]
30. Tan ASL, Bigman CA. Misinformation about commercial tobacco products on social media-implications and research opportunities for reducing tobacco-related health disparities. Am J Public Health 2020 Oct;110(S3):S281-S283. [doi: 10.2105/AJPH.2020.305910] [Medline: 33001728]
31. Erz A, Marder B, Osadchaya E. Hashtags: motivational drivers, their use, and differences between influencers and followers. Comput Human Behav 2018 Dec;89:48-60. [doi: 10.1016/j.chb.2018.07.030]
32. Newton R. Tiktokapipy 0.2.4.post2. PyPI. URL: https://pypi.org/project/tiktokapipy/ [accessed 2026-06-04]
33. About Apify. Apify. 2015. URL: https://apify.com/about [accessed 2026-06-04]
34. Heft A, Buehling K, Zhang X, Schindler J, Milzner M. Challenges of and approaches to data collection across platforms and time: conspiracy-related digital traces as examples of political contention. J Inf Technol Politics 2024 Jul 2;21(3):323-339. [doi: 10.1080/19331681.2023.2250779]
35. Dobbs PD, Schisler ED, McCormick C. #Discreetshipping: selling E-cigarettes on TikTok. Nicotine Tob Res 2025 Mar 24;27(4):748-752. [doi: 10.1093/ntr/ntae081] [Medline: 38778482]

XSL•FO
RenderX

36. Swain J. A hybrid approach to thematic analysis in qualitative research: using a practical example. In: Sage Research Methods Cases Part 2: SAGE Publications Ltd; 2018. [doi: 10.4135/9781526435477]
37. Sidani JE, Hoffman BL, Colditz JB, et al. E-cigarette-related nicotine misinformation on social media. Subst Use Misuse 2022;57(4):588-594. [doi: 10.1080/10826084.2022.2026963] [Medline: 35068338]
38. Sidani JE, Colditz JB, Barrett EL, Chu KH, James AE, Primack BA. JUUL on Twitter: analyzing tweets about use of a new nicotine delivery system. J Sch Health 2020 Feb;90(2):135-142. [doi: 10.1111/josh.12858] [Medline: 31828791]
39. Sidani JE, Colditz JB, Barrett EL, et al. I wake up and hit the JUUL: analyzing Twitter for JUUL nicotine effects and dependence. Drug Alcohol Depend 2019 Nov 1;204:107500. [doi: 10.1016/j.drugalcdep.2019.06.005] [Medline: 31499242]
40. Pretorius K. A simple and systematic approach to qualitative data extraction from social media for novice health care researchers: tutorial. JMIR Form Res 2024 Jul 9;8:e54407. [doi: 10.2196/54407] [Medline: 38980712]
41. Kresovich A, Norris AH, Carter CC, Kim Y, Kostygina G, Emery SL. Deciphering influence on social media: a comparative analysis of influential account detection metrics in the context of tobacco promotion. Soc Media Soc 2024 Jan;10(1). [doi: 10.1177/20563051231224268] [Medline: 39575100]
42. Step MM, Bracken CC, Trapl ES, Flocke SA. User and content characteristics of public tweets referencing little cigars. Am J Health Behav 2016 Jan;40(1):38-47. [doi: 10.5993/AJHB.40.1.5] [Medline: 26685812]
43. De Veirman M, Cauberghe V, Hudders L. Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. Int J Advert 2017 Sep 3;36(5):798-828. [doi: 10.1080/02650487.2017.1348035]
44. Fernández-Gómez E, Neira Placer P, Feijoo Fernández B. New mobile advertising formats targeting young audiences: an analysis of advertainment and influencers' role in perception and understanding. Humanit Soc Sci Commun 2024 Nov;11(1):1488. [doi: 10.1057/s41599-024-04003-3]
45. Chen Y, Fowler CH, Papa VB, et al. Adolescents' behavioral and neural responses to e-cigarette advertising. Addict Biol 2018 Mar;23(2):761-771. [doi: 10.1111/adb.12510] [Medline: 28401670]
46. Albert J. Zyn nicotine pouches gain popularity among youth through social media. Experts explain rising concerns. CBS News. 2024 Jun 3. URL: https://www.cbsnews.com/baltimore/news/zyn-influencers-blamed-for-enticing-youth-on-social-media-partake-in-unauthorized-nicotine-pouches/ [accessed 2026-06-04]
47. Pepper JK, Ribisl KM, Brewer NT. Adolescents' interest in trying flavoured e-cigarettes. Tob Control 2016 Nov;25(Suppl 2):ii62-ii66. [doi: 10.1136/tobaccocontrol-2016-053174] [Medline: 27633762]
48. Zheng X, Li W, Li R, Yang M, Lin HC. Exposure to user-generated e-cigarette content on social media associated with greater vulnerability to e-cigarette use among youth non-users. Addict Behav 2024 Sep;156:108061. [doi: 10.1016/j.addbeh.2024.108061] [Medline: 38744213]
49. Lovato C, Linn G, Stead LF, Best A. Impact of tobacco advertising and promotion on increasing adolescent smoking behaviours. Cochrane Database Syst Rev 2003(4):CD003439. [doi: 10.1002/14651858.CD003439] [Medline: 14583977]
50. Jung S, Murthy D, Bateineh BS, Loukas A, Wilkinson AV. The normalization of vaping on TikTok using computer vision, natural language processing, and qualitative thematic analysis: mixed methods study. J Med Internet Res 2024 Sep 11;26:e55591. [doi: 10.2196/55591] [Medline: 39259963]
51. Barron ATJ, Bollen J. Quantifying collective identity online from self-defining hashtags. Sci Rep 2022 Sep 3;12(1):15044. [doi: 10.1038/s41598-022-19181-w] [Medline: 36057691]
52. Lee J, Suttiratana SC, Sen I, Kong G. E-cigarette marketing on social media: a scoping review. Curr Addict Rep 2023 Mar;10(1):29-37. [doi: 10.1007/s40429-022-00463-2] [Medline: 41001189]
53. Nicotine pouches. Australian Government, Department of Health, Disability, and Ageing. URL: https://www.tga.gov.au/products/unapproved-therapeutic-goods/vaping-hub/nicotine-pouches [accessed 2026-06-04]
54. Only use authorized nicotine pouches as directed, and do not use unauthorized nicotine pouches. Government of Canada. URL: https://recalls-rappels.canada.ca/en/alert-recall/only-use-authorized-nicotine-pouches-directed-and-do-not-use-unauthorized-nicotine [accessed 2026-06-04]

Abbreviations

FDA: US Food and Drug Administration
MRTP: modified-risk tobacco product
NTP: nicotine and tobacco product
ONP: oral nicotine pouch

XSL•FO
RenderX

Edited by T Mackey; submitted 02.Dec.2025; peer-reviewed by J Jancey, K LaCapria, W Qin; revised version received 25.May.2026; accepted 26.May.2026; published 23.Jul.2026.

Please cite as:
Tripathi A, Hoffman B, Larkin C, Narendorf P, Kittredge C, Kunz C, Sidani JE
Exploring the Discourse Around Zyn Nicotine Pouches on Instagram and TikTok: Content Analysis
JMIR Infodemiology 2026;6:e88825
URL: https://infodemiology.jmir.org/2026/1/e88825
doi:10.2196/88825

© Arpita Tripathi, Beth Hoffman, Christine Larkin, Piper Narendorf, Chaim Kittredge, Coltin Kunz, Jaime E Sidani. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 23.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Investigation of Differences in Tobacco Use Language Between Groups: Corpus-Assisted Analysis

Iona Fitzpatrick[1], PhD; Xinmei Sun[2], MA

[1]Tobacco Control Research Group, Department of Health, University of Bath, Claverton Down, Bath, United Kingdom

[2]Linguistics and the English Language, School of Social Sciences, Lancaster University, Lancaster, United Kingdom

Corresponding Author:
Iona Fitzpatrick, PhD
Tobacco Control Research Group, Department of Health, University of Bath, Claverton Down, Bath, United Kingdom

Abstract

Background: The variation of language concerning tobacco products and tobacco use is known to impact the understanding of related risks and influence behaviors including use uptake and product cessation. Transnational tobacco companies can use such complexities to change the acceptability of tobacco use and influence public understanding of related risks. These changes, in turn, impact tobacco use behaviors. Looking at variations in the language used by different groups can therefore offer helpful insights into tobacco use cultures and make imbalances of information between groups plain. This paper examines the language of tobacco use, specifically smoking, across a sample of health organizations (the National Health Service, the World Health Organization, the National Institute for Health and Care Excellence, and the Centers for Disease Control and Prevention), the tobacco industry (British American Tobacco and Philip Morris International), and in "general English."

Objective: Through corpus-assisted analysis, the study aims to illustrate differences in tobacco use and tobacco user characterizations by 2 transnational tobacco companies, health organizations, and users of general English. The study assesses the possible implications of these differences for public health.

Methods: We queried 4 bodies of text (corpora) from 3 different groups; 2 of these corpora were preexisting corpora of general spoken and written English in the United Kingdom and the United States. The remaining 2 sampled tobacco-related documents from 2 transnational tobacco companies and from the United Kingdom and international organizations with a focus on health between 2003 and 2023. The 2 sampled corpora contained 1355 documents and 10,023,538 words. We used the corpus analysis software LancsBox (Lancaster University) to identify variations in characterizations of tobacco users and tobacco use behaviors between these groups, using the stems "smoker*" and "smok*."

Results: Frequency and collocation analysis showed clear differences in how the 3 groups described smokers and smoking, with only limited overlap in the terms they used. Only 7 of 23 unique categorizations of "smoker*" were shared. There was a significant association (P<.001) between individual corpora and singular or plural noun forms. Health organization texts more often used clinical and population-focused classifications, whereas tobacco industry texts more often framed smokers as consumers or people involved in legal disputes. General English corpora showed the widest range of labels for tobacco users.

Conclusions: While there was some overlap in terminology used between corpora, the most common categorizations in each corpus were highly varied, showing very little shared language between groups in their descriptions of either tobacco users or use behaviors. This variance indicates that these groups may not share the same sense-making resources related to tobacco use, which may render information flows around tobacco vulnerable to distortion.

(JMIR Infodemiology 2026;6:e86593) doi:10.2196/86593

KEYWORDS

corpus linguistics; health information; tobacco use; commercial determinants of health; health information

Introduction

Tobacco is a leading cause of preventable death globally, killing more than 8 million people annually [1]. Despite sustained international efforts to curtail the manufacture, sale, and use of tobacco, tobacco use is not in universal decline [2]. The rapidly evolving landscape of tobacco and nicotine products, including e-cigarettes, heated tobacco products, and nicotine pouches, presents an ongoing challenge to policymakers, regulators [3], and consumers [4] alike, who are faced with an ever-expanding range of nicotine and tobacco products and product-based information.

The volume of available products and the proliferation of brands driven by the tobacco industry are known to confuse and mislead consumers [3,5] and contribute to the spread of misinformation regarding usage risks [6]. At the same time, language relating

XSL•FO
RenderX

to tobacco varies significantly across user categories—for example, between "non-smokers" and "never-smokers" in their personal narratives, including descriptions of risk and addiction associated with product use [7]. Some features of newer nicotine and tobacco products, such as flavorings, connect nicotine and tobacco use to the relatively innocuous consumption of sweets [8]. It is through practices like this that what it means to use tobacco and to be a "smoker" is increasingly contested [9]. Different groups also vary in their understandings of risk and in their ability to overcome structural constraints in their tobacco use behaviors [5]. Individuals living with mental ill health, for example, are especially vulnerable to misconceptions of harm relating to nicotine and tobacco product use [10].

Uncertainty surrounding nicotine and tobacco products, as well as the risks associated with their use, is not exclusively a problem for consumers. Efforts have been made to shift language away from terms that are imprecise, inherently negative, or derisive [11,12]. In public health and scientific communication, for example, the rise in calls for "person-first" language is notable [13,14], but the route to consensus is not clear. Some alternatively advocate for a focus on personal agency in reporting on tobacco use [15]. The shift to person-first language has not been universally adopted and has come under criticism for being overly simplistic and for perpetuating stigma [16].

Health-related language used by the scientific community and the public does not consistently overlap, sometimes relying on outdated conceptualizations of ill-health [17] and tobacco use [18]. Policymakers and regulators share in these dilemmas. Lempert et al [19] highlighted disparate, overly specific definitions as a problem for effective regulation as early as 2014. It is not surprising, then, that existing research has highlighted a "jingle-jangle" problem [11] and describes the muddled terminological landscape as a hurdle in "attempting to integrate evidence internationally" [20].

Guidance published in 2024 by Maddox et al [21] suggests that better alignment of language on the part of public health could help counter the narratives of powerful industries more effectively, including by calling out "co-opted and misleading terms and phrases" and through the use of "person-centered language." Both of these strategies, they argue, could serve to help in "disarming" harmful industries like the tobacco industry.

As tobacco use is both an individual-level behavior and a socially shaped practice [22], intertwined with cultural values and expressions of identity, we make use of corpus-assisted methods to provide a quantitative evidence base for the examination of tobacco use discourse [23]. Previous applications of these methods have highlighted the importance of linguistic choices in the negotiation of responsibility [24] and the cultivation of stigma [25] in the context of health communication.

Our querying of 2 tailor-made tobacco-focused corpora and 2 ready-made corpora of general English explores and clarifies the linguistic variations surrounding tobacco use status and tobacco use behaviors among groups (the tobacco industry, health organizations, and users of general English).

Methods

Overview

We applied methods from corpus linguistics to investigate language variance between three preselected groups with converging interests in tobacco control. The 2 largest transnational tobacco companies, British American Tobacco (BAT) and Philip Morris International (PMI) [26], were selected to sample tobacco industry language. We selected health organizations from the United Kingdom, the United States, and internationally to sample institutional health communication: the National Health Service (NHS), the National Institute for Health and Care Excellence (NICE), the Centers for Disease Control and Prevention (CDC), and the World Health Organization (WHO). These English-language samples matched the English-language balanced corpora we chose: the 27 British National Corpus 2014 (BNC2014) [27,28] and the Corpus of Contemporary American English (COCA) [29]. We used both the written and spoken components of the 2 corpora, spanning genres such as informal speech (transcribed to text), newspapers, fiction, television scripts, and social media. To compile new bodies of language for querying, we constructed detailed search queries for the tobacco industry and health organization corpora.

Setting the Frame and Searching for Sources

The time frame for data collection was set between January 1, 2003, and December 31, 2023. The period from 2003 to 2023 was marked by huge upheaval in both the tobacco consumer product landscape and information ecosystems. In May 2003, member states of the WHO adopted the WHO Framework Convention on Tobacco Control. Although it did not enter into force until 2005, the adoption marked a concerted effort "for the promotion of public health" on a global scale. Simultaneously, a consistent consolidation of tobacco companies meant that the tobacco industry had also undergone some major shifts—from a number of small and medium enterprises to just a few global powerhouses [18], of which PMI and BAT are the largest.

The 2 ready-made, balanced corpora of English—BNC2014 and COCA—were chosen for the insights they might offer into the way a "general" English-speaking public talks about tobacco products and tobacco use behaviors. The BNC2014 is a 100-million-word collection representing British English, with texts dating between 2007 and 2020, and the COCA is a 1-billion-word collection representing American English, with texts dating between 1990 and 2019.

For feasibility purposes, the construction of the Tobacco Industry Corpus (TIC) was restricted to BAT and PMI, the 2 largest transnational companies based on 2021 market value [26]. The search parameters for the TIC included web pages and documents published by BAT and PMI during the set time frame. The search query was designed to capture content containing multiple terms related to smoking status and smoking behavior. Asterisks indicate "wildcard" searches, which capture various word endings following a given stem:

smok* OR quit* OR cessation OR switch* OR vap* OR experiment* OR relaps* OR "us* NRT*" OR "dual" OR addict*

XSL•FO
RenderX

Table 1 provides examples of terms captured by different parts of the search query. We conducted three separate searches: (1) a Google site search, (2) using the search function within the two tobacco company websites, and (3) searching web archive tools—the Internet Archive [30] and Archive Today [31]—to collect backdated content not captured in searches one or two. All documents were available in the public domain.

Table . Example terms captured by wildcard search queries[a].

Search term	Examples of terms captured
smok*	smoker, smokers, smoke, smoking, smoked
quit*	quitter, quit, quits, quitting
switch*	switcher, switch, switching, switched
vap*	vaper, vape, vaping, vaped
experiment*	experimenter, experiment, experimentation
relaps*	relapse, relapsed, relapsing
us* NRT*[b]	use NRT, using NRT, using NRTs
dual	dual users, dual use, dual usage, dual e-vapor combustible users
addict*	addict, addicted, addiction

[a]Asterisks indicate "wildcard" searches, which capture various word endings following a given stem.

[b]NRT: nicotine replacement therapy.

The search parameters for the Health Organizations Corpus (HOC) were webpages and documents published by 4 major health organizations: NICE, NHS, CDC, and WHO. For the HOC, we added the following to the previous search query to refine the content returned to be about tobacco:

AND (tobacco OR smok* OR vap*)

The study conducted two searches: (1) a Google site search of the named organizations and (2) the WHO Institutional Repository for Information Sharing [32]. The study excluded workbooks, research papers, replacement guidelines, and any draft documents.

Preprocessing and Compiling Corpora

When compiling both corpora, lists of returned results were saved as structured CSV files containing metadata, which included URLs, webpage headings, and publication dates. This list was manually filtered to exclude job advertisements, pages spotlighting named individuals, and pages or documents in which search terms only appeared in legal disclaimers. Region-specific or geographically local content pages were also excluded, as our focus was on organizational rather than local characterizations. Documents or webpages consisting mainly of figures or videos were also excluded from both the TIC and the HOC for reasons of feasibility. Descriptions of smoking status and smoking behavior were our primary focus, so we manually filtered the data files to ensure documents focusing on individuals (profiles or personal stories) were not included in the corpora. The TIC includes multiple document types directed at different audiences, including investors and consumers. This was intentional, as the aim was to capture the tobacco industry's discourse at the organizational level rather than to compare communication strategies across internal genres.

After manual filtering and the application of the exclusions listed above, we used Python (Python Software Foundation) scripts to process the CSV files. The scripts included routines for downloading linked PDF documents, extracting and cleaning text from PDF documents using the *pdfminer* library [33], and exporting web-based content as individual plain text (.txt) files. Annotations of page numbers were included in the files to preserve cross-references to the original PDF documents. The author and publication year were included in the given file names. The sources used to compile the corpora are available via the University of Bath Research Data Archive [34]. Once processed, corpus documents were imported into LancsBox using automatic part-of-speech tagging, with LancsBox (Lancaster University) default settings applied.

Table 2 shows the composition of the 2 tailor-made corpora, the TIC and the HOC.

XSL•FO
RenderX

Table . Composition of tailor-made corpora.

Organization and document type	Files	Unique words (types)	Total words (tokens)
TIC[a]			
BAT[b]			
Annual and sustainability reports	40	32,599	2,893,751
Webpages	170	8306	127,129
Other documents	35	19,078	1,123,665
Total	245	59,983	4,144,545
PMI[d]			
Annual and sustainability reports	24	20,393	1,415,877
Webpages	406	15,031	457,267
Other documents	102	19,569	629,801
Total	532	54,993	2,502,945
Corpus total	777	114,976	6,647,490
HOC[e]			
NICE[f]			
Documents	15	5966	94,513
Webpages	20	2254	21,422
Total	35	6641	115,935
NHS[g]			
Documents	1	2173	11,186
Webpages	23	1861	13,790
Total	24	3189	24,976
World Health Organization			
Documents	124	69,175	2,428,207
Webpages	120	7239	81,256
Total	244	69,947	2,509,463
CDC[h]			
Documents	69	33,757	583,435
Webpages	206	9406	124,628
Total	275	37,012	725,674
Corpus total	578	116,789	3,376,048

[a]TIC: Tobacco Industry Corpus.

[b]BAT: British American Tobacco.

[c]PMI: Philip Morris International.

[d]HOC: Health Organizations Corpus.

[e]NICE: National Institute for Health and Care Excellence.

[f]NHS: National Health Service.

[g]CDC: Centers for Disease Control and Prevention.

Analytical Approach

The data were analyzed using a corpus-assisted discourse approach, which combines quantitative corpus techniques with qualitative discourse interpretation to examine patterns in large collections of text [35]. In practical terms, statistical measures were used to identify potentially meaningful patterns in how people talked about tobacco use status and tobacco use behavior,

XSL•FO
RenderX

and then those patterns were inspected in their immediate context to interpret what they were doing in the discourse.

The analysis drew on 2 complementary techniques—collocation analysis and concordance analysis. For collocation analysis, we examined which words occurred unusually often near a target term to identify recurring associations in the corpus. Collocational strength was measured through mutual information using LancsBox (TIC, HOC, and BNC2014) and English-corpora.org (COCA). Collocations were calculated within a window of 5 words to the left and 5 words to the right of the target term; this is the norm in corpus research and has been shown to be productive in corpus-based discourse studies [36]. A minimum raw co-occurrence frequency of 10 was used to keep the number of collocates manageable for detailed qualitative analysis. Common function words (eg, the, and, of) were excluded using the Natural Language Toolkit's stop-word list [37]. We then used concordancing to interpret these quantitative patterns by viewing every occurrence of a search term in its immediate context, supporting qualitative examination and interpretation of differences and similarities in discourse.

Ethical Considerations

Patient engagement or involvement was not applicable or relevant to the project, as it made use of documents available in the public domain. Institutional ethical clearance was provided by the University of Bath Biomedical Ethics Committee with approval EP18/19 004. The research was conducted in accordance with the guidelines set forth in the Declaration of Helsinki.

Results

Frequencies

The singular form "smoker" and the plural form "smokers" were analyzed separately to assess whether any of the corpora showed strong habits of reference. Table 3 presents the raw and relative frequencies of the terms "smoker*" as a noun and "smok*" as a verb in each corpus.

Table . Frequency of "smoker* N*" and "smok* V*" in each corpus[a].

	Raw and relative (per million) frequency			
	TIC[b]	HOC[c]	BNC2014[d]	COCA[e]
smoker* N*				
smokers N*	4924 (7.39)	7577 (22.44)	458 (4.48)	4120 (4.15)
smoker N*	936 (1.40)	648 (1.92)	206 (2.01)	2355 (2.37)
smok* V*				
smoke V*	2042 (3.06)	2749 (8.14)	1138 (11.12)	43,675 (43.98)
smoking V*	1126 (1.69)	1696 (5.02)	1135 (11.09)	30,393 (30.61)
smoked V*	272 (0.41)	1330 (3.94)	475 (4.64)	8866 (8.93)
smokes V*	17 (0.03)	130 (0.39)	131 (1.28)	2278 (2.29)

[a]Asterisks indicate "wildcard" searches, which capture various word endings following a given stem, while the asterisks after V and N indicate "smart searches" where the term is tagged as either a Noun (N*) or VERB (V*).

[b]TIC: Tobacco Industry Corpus.

[c]HOC: Health Organizations Corpus.

[d]BNC2014: British National Corpus 2014.

[e]COCA: Corpus of Contemporary American English.

A chi-square test showed a significant association between corpus and noun form (singular vs plural; n=21,224; χ^2_3=1982.01; P<.001; Cramér's V=0.31), indicating a large effect. Bonferroni-adjusted pairwise comparisons (α=.0083) showed significant differences between all corpus pairs (all P≤.007). HOC had the highest proportion of plural forms, with the largest pairwise difference observed between HOC and COCA (ϕ=0.35). By contrast, the difference between BNC2014 and COCA, although statistically significant, was negligible in magnitude (ϕ=0.03). The predominance of plural forms in the HOC constructs smoking as a population-level phenomenon, while comparatively greater use of singular forms elsewhere individualizes the smoker.

The distribution of verb forms also differed significantly across corpora (n=97,453; χ^2_9=1173.18; P<.001); however, the effect size was small (Cramér's V=0.06). TIC showed the highest proportion of the base form "smoke," whereas HOC favored the past tense "smoked," consistent with the retrospective orientation of clinical and epidemiological reporting.

Collocation

To explore how smoking status and behavior were constructed across corpora, we investigated the semantic themes of collocates of "smoker*" as a noun and "smok*" as a verb using the complex queries "smoker* N*" and "smok* V*." Themes were generated through an iterative examination of concordance lines of the identified collocates, with provisional categories developed, reviewed, and refined throughout the analytic process. Table 4 shows the top collocates of "smoker*N*."

XSL•FO
RenderX

Table . Top lexical collocates of "smoker* N*" (ranked by mutual information [MI] score) and their themes[a].

Theme	Top 20 collocates (MI score) sorted in observed themes			
	TIC[b]	HOC[c]	BNC2014[d]	COCA[e]
Smoker* specific categorization	• deceased (9.9) • adult (9.6) • alive (9.5) • ASU30[f] (9.5) • former (8.7)	• some-day (9.4) • former (8.8) • never (8.3) • nondaily (8.2) • cigarette-only (8.2) • occasional (8.1) • heavier (8.0) • african-american (8.0) • non (8.0) • novice (7.9) • current (7.6) • ex-smokers (7.5) • persistent (7.4)	• ex-smokers (16.8) • non-smokers (15.8) • smoker (13.1) • smokers (12.7) • non (9.5) • heavy (9.3) • current (7.5) • former (7.1) • never (5.6)	• nondaily (13.2) • two-pack-a-day (12.8) • ex-smokers (12.4) • nonsmokers (12.1) • non-smokers (11.7) • late-onset (10.6) • nonsmoker (10.2) • smokers (9.7) • early-onset (9.6) • smoker (9.0)
Actors	• heir (10.1) • dependents (9.3)	• sexes (7.9) • quitters (7.3)	—[g]	• insley (12.2) • drinker (8.8) • drinkers (8.4)
Products	• ultra-low-price (10.0)	• non-menthol (7.5)	• tobacco (11.0) • cigarettes (10.8) • cigarette (9.6)	• e-cigarettes (9.1) • nicotine (8.5) • cigarettes (7.5) • cigarette (7.4)
Behavior	—	• intakes (7.9)	• smoking (10.9) • smoked (10.9) • smoke (9.3)	—
Cessation	• assisting (9.1) • switch (8.8) • convert (8.7)	• switched (7.8) • tips (7.5)	• quit (11.5) • help (6.0)	• cessation (7.4)
Health	—	• HrQol (7.6)	• dust (9.5)	• methylation (7.9) • emphysema (7.6)
Legal	• r$1,000 (10.1) • alleges (9.3) • smoker's (9.2) • antismoking (9.1) • estates (9.1) • expended (9.0) • comprised (8.8)	—	—	—
Measures	—	—	• likely (6.6) • higher (6.5)	—
Other	• convince (9.6) • diminished (9.3)	—	—	—

[a]Asterisks indicate "wildcard" searches, which capture variable word endings, and in "smart" searches, they limit word form searches within a specified part of speech category.

[b]TIC: Tobacco Industry Corpus.

[c]HOC: Health Organizations Corpus.

[d]BNC2014: British National Corpus 2014.

[e]COCA: Corpus of Contemporary American English.

[f]ASU30: adult smokers under 30 years.

[g]Not applicable.

Across all 4 corpora, the plural term "smokers" was used to describe either a class ("smokers that"; raw frequencies: TIC=32, HOC=21, BNC2014=3, and COCA=19) or a group of individuals ("smokers who"; raw frequencies: TIC=995, HOC=731, BNC2014=22, and COCA=176), with the group classification ("who") used more often.

XSL•FO
RenderX

Otherwise, overlap in the lexical construction of smokers was limited. Of the 23 specific categorizations of "smoker* N" found among the top collocates, only 7 were shared across all corpora: "former," "never," "non," "current," "ex," "heavy," and "regular," indicating that most high-ranking descriptors were corpus-specific. Different groups, therefore, rely on largely distinct vocabularies when talking about smokers.

Tobacco Users

Tobacco Industry Corpus

In relation to smokers, the TIC was the only corpus that included references to mortality in the lists of top collocates: "deceased" and "alive" are both top collocates of smoker* N*. Together with the prominence of associated actors in the top collocates, such as "heir" and "dependents," this shows a tendency toward impersonal descriptions of tobacco users in legal contexts. On the other hand, the classifier "ASU30" (adult smokers under 30 y) positions smokers as consumers. Health-related collocates were absent among the top results.

Several TIC collocates reflect the tobacco industry's emphasis on regulatory positioning. The term "diminished" appeared in several different contexts, but its key association with "smoker*" was in reference to risk, particularly in the context of regulatory obstacles, for example, "significant risks include our diminished ability to convert adult smokers to our reduced-risk products (RRPs)." Similarly, "convince" was used in reference to a transition to other products, often discussed in the context of regulation and coinciding with calls for "supportive government policy." Additional TIC-exclusive collocates, such as "assisting" and "convert," further supported this trend.

We noted the mention of adult smokers' intentions in these discussions, which suggested that the use of "RRPs" was behaviorally equivalent to smoking, as in "convince all adult smokers who intend to continue smoking to switch to RRPs as soon as possible."

These constructions simultaneously emphasized smoker autonomy ("convince adult smokers") while normalizing continued nicotine consumption ("would otherwise continue to smoke"), suggesting a strategic distribution of responsibility between individual choice and the regulatory context.

Health Organizations Corpus

The HOC foregrounded clinical and behavioral classifications, including "cigarette-only," "persistent," and "nondaily," and positioned smokers within cessation and research contexts through terms such as "quitters," "intakes," and "HrQOL." Many of these classifications were uncommon in the general English corpora, indicating limited uptake of institutional public health terminology in everyday language use. Despite this clinical orientation, disease-specific collocates were not prominent among the HOC's top results.

General English Corpora: BNC2014 and COCA

In the general English corpora, smokers were frequently referred to through contrastive categorization, particularly in association with "nonsmokers", and through strong associations with tobacco and nicotine products. Product-related collocates were more prominent here than in either the TIC or HOC, positioning smokers primarily within consumer and behavioral contexts.

References to disease ("emphysema") and probabilistic risk ("higher","likely") were also present in the general English corpora, which is distinct from both the TIC's legal-consumer framing and the HOC's clinical and research-centered categorizations. Table 5 shows the top collocates of "smok*V*."

XSL•FO
RenderX

Table . Top 20 lexical collocates of smok* V* (ranked by mutual information [MI] score) and their themes[a].

	Top 20 collocates (MI score) sorted in observed themes			
	TIC[b]	HOC[c]	BNC2014[d]	COCA[e]
Smok* specific categorization	—[f]	• some-day (8.6) • day (7.0) • don't (7.0)	• smokers (10.7)	• nonsmokers (8.8) • ex-smokers (8.6) • secondhand (8.5) • non-smokers (8.3)
Actors	• dependents (9.4) • men (8.9)	• mothers (7.6) • anyone (6.9)	—	—
Products	—	• snuffing (8.9) • sucking (8.9) • narghile (7.0)	• spliff (12.7) • cigars (12.3) • fags (12.2) • cigarettes (12.2) • cigar (11.7) • weed (11.7) • cigarette (11.5) • fag (10.7) • pipe (10.5) • marijuana (10.3) • cannabis (9.7)	• menthols (8.4) • kools (8.4) • luckies (8.3)
Behavior	• secondhand (9.5) • gateway (8.4)	• urge (8.0) • smokes (7.5) • involuntary (7.1) • temporarily (7.0) • secondhand (7.0)	• smokes (11.2) • smoked (10.2) • drank (9.5) • smoke (9.5) • smoking (9.4)	• early-experiment (9.7) • cigarettes/day (9.7) • pack-years (9.6) • nondaily (9.6) • two-pack-a-day (8.9) • pack-a-day (8.9)
Cessation	• otherwise (9.2) • switch (8.8) • continuing (8.7) • complete (8.5) • completely (8.5)	—	—	• cessation (8.3)
Health	• respirable (10.6) • multidisciplinary (10.3) • dangers (10.1) • scientifically-substantiated (9.9) • ventilated (9.8) • injured (9.5)	• recode (8.3) • machine-smoking (7.5) • column (7.3) • admitted (7.1) • unknown (7.0)	• BMI (9.4)	• sidestream (10.0) • F2RL3 (8.8) • varenicline (8.5)
Legal	• estates (9.0)	—	—	—
Other	• dissuade (10.0) • competes (9.4) • deserve (8.5)	• waking (7.6) • fewer (7.0)	• deprivation (9.8)	• third (8.6)
Measure	25,000 (10.4)	—	—	—
Other nonrelevant	—	—	• chimney (9.7)	• hoggz (9.9) • continus (9.0)

[a]Asterisks indicate "wildcard" searches, which capture various word endings following a given stem.

[b]TIC: Tobacco Industry Corpus.

[c]HOC: Health Organizations Corpus.

[d]BNC2014: British National Corpus 2014.

[e]COCA: Corpus of Contemporary American English.

[f]Not applicable.

XSL•FO
RenderX

Tobacco Use Behaviors

Tobacco Industry Corpus

As with smoker* N*, smoking behaviors (smok* V*) were associated with legal discourses in the TIC. "Dependents" was among the associated actors, reinforcing litigation-related narratives.

Health-related collocates in the TIC, such as "respirable," "dangers," and "injured," were primarily situated within discussions of liability rather than prevention. The collocates "multidisciplinary" and "scientifically-substantiated" reflect a strategic deployment of scientific language in support of regulatory and product positioning.

Health Organizations Corpus

In the HOC, smoking behaviors were linked to clinical and population-level concerns. The collocate "mothers" points to maternal and infant health contexts, while collocates such as "recode," "machine-smoking," and "unknown" relate to medical research discourse. An exception to this is "admitted," which relates to the provision of health care related to tobacco use.

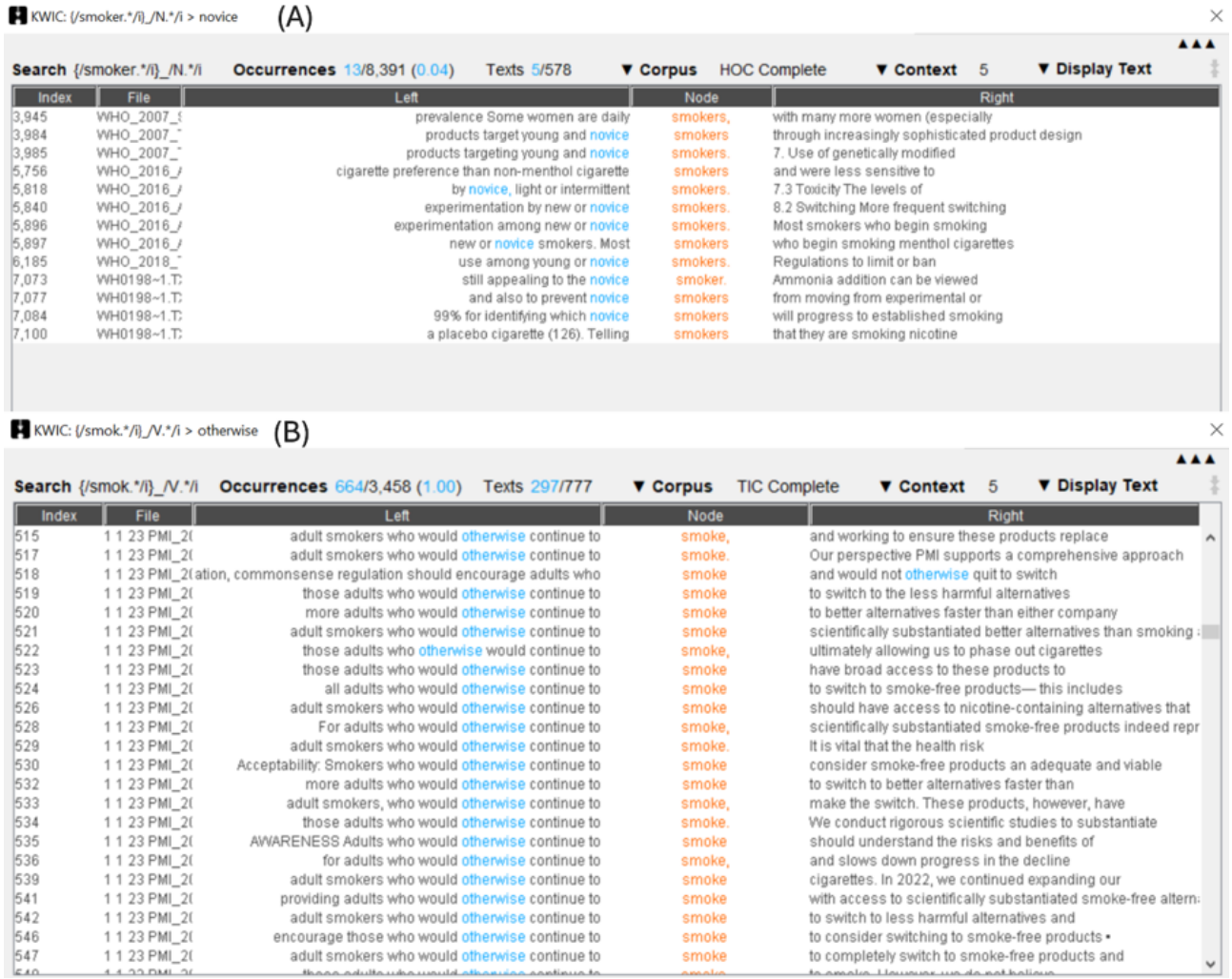
General English Corpora: BNC2014 and COCA

Similar to the collocates of smoker* N*, smoking behavior (smok* V*) is strongly associated with tobacco and nicotine products. In the BNC2014, smoking behavior is also associated with cannabis products. Brand or product names did not feature in the top collocates in the TIC, but both "Luckies" and "Kools" were top collocates in one of the general English corpora (COCA).

Use in Context

Iterative sorting of collocates into themes permitted the exploration of language patterns that either recurred across corpora or were distinctive to particular organizations. For example, the categorization of some tobacco users as "novice smokers" in the HOC, or the classification of tobacco use behaviors such as "otherwise" in the TIC, as shown in Figure 1.

Figure 1. Key Word In Context (KWIC) analysis showing collocations in 2 corpora: (A) "novice" (smoker*N*) in HOC and (B) "otherwise" (smok*V*) in TIC. These results are from a complex "smart search" that includes both a wildcard (open word ending) and tagging for part-of-speech (N* for any noun and V* for any verb). HOC: Health Organizations Corpus; TIC: Tobacco Industry Corpus.



The HOC and COCA both included the term "novice smoker*," but this term was exclusive to the WHO in the HOC and found across 3 genres in COCA (fiction, academic, and web). The threshold for this categorization is not clearly described in either case; it is variably used by the WHO, preceded by either "young and" or "new or."

XSL•FO
RenderX

In the TIC, "otherwise" as a collocate of "smok*V*" was used formulaically in a phrase repeated more or less verbatim from one document to another to talk about a group of people (variably described) "who would otherwise continue to smoke."

Discussion

Principal Findings

Our analysis highlights a high degree of variation in the language used to describe smoking and smokers among the tobacco industry (BAT and PMI), national and international health bodies and services (WHO, NICE, CDC, and NHS), as well as in samples of general English language use (BNC2014 and COCA).

The HOC included more fine-grained classifications for smoker* N* than the other corpora, but each corpus we queried contained descriptors that were not shared with other corpora. This indicates significant and continuing misalignment in the ways in which different groups talk about tobacco use, in spite of calls for change [11,14,15]. It also provides clear evidence of an ongoing "jingle-jangle" problem noted by previous work [11], highlighting key patterns of language held by individual organizations.

The prominence of partially nonoverlapping vocabularies in our analysis, including the differentiation of plural (smokers) and singular (smoker) forms between the corpora, highlights the progress yet to be made in aligning shared understandings of tobacco use risk and determining structural solutions to overcome tobacco-related harms. Indeed, it is evident that addressing tobacco use as a population-level matter is difficult when only 1 group (HOC) frames it as such.

What Is a Smoker and What Is Smoking?

Frequency analysis highlighted that the plural or singular construction of smoker* as a noun diverged between the general English corpora (BNC2014 and COCA) and the tailor-made corpora: "smokers" were significantly more common in the HOC compared to the TIC, BNC2014, and COCA. The more common reference to smokers in the plural by the HOC, as well as the less frequent use of either "people who" or "people that" (smoke*) across all the corpora, indicates that habits of using person-first phrasing have not yet been established in general English, by health organizations or by the industry [13].

Our results show that BAT and PMI consistently frame smokers as consumers ("adult" and "ASU30") and as players in legal dramas ("dependents" and "estates") over dimensions seen in the other corpora ("mothers"). All corpora included categories of behavioral or status descriptions focusing on time or frequency ("never," "daily," and "former"). This is an important reminder that tobacco is a commercially driven industry. The characterization of the agency of the business demonstrated a focus on the provision of access to products and information, while the agency of individuals was emphasized in their will to quit or to "continue to smoke." The characterizations of smoker agency by the TIC were complex, emphasizing the agency of individuals in their phrasing, such as "won't" and "choose not to," for example, while deterministically characterizing the smokers' futures by talking about those who "would otherwise continue to smoke," presenting continued consumption as both chosen and structurally anticipated.

It was not obvious from our analysis that tobacco use uptake was a central concern for any of the groups. It was reported indirectly by PMI, which noted it as a concern of others regarding its "noncombustible" product portfolio acting as a "gateway to smoking." Other terms we reported as top collocates, such as "novice," "early-experiment," and "early-onset," may be useful in future work exploring the narrativization of smoking trajectories by different groups and the role of personal agency and discourse in navigating those trajectories.

The high number of product terms in the general English corpora bolsters existing arguments that product variation is a key driver of nebulous understandings of tobacco use behaviors and combustible tobacco products [8]. The prevalence of associated actors and behaviors in the general English corpora also indicates that smokers and smoking are consistently linked to other behaviors, such as drinking, by general audiences. The plural form "smokers" commonly collocates with "drinkers," which is indicative of a habitual grouping of people according to their behaviors in general English.

Person-first language (people who or that) is not being equally taken up by different groups, calling into question the effectiveness of calls for its use, especially among nonspecialist audiences (those speaking general English).

Implications for Public Health

The uniquely appalling toll the tobacco industry has had on human health is writ large in our collocational analysis of the TIC; key examples include repeated mention of a legal case focused on a "deceased" smoker and the need to categorize "adult" smokers as a distinct use category (which would not be necessary if there were no nonadult smokers). The data also highlight specific categories of "smoker" that may be a key strategic focus for the industry—ASU30—and confirm key areas of concern for the company, such as PMI's "diminished ability" to convert adult smokers to their "RRPs," which was listed as a risk in business reports from 2019 until 2022. This "diminished ability" contrasts with the same organization's reliance on capacity and resource-based framings in their discussions of tobacco-related harm [38] and highlights how access to policy arenas and an ability to shape regulatory policy are key to addressing business risks for transnational tobacco companies.

In a world where even health-focused organizations like the WHO have been subject to critique for their inconsistent and ambiguous language [39], we must endeavor to remain vigilant to linguistic patterns and the epistemic power of organizational actors. The predictability of downstream impacts of product availability and regulation [7,19] may prove elusive given the lack of common or shared phrasing between consumers, policymakers, and tobacco manufacturers, especially in relation to electronic devices [9]. Our results show that these terminological inconsistencies persist across groups, even when the analysis is highly tailored to focus on a single topic (tobacco). These differences are potentially dangerous at both

XSL•FO
RenderX

the individual and population levels. They have been shown to lead to misunderstandings of product harm [4,40] and undermine the consumers' ability to make informed choices about tobacco product use. A powerful tobacco industry can leverage information gaps and misalignments of understanding to its own advantage [41-44] and begin to embed expectations for those who use or regulate tobacco—including expectations about product choice and product use trajectories, for example, in the portrayal of the use of "RRPs" as behaviorally equivalent to smoking.

Our work shows that linguistic clarity and a common understanding of what forms of tobacco use constitute "smoking," as well as what it means to use tobacco (to be "a smoker"), are lacking between key groups. Given that information-seeking is one of many processes of interpersonal communication that can impact personal health [45], and the recognized association between exposure to tobacco-related media content and tobacco use [46], understanding how flows of information in the public domain interact with (or indeed contradict) one another is paramount to the reduction of structural health inequalities. This includes the reduction of informational access barriers and working collaboratively to overcome imbalances resulting from the distortion of "sensemaking resources" [47], whether by intent or by design.

Public information campaigns are one mechanism by which international-level health organizations attempt to mitigate the deleterious effects of tobacco use, and yet we have noted that these organizations use highly variable language in their communications. The tobacco industry, too, fosters its own characterizations of what smoking is and who smokers are. In the background, we have general English—a multitude of different genres (books, news articles, and so on) where individuals are exposed to different characterizations of tobacco products and their use. By highlighting the linguistic variation that characterizes public-domain discussions of tobacco use across key groups, we hope that our work can contribute to public health efforts to stem flows of misinformation and help disarm a tobacco industry with a history of concealing known harms associated with tobacco products and manipulating policy design and implementation for their own ends.

Limitations

As our summaries of the corpora show, there was a significant variance in the size of the corpora we deployed in our analyses, and while we endeavored to mitigate the effect of this by reporting relative frequencies in addition to raw frequencies and by foregrounding other relatively adjusted scores such as mutual information, we recognize that this variance impacts the data. We worked to match our corpora to one another as best we could, including in size and date range, but it is important to point out that the dates of our tailored corpora and the prebuilt corpora do not exactly overlap.

We intentionally excluded personal narratives focused on individuals from the TIC corpora to protect the privacy of named individuals, but existing publications show that these can be valuable in exploring personal narratives relating to tobacco use, especially with regard to reducing hermeneutical injustices through the inclusion of different sense-making practices and resources [47].

The inclusion of documents from organizations within a limited time frame is inherently constrained with regard to the carrying forward of research conclusions. All corporate entities are subject to change, including significant internal restructuring, which can involve reductions in the available specialized expertise used to shape public policy. Such changes may pose a threat to expertise-led information provision in key informational spaces available to consumers. Monitoring linguistic change specific to certain disease areas or product landscapes, such as tobacco, could therefore be helpful in shaping the design of population-relevant and impactful public information campaigns that can be safeguarded against such organizational volatility.

Future Work

Possible valuable future work could seek to query product-specific narratives or focus on the development of narratives in web-based communication to determine how personal technology use, such as information-seeking or dispersed content production, has impacted the information landscape. It would also be valuable to compare discourse across document types within TIC—for example, between customer-facing webpages and more formal corporate or legal reports—to assess how genre and audience influence narrative strategies. Tracing the development of specific terms over time and considering their distribution across genres could also provide insight into the dispersion of industry-constructed language into popular usage and identify key informational asymmetries and key waypoints in their evolution.

Acknowledgments

The authors would like to thank Karin Silver and Dr Allen Gallagher for their expertise and guidance. No generative artificial intelligence tools were used at any stage in the preparation of this manuscript.

Funding

IF receives funding from Bloomberg Philanthropies (https://www.bloomberg.org/) under the Stopping Tobacco Organisations and Products grant. Xinmei Sun was supported by the Bloomberg Stopping Tobacco Organisations and Products grant. All authors were funded by Bloomberg Philanthropies as part of the Bloomberg Initiative to Reduce Tobacco Use. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

XSL•FO
RenderX

Data Availability

The dataset for this study is available under a CC license at the University of Bath Research Data Archive.

Authors' Contributions

Conceptualization: IF (lead), XS (supporting)
Formal analysis: IF (supporting), XS (lead)
Methodology: IF (equal), XS (equal)
Visualization: IF (equal), XS (equal)
Writing—original draft: IF (lead), XS (supporting)
Writing—review & editing: IF (equal), XS (equal)

Conflicts of Interest

None declared.

References

1. WHO global report on trends in prevalence of tobacco use 2000–2025. : World Health Organization; 2019 URL: https://iris.who.int/server/api/core/bitstreams/a64f9b44-346d-465c-895f-376e8709c379/content [accessed 2026-05-20]
2. Dai X, Gakidou E, Lopez AD. Evolution of the global smoking epidemic over the past half century: strengthening the evidence base for policy action. Tob Control 2022 Mar;31(2):129-137. [doi: 10.1136/tobaccocontrol-2021-056535] [Medline: 35241576]
3. Ling PM, Kim M, Egbe CO, Patanavanich R, Pinho M, Hendlin Y. Moving targets: how the rapidly changing tobacco and nicotine landscape creates advertising and promotion policy challenges. Tob Control 2022 Mar;31(2):222-228. [doi: 10.1136/tobaccocontrol-2021-056552] [Medline: 35241592]
4. Byron MJ, Jeong M, Abrams DB, Brewer NT. Public misperception that very low nicotine cigarettes are less carcinogenic. Tob Control 2018 Nov;27(6):712-714. [doi: 10.1136/tobaccocontrol-2017-054124] [Medline: 29363610]
5. Antin TMJ, Sanders E, Lipperman-Kreda S, Annechino R, Peterkin E. “I Can’t Make Perfect Choices All the Time”: perspectives on tobacco harm reduction among young adults who identify as sexual and gender minorities. Nicotine Tob Res 2023 May 22;25(6):1090-1098. [doi: 10.1093/ntr/ntac291] [Medline: 36548953]
6. Tan ASL, Bigman CA. Misinformation about commercial tobacco products on social media-implications and research opportunities for reducing tobacco-related health disparities. Am J Public Health 2020 Oct;110(S3):S281-S283. [doi: 10.2105/AJPH.2020.305910] [Medline: 33001728]
7. Loud EE, Duong HT, Henderson KC, et al. Addicted to smoking or addicted to nicotine? A focus group study on perceptions of nicotine and addiction among US adult current smokers, former smokers, non-smokers and dual users of cigarettes and e-cigarettes. Addiction 2022 Feb;117(2):472-481. [doi: 10.1111/add.15634] [Medline: 34227709]
8. Measham F, O’Brien K, Turnbull G. “Skittles & Red Bull is my favourite flavour”: E-cigarettes, smoking, vaping and the changing landscape of nicotine consumption amongst British teenagers – implications for the normalisation debate. Drugs Educ Prev Policy 2016 May 3;23(3):224-237. [doi: 10.1080/09687637.2016.1178708]
9. East KA, Miller CR, Hitchman SC, McNeill A, Tompkins CNE. “It’s not what you’d term normal smoking”: a qualitative exploration of language used to describe heated tobacco product use and associated user identity. Addiction 2023 Mar;118(3):533-538. [doi: 10.1111/add.16051] [Medline: 36148622]
10. Perman-Howe PR, Horton M, Robson D, McDermott MS, McNeill A, Brose LS. Harm perceptions of nicotine-containing products and associated sources of information in UK adults with and without mental ill health: a cross-sectional survey. Addiction 2022 Mar;117(3):715-729. [doi: 10.1111/add.15657] [Medline: 34338387]
11. Notley C. Words matter-a call for consistent and accurate scientific language. Nicotine Tob Res 2025 Jun 23;27(7):1155-1156. [doi: 10.1093/ntr/ntaf073] [Medline: 40178382]
12. Goodwin RD, Walker LK. Time to stop using the word “Smoker”: reflecting on the role of language in advancing the field of nicotine and tobacco research. Nicotine Tob Res 2022 Nov 12;24(12):1847-1848. [doi: 10.1093/ntr/ntac218] [Medline: 36130319]
13. People-first language policy. Tobacco Control (BMJ journal). URL: https://tobaccocontrol.bmj.com/pages/people-first-language-policy [accessed 2025-07-15]
14. Botticelli MP, Koh HK. Changing the language of addiction. JAMA 2016 Oct 4;316(13):1361-1362. [doi: 10.1001/jama.2016.11874] [Medline: 27701667]
15. Williamson TJ, Riley KE, Carter-Harris L, Ostroff JS. Changing the language of how we measure and report smoking status: implications for reducing stigma, restoring dignity, and improving the precision of scientific communication. Nicotine Tob Res 2020 Dec 12;22(12):2280-2282. [doi: 10.1093/ntr/ntaa141] [Medline: 32756966]
16. Chaiton M. Embrace the smoker: person-first language is not a solution to stigma. Nicotine Tob Res 2023 Jul 14;25(8):1509-1510. [doi: 10.1093/ntr/ntad047] [Medline: 36964765]

XSL•FO
RenderX

17. van Druten VP, Bartels EA, van de Mheen D, de Vries E, Kerckhoffs APM, Nahar-van Venrooij LMW. Concepts of health in different contexts: a scoping review. BMC Health Serv Res 2022 Mar 24;22(1):389. [doi: 10.1186/s12913-022-07702-2] [Medline: 35331223]
18. O'Connor R, Schneller LM, Felicione NJ, Talhout R, Goniewicz ML, Ashley DL. Evolution of tobacco products: recent history and future directions. Tob Control 2022 Mar;31(2):175-182. [doi: 10.1136/tobaccocontrol-2021-056544] [Medline: 35241585]
19. Lempert LK, Grana R, Glantz SA. The importance of product definitions in US e-cigarette laws and regulations. Tob Control 2016 Apr;25(e1):e44-e51. [doi: 10.1136/tobaccocontrol-2014-051913] [Medline: 25512432]
20. Cox S, West R, Notley C, Soar K, Hastings J. Toward an ontology of tobacco, nicotine and vaping products. Addiction 2023 Jan;118(1):177-188. [doi: 10.1111/add.16010] [Medline: 35971622]
21. Maddox R, Diaz A, Waa A, et al. Resisting industry narratives: guidance to avoid tobacco and nicotine industry framing. Health Promot Int 2024 Dec 1;39(6):daae143. [doi: 10.1093/heapro/daae143] [Medline: 39569484]
22. Poland B, Frohlich K, Haines RJ, Mykhalovskiy E, Rock M, Sparks R. The social context of smoking: the next frontier in tobacco control? Tob Control 2006 Feb;15(1):59-63. [doi: 10.1136/tc.2004.009886] [Medline: 16436407]
23. McEnery T, Hardie A. Corpus Linguistics: Method, Theory and Practice: Cambridge University Press; 2011. [doi: 10.1017/CBO9780511981395]
24. Ma W, Liu Q. Language and health studies in the era of holistic health: achievements and prospects. Lang Health 2023 Dec;1(2):1-11. [doi: 10.1016/j.laheal.2023.11.001]
25. P Goddu A, O'Conor KJ, Lanzkron S, et al. Do Words Matter? stigmatizing language and the transmission of bias in the medical record. J Gen Intern Med 2018 May;33(5):685-691. [doi: 10.1007/s11606-017-4289-2] [Medline: 29374357]
26. Leading tobacco companies worldwide in 2025, based on market value (in billion U.S. dollars). Statista. URL: https://www.statista.com/statistics/942132/leading-10-tobacco-companies-worldwide-based-on-net-sales [accessed 2022-07-08]
27. Brezina V, Hawtin A, McEnery T. The Written British National Corpus 2014—design and comparability. Text Talk 2021 Oct 26;41(5-6):595-615. [doi: 10.1515/text-2020-0052]
28. Love R, Dembry C, Hardie A, Brezina V, McEnery T. The spoken BNC2014: designing and building a spoken corpus of everyday conversation. Int J Corpus Linguist 2017;22(3):319-344. [doi: 10.1075/ijcl.22.3.02lov]
29. The Corpus of Contemporary American English (COCA). URL: https://www.english-corpora.org/coca/ [accessed 2026-05-20]
30. Wayback Machine. Internet Archive. URL: https://web.archive.org/ [accessed 2026-05-20]
31. archive.today. URL: https://archive.ph/ [accessed 2026-06-03]
32. Institutional repository for information sharing (IRIS). World Health Organization. URL: https://iris.who.int/home [accessed 2026-05-20]
33. Pdfminer 20191125. Python Package Index (PyPI). URL: https://pypi.org/project/pdfminer/20191125/ [accessed 2026-05-20]
34. Fitzpatrick I, Sun X. Dataset for "Investigation of differences in tobacco use language between groups: corpus assisted analysis". University of Bath Research Data Archive. 2026.
35. Baker P. Using Corpora in Discourse Analysis: Continuum; 2006. [doi: 10.5040/9781350933996]
36. Brezina V. Statistics in Corpus Linguistics: A Practical Guide: Cambridge University Press; 2018.
37. NLTK's list of English stopwords. GitHub Gist. URL: https://gist.github.com/sebleier/554280?permalink_comment_id=3744516#gistcomment-3744516 [accessed 2026-05-20]
38. Fitzpatrick I, Dance S, Silver K, Violini M, Hird TR. Tobacco industry messaging around harm: narrative framing in PMI and BAT press releases and annual reports 2011 to 2021. Front Public Health 2022;10:958354. [doi: 10.3389/fpubh.2022.958354] [Medline: 36330126]
39. Amri M, Enright T, O'Campo P, Di Ruggiero E, Siddiqi A, Bump JB. Health promotion, the social determinants of health, and urban health: what does a critical discourse analysis of World Health Organization texts reveal about health equity? BMC Glob Public Health 2023 Dec 1;1(1):25. [doi: 10.1186/s44263-023-00023-4] [Medline: 39681886]
40. Yang B, Massey ZB, Popova L. Effects of modified risk tobacco product claims on consumer comprehension and risk perceptions of IQOS. Tob Control 2022 Aug;31(e1):e41-e49. [doi: 10.1136/tobaccocontrol-2020-056191] [Medline: 33688084]
41. Balbach ED, Smith EA, Malone RE. How the health belief model helps the tobacco industry: individuals, choice, and "information". Tob Control 2006 Dec;15 Suppl 4(Suppl 4):iv37-iv43. [doi: 10.1136/tc.2005.012997] [Medline: 17130622]
42. Lee CH. The effects of tobacco company's image restoration communication strategies on smoking attitudes and company images [Article in Korean]. Korean J Advert Public Relat 2007;9(4):135-161 [FREE Full text]
43. Brown C, Rubin DL. Causal markers in tobacco industry documents: the pragmatics of responsibility. J Pragmat 2005 Jun;37(6):799-811. [doi: 10.1016/j.pragma.2004.10.015]
44. Cappella JN, Maloney E, Ophir Y, Brennan E. Interventions to correct misinformation about tobacco products. Tob Regul Sci 2015 Jul;1(2):186-197. [doi: 10.18001/TRS.1.2.8] [Medline: 27135046]
45. Ackerson LK, Viswanath K. The social context of interpersonal communication and health. J Health Commun 2009;14 Suppl 1(sup1):5-17. [doi: 10.1080/10810730902806836] [Medline: 19449264]

XSL•FO
RenderX

46. Hassanein ZM, Barker AB, Murray RL, Britton J, Agrawal S, Leonardi-Bee J. Impact of smoking and vaping in films on smoking and vaping uptake in adolescents: systematic review and meta-analysis. Health Educ Behav 2022 Dec;49(6):1004-1013. [doi: 10.1177/10901981221086944] [Medline: 35502561]
47. Ajmani LH, Foriest JC, Taylor J, Pittman K, Gilbert S, DeVito MA. Whose knowledge is valued? Epistemic injustice in CSCW applications. Proc ACM Hum-Comput Interact 2024 Nov 7;8(CSCW2):1-28. [doi: 10.1145/3687062]

Abbreviations

ASU30: adult smokers under 30 years
BAT: British American Tobacco
BNC2014: British National Corpus 2014
CDC: Centers for Disease Control and Prevention
COCA: Corpus of Contemporary American English
HOC: Health Organizations Corpus
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
PMI: Philip Morris International
RRP: reduced-risk product
TIC: Tobacco Industry Corpus
WHO: World Health Organization

Edited by T Mackey; submitted 27.Oct.2025; peer-reviewed by F Watson, RR Jablonkai; revised version received 22.Apr.2026; accepted 04.May.2026; published 15.Jul.2026.

Please cite as:
Fitzpatrick I, Sun X
Investigation of Differences in Tobacco Use Language Between Groups: Corpus-Assisted Analysis
JMIR Infodemiology 2026;6:e86593
URL: https://infodemiology.jmir.org/2026/1/e86593
doi:10.2196/86593

© Iona Fitzpatrick, Xinmei Sun. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 15.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Portrayals of Depression on TikTok: Content Analysis of Diagnostic Accuracy, Creator Type, and Stylistic Features

Elena Rainer, MSc; Amber van der Wal, PhD; Ine Beyens, PhD

Amsterdam School of Communication Research, University of Amsterdam, P.O. Box 15791, Amsterdam, NH, The Netherlands

Corresponding Author:
Amber van der Wal, PhD
Amsterdam School of Communication Research, University of Amsterdam, P.O. Box 15791, Amsterdam, NH, The Netherlands

Abstract

Background: Youths are increasingly turning to TikTok for mental health information, making the platform an important space where young people encounter portrayals of mental illness. While such visibility can raise awareness, reduce stigma, and make young people feel more connected and understood in their experiences, concerns have been raised about the diagnostic accuracy of this content, which is often produced by nonprofessionals and presented using emotionally appealing stylistic features. Although prior research has examined mental health content on TikTok broadly, little is known about how depression-related symptoms are portrayed by creators on the platform.

Objective: Given depression's rising prevalence among youth and its prominent presence on TikTok, this study examined (1) the diagnostic accuracy of TikTok videos about depression, (2) differences in diagnostic accuracy and stylistic features by creator type (medical professionals vs nonprofessionals), and (3) how diagnostic accuracy, stylistic features (personal experiences, emotional appeals, and background music), and creator type relate to user engagement.

Methods: A quantitative content analysis was conducted of 210 English-language TikTok videos retrieved using symptom-focused search terms (eg, "depression symptoms"). Videos were coded for diagnostic accuracy using a standardized coding scheme based on the *International Classification of Diseases, 11th Revision* diagnostic criteria for depressive episodes. In addition, videos were coded for creator type, presentation style, and the presence of emotionally appealing stylistic features. Engagement was operationalized as the sum of a video's likes, comments, saves, and shares. Intercoder reliability was assessed using Krippendorff α, percent agreement, and Gwet AC1 (agreement coefficient 1). Analyses included Mann-Whitney *U* tests, chi-square tests, and hierarchical regression.

Results: Diagnostic accuracy was low overall (mean score 1.21, SD 1.04, on a 0 - 4 scale) and did not differ significantly between medical professionals and nonprofessionals (median 1.40 [IQR 1-2] vs 1.11 [IQR 0-2]; P=.06). Hierarchical regression analysis showed that diagnostic accuracy did not predict engagement (B=−0.10; P=.19). In contrast, engagement was higher for videos containing personal experiences (B=0.41; P=.02), emotional appeals (B=0.73; P=.001), and background music (B=0.54; P=.01). Across regression models, direct-to-camera formats (Bs −0.49 to −0.69; .003≤P≤.04) and text-centered videos (Bs −0.56 to −0.64; .002≤P<.01) were associated with lower engagement.

Conclusions: Depression-related content on TikTok is characterized by limited diagnostic completeness, regardless of creator type. Engagement appears to be driven primarily by stylistic features rather than diagnostic accuracy. These patterns raise concerns about concept creep—the gradual expansion of the psychological concept for depression—and the potential for premature self-diagnosis among young users, while also highlighting opportunities for medical professionals to adapt their communication styles on TikTok to increase both accuracy and engagement.

(JMIR Infodemiology 2026;6:e85323) doi:10.2196/85323

KEYWORDS

TikTok; mental health; depression; diagnostic accuracy; stylistic features; user engagement

Introduction

Background

Social media platforms have become central spaces where young people are exposed to and actively search for information on mental health [1]. Among these platforms, TikTok stands out for its unparalleled reach and influence among youth, with mental health hashtags like #mentalhealth or #mentalillness amassing billions of views [2,3]. This increased accessibility to mental health information can have important benefits: It may raise awareness, reduce stigma, help young people make sense of their own experiences, foster feelings of validation and community, and encourage help-seeking among those who might otherwise feel isolated or reluctant to seek support [4]. At the same time, concerns have been raised about the accuracy

XSL•FO
RenderX

of mental health content on TikTok. Much of this content is created by nonmedical professionals and is often presented using emotionally appealing stylistic features—such as background music, personal storytelling, and expressions of fear, hope, or humor [3,5-8]—which may increase relatability but also raise the risk that portrayals diverge from clinical definitions of mental illness [6,9].

Of particular concern is that mental health content characterized by diagnostic inaccuracy and appealing stylistic features tends to attract the highest levels of audience engagement [5-7,9,10]. On TikTok, engagement refers to visible indicators of user interaction, such as views, likes, comments, and shares. Importantly, such engagement metrics not only reflect a video's popularity but also steer the platform's recommendation system, increasing the likelihood that a video is surfaced more prominently and repeatedly within users' feeds and distributed more widely across the platform [5]. As a result, diagnostically imprecise yet emotionally appealing mental health portrayals from nonprofessionals are likely to be disproportionately promoted across the platform, increasing their visibility and shaping young viewers' perceptions and understanding of mental illness.

Widespread engagement with inaccurate mental health content can have several downstream consequences. It may lead to *concept creep* [11]: the gradual expansion of psychological concepts over time, broadening the accepted range of what experiences or situations encompass certain mental illnesses. When symptoms are vaguely or inaccurately portrayed, the boundaries of what constitutes a disorder may become increasingly diffuse, and the concept of that illness broadens [12]. This, in turn, can lower the threshold for *self-diagnosing*: the act of independently labeling oneself with a medical diagnosis without an official assessment by a medical professional [13]. For some individuals—particularly those facing barriers to professional care, such as financial constraints or long waiting lists—self-diagnosing may benefit their mental health. However, when diagnostic labels are applied on the basis of incomplete or inaccurate information, this may also contribute to confusion about clinical thresholds, misinterpretation of symptoms, and added pressure on already overstretched mental health systems [14].

Of particular interest in this context is depression-related content on TikTok. Not only is depression one of the most-discussed mental health issues on TikTok [2,3,7], but it is also one of the most prevalent and rising mental health disorders among young people worldwide [15]. Yet, most research on mental health content on TikTok has focused on heterogeneous mental health content, which does not permit condition-specific conclusions for depression. To the best of our knowledge, only one study has focused specifically on depression-related content on TikTok or a comparable short-form video platform; however, this study was limited to personal self-disclosure videos and did not assess the diagnostic accuracy of symptom portrayals [16]. A substantial body of research has examined depression portrayals on other social media platforms [17]—primarily text-based platforms such as Twitter (subsequently rebranded as X)—but this literature likewise has not evaluated diagnostic accuracy. This highlights the need for a comprehensive examination of depression portrayals on TikTok.

This Study

Through a quantitative content analysis of 210 TikTok videos, this study analyzes how accurately depression symptoms are portrayed in relation to objective diagnostic criteria, what emotionally stylistic features are used to convey the content, and how these elements vary depending on whether the creator is a medical professional or not. In addition, this study explores how the diagnostic accuracy, emotionally appealing stylistic features, and creator type are associated with levels of user engagement.

Diagnostic Accuracy of TikTok Videos About Depression

In this study, TikTok videos about depression are considered low in diagnostic accuracy when they misrepresent depression by portraying experiences that fall outside established clinical definitions—for example, labeling brief or situational sadness as "depression"—or when they rely on vague, oversimplified language (eg, "feeling down") without providing additional clinical context. Such portrayals omit key diagnostic distinctions required to differentiate depression from everyday emotional experiences.

Little is known about the diagnostic accuracy of TikTok videos about depression and the factors determining it. Prior research on broader mental health content suggests that creator type is an important determinant of diagnostic accuracy, with videos created by medical professionals being more accurate than those created by nonmedical professionals [5,18]. We expect that this pattern also applies to depression portrayals on TikTok, leading to the following hypothesis:

- H1a: The diagnostic accuracy of TikTok videos about depression is higher when the creator is a medical professional compared to a nonprofessional.

Although videos about depression that are created by medical professionals may be higher in diagnostic accuracy, this does not necessarily translate into greater user engagement, as prior research on general mental health content on TikTok has shown that videos low in diagnostic accuracy often achieve higher levels of engagement [6,9,10]. The elaboration likelihood model (ELM) [19] offers a framework for understanding this pattern. According to the ELM, there are 2 routes of information processing: the central route, involving deliberate and thoughtful evaluation, and the peripheral route, which relies on surface-level cues, such as source attractiveness. Which processing route is taken depends largely on the individual's motivation and ability to process the information. On TikTok, a platform designed for fast-paced, entertaining content consumption, users are more likely to process information peripherally [20]. As a result, TikTok videos that emphasize clinically detailed, high-accuracy information may receive lower engagement because they place higher cognitive demands on viewers. This leads to the following hypothesis:

XSL•FO
RenderX

- H1b: TikTok videos about depression that have lower diagnostic accuracy achieve higher engagement compared to those that have higher diagnostic accuracy.

Personal Experiences and Emotional Appeals in TikTok Videos About Depression

Prior research on general mental health content on TikTok has shown that nonmedical professionals use more emotionally appealing stylistic features—in particular, personal experiences and emotional appeals—than medical professionals [3,5,7,8,16]. In addition, mental health videos with emotionally appealing stylistic features have consistently been linked to higher user engagement [5,7,10]. Personal experiences—often relayed with vivid emotional detail—tend to receive particularly high engagement on TikTok [2,5-7,10]. Mental health videos are also frequently emotionally charged [6] and both positive (eg, affiliation, hope, and ease [5]) and negative emotional appeals (eg, peril, fear, and status loss [5,16]) have been found to increase engagement. This is in line with identification theory, which posits that engagement is more likely when audiences psychologically and emotionally align with a media persona—such as a TikTok creator—by momentarily adopting their "identity, goals, and perspectives" [21, p. 261]. This leads to the following hypotheses:

- H2a: TikTok videos about depression will contain fewer personal experiences when the creator is a medical professional compared to a nonprofessional.
- H2b: TikTok videos about depression will contain fewer emotional appeals (positive or negative) when the creator is a medical professional compared to a nonprofessional.
- H2c: TikTok videos about depression that include personal experiences achieve higher engagement compared to those that do not include personal experiences.
- H2d: TikTok videos about depression that include emotional appeals (positive or negative) achieve higher engagement compared to those that do not include emotional appeals.

Background Music in TikTok Videos About Depression

Background music is a third appealing stylistic feature commonly used in mental health TikTok videos [6]. As with personal experiences and emotional appeals, we expect that the use of background music in TikTok videos about depression may also differ between creator types. Where prior studies have compared the use of emotional appeals and personal experiences across creator types [5,7], no research to date has explicitly examined differences in background music use between medical professionals and nonmedical creators on TikTok. However, given their more formal communication style, medical professionals—who already tend to avoid personal narratives and emotional appeals—we expect they may likewise be less inclined to incorporate background music. To test this assumed relationship, the following hypothesis is proposed:

- H3a: TikTok videos about depression are less likely to contain background music when the creator is a medical professional compared to a nonprofessional.

It has also not been empirically tested whether background music influences engagement with mental health content. Yet, transportation theory suggests that stylistic features such as music can enhance emotional immersion into a narrative, thereby increasing engagement [22]. In addition, from the perspective of the ELM [19], background music can be interpreted as a peripheral cue that may influence engagement without requiring extensive cognitive processing. On a fast-paced platform like TikTok, where users make engagement decisions in a matter of seconds, the presence of background music may help content stand out, prompting greater user engagement. Accordingly, we hypothesize:

- H3b: TikTok videos about depression with background music achieve higher engagement compared to those without background music.

Methods

Ethical Considerations

This study did not involve direct interaction with human participants, as it was based on a content analysis of publicly available TikTok videos. Accordingly, informed consent and participant compensation were not applicable. The study protocol was reviewed and approved by the Ethics Review Board of the authors' university (FMG-14125) prior to data collection. Given that the analyzed content may include identifiable information, appropriate measures were taken to ensure privacy and confidentiality. All data were collected and stored on secure university infrastructure, and no personally identifiable information is reported in this study. The study design, data collection, coding procedures, and reporting adhere to established standards for quantitative content analysis [23].

Sample and Procedure

Using G*Power 3.1, we conducted a priori power analyses corresponding to the main analyses of the study, which indicated that a sample of 210 videos was required to detect medium effect sizes with 95% power and an α of 5%. This sample size also aligns with sample sizes of previous content analyses on general mental health portrayals on TikTok [2,5,24].

To collect our sample of videos, we relied on search term–based sampling rather than hashtag-based sampling to better approximate how users actively seek information about depression-related symptoms on TikTok. Whereas hashtag searches primarily reflect creator-driven labeling practices and platform trends—often optimized for visibility, community signaling, or reach—search queries more closely capture user-driven informational intent. To identify appropriate search terms, we conducted a pilot review designed to mimic typical user search behavior. We compared broader queries such as "depression" and "depressed" with more symptom-focused terms like "depression symptoms" and "symptoms of depression," reviewing the first 20 - 30 videos returned for each. The broader searches primarily surfaced clips that conveyed the general affective tone of depression—often through mood or music—without explicitly referencing specific symptoms. In contrast, the symptom-related queries yielded videos more directly focused on naming, describing, or illustrating depressive symptoms. Based on these findings, and in alignment with the study's diagnostic focus, we selected 4 symptom-specific search terms that also reflect plausible user

XSL•FO
RenderX

queries: “depression symptoms,” “symptoms of depression,” “depression signs,” and “signs of depression.”

To minimize preexisting algorithmic biases—such as personalized content curation based on prior user behavior, interactions, or viewing history—a new TikTok account was created for data collection. This ensured that the search results were not influenced by an established user profile and instead reflected a more neutral presentation of content. Following the approach of Jerin and colleagues [5], the web scraping tool Octoparse was used to extract an initial sampling frame of 731 TikTok videos by using the aforementioned search terms. This larger initial scrape was intended to ensure a broad enough pool of videos that would meet all inclusion criteria, allowing for the selection of the desired sample size.

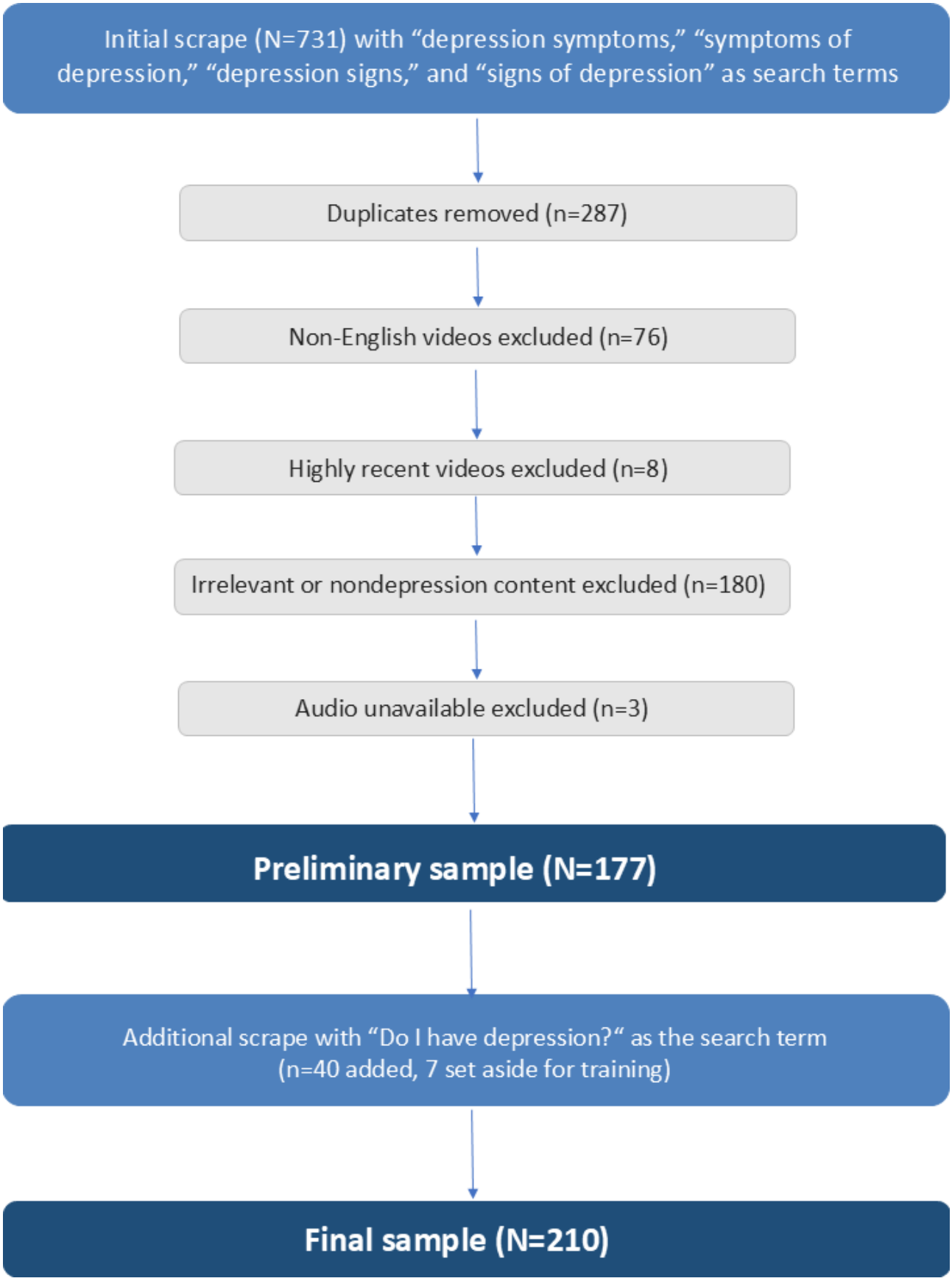
To ensure that only relevant and analyzable videos were included, a multistep filtering process was applied based on predefined inclusion and exclusion criteria related to language, recency, content focus, and diagnostic relevance (Figure 1). First, the sampling frame was screened to exclude duplicates (n=287) and non-English language videos (n=76), as TikTok videos had to be in English to maintain consistency in analysis and ensure clarity in interpretation. Additionally, videos posted within the last week prior to data collection were excluded (n=8) to avoid premature capture of underexposed content, since TikTok videos can continue to gain engagement several days after posting.

The remaining sampling frame of 363 was then manually filtered to include only videos that explicitly discussed or displayed at least one symptom of depression, regardless of whether the symptom was diagnostically accurate. Videos had to refer to general depression and were excluded if they focused on specific subtypes of depression (eg, postpartum depression) or addressed other mental health problems (eg, anxiety) without clearly distinguishing them from depression. In cases where the presence of depression-related symptoms was unclear, captions and hashtags were consulted; if indicators such as “#depression” were present, the video was retained under the assumption that the symptom depiction referred to depression. This filtering step resulted in the exclusion of 180 videos. Videos were included regardless of whether depressive symptoms were presented as first-person self-disclosures, secondhand descriptions of another individual’s symptoms, or general informational content.

Finally, 3 videos were excluded because the audio was not available anymore. This led to a preliminary final sample of 177. To achieve the required sample size of 210, one additional scrape using Octoparse with the search term “do I have depression” was conducted. This search phrase reflects a more personalized and diagnostic-seeking query style compared to the initial symptom-focused terms, yet still yielded content centered on symptom portrayals. After the same filtering procedure and application of exclusion criteria, this added 40 new TikTok videos to the sample, of which 7 were randomly selected to be put aside for the purpose of coder training. Consequently, through this purposive sampling method, the final sample of 210 was obtained. The sample was collected between May 5 and May 11, 2025, and all TikTok videos were either screen-recorded or downloaded (if allowed by the video’s settings) at the time of sampling.

XSL•FO
RenderX

Figure 1. Flowchart of the TikTok video selection process, including initial scraping, exclusion steps, and the final analytic sample. “N” denotes the sample size at each stage, and “n” denotes the number of videos excluded at each step.



XSL•FO
RenderX

Coding Procedure

Unit of Analysis

The individual TikTok video was the unit of analysis [23], and the entire video with its caption and hashtags was considered the context frame [23]. This context frame included spoken content, on-screen overlay text, and visual depictions, as well as caption text and hashtags, but did not include user comments. A full description of the coding process, including the coding manual, can be found on the Open Science Framework [25].

Video Characteristics

The posting date and all engagement metrics (views, likes, comments, shares, and saves) were manually recorded at the time of sampling to capture engagement data as it appeared at the time of data collection. This procedure allowed for accurate engagement comparisons that were not affected by differences in time of coding. From this information, engagement was then calculated as the sum of likes, comments, saves, and shares (ie, active metrics that reflect a user's deliberate decision to interact with content), while the passive metric of views was omitted. This approach follows prior work conceptualizing engagement as user‐initiated responses rather than mere exposure [7,10] and represents the outcome variable of interest for this study. During this stage, it was also recorded whether the creator self-identified as a medical professional or not by assessing explicit statements about professional status in their username, biography, caption of the video, or mentions in the video.

Diagnostic Accuracy

General Depressive Episodes

In this study, *diagnostic accuracy* was assessed through clinically recognized criteria from the *International Classification of Diseases, 11th Revision* (*ICD-11*) *for Mortality and Morbidity Statistics* [26]. The *ICD-11* is a globally accepted diagnostic system used by medical professionals and researchers to define and classify mental health disorders, including general depressive episodes [27]. According to the *ICD-11*, a depressive episode is defined as a period of persistently depressed mood and/or loss of interest in activities, lasting for a minimum of 2 weeks, accompanied by at least one other symptom like fatigue, feelings of worthlessness or guilt, or difficulty concentrating [26]. Hence, three *ICD-11*–based criteria were used: presence of core and associated symptoms (A), symptom frequency (B), and symptom duration (C). A fourth criterion regarding the presence of nondiagnostic symptoms was developed for the purpose of this study.

Criterion A: Presence of Core and Associated Symptoms

Based on the *ICD-11*, a depressive episode requires at least one of the core symptoms *depressed mood* and *diminished interest in activities*. The core symptoms must be accompanied by at least one other symptom from a list of 8 further diagnostic symptoms: *difficulty concentrating*, *feelings of worthlessness or guilt*, *hopelessness*, *recurrent thoughts of death or suicide*, *changes in sleep*, *changes in appetite*, *psychomotor agitation or retardation*, or *reduced energy or fatigue*. Each video was coded for the presence of these 10 symptoms. The codings were then combined to determine whether the video fulfilled this first diagnostic criterion, that is, whether it depicted at least one core symptom in conjunction with at least one additional symptom.

Criterion B: Symptom Frequency

Second, informed by the *ICD-11*, a depressive episode needs to be present for most of the day, nearly every day, to fulfill diagnostic criteria. This criterion was coded as 1 if it was mentioned in the video and as 0 if the mention was absent.

Criterion C: Symptom Duration

The third criterion of the *ICD-11* is the duration of the symptoms, as they must be present for at least 2 weeks. This duration criterion was again coded as 1 for present and 0 for absent.

Criterion D: Mention of Nondiagnostic Symptoms

This study-specific criterion assessed whether the video included portrayals of symptoms not recognized by the *ICD-11*. In our initial exploratory screening process, we cataloged the most common and recurring non–*ICD-11* symptoms mentioned and constructed a predefined list of 17 nondiagnostic symptoms. An "other" category was provided, allowing coders to note any additional symptoms not covered by the predetermined list. Examples of nondiagnostic symptoms include "putting on a front" or "messy room." Nondiagnostic symptoms were coded as present (=1) or absent (=0); if no nondiagnostic symptom was present in a video, it was later recoded as 1 for the purposes of the diagnostic accuracy score.

To calculate a diagnostic accuracy score for every video, the 4 criteria (A-D) were summed to create a composite score ranging from 0 to 4, with higher values reflecting a more accurate diagnostic depiction of depression. Because the presence of nondiagnostic symptoms indicates inaccuracy, their absence contributed positively to the overall score.

Personal Experiences, Emotional Appeals, and Background Music

Each video was coded for 3 stylistic features: personal experiences, emotional appeals, and background music. Personal experiences were coded as present when the video included personal storytelling elements (eg, the creator's own experience or that of someone they know). Emotional appeals were coded as present when the video conveyed emotional expressions, categorized as negative (eg, sadness, fear, shame), humorous (eg, sarcasm, irony), or positive (eg, hope, encouragement, empowerment). Background music was coded for its presence or absence, defined as any musical track or sound layered beneath spoken or visual content, and its valence was classified as sad, neutral, or uplifting.

Control Variables

Following prior TikTok content analyses [3,6,8], we coded for presentation style to account for basic differences in how videos are delivered, which may independently shape engagement. Presentation style categories were adopted from previous work [3,8] and slightly modified to better fit the corpus under study. Each video was coded for its dominant presentation style, categorized into "speaking directly to camera," "acted-out scenes," "animation," "movie or TV scene," "text-centered

XSL•FO
RenderX

videos," "someone giving a speech," "podcast interview," or an open "other" category for formats not captured by these options.

Reliability Testing

For intercoder reliability (ICR) testing of the coding scheme, an independent coder was trained by the first author, who then participated in the coding process. A set of 7 videos was used for coder training, which were not part of the final dataset. The training began with an introduction session in which 2 TikTok videos were used as examples to explain the coding process. The second coder then independently coded 3 TikTok videos and received detailed feedback. Any uncertainties were discussed, and the coding scheme was refined accordingly. Finally, 2 TikTok videos were used for a concluding training session, after which no further questions remained.

Following the training phase, each coder coded a randomly selected 10% of the sample (n=21), a threshold that adheres to existing guidelines to calculate ICR [23]. Krippendorff α was used as the reliability coefficient due to its commonly argued superiority for applicability to relatively small sample sizes and all levels of data, including dichotomous variables [23,28,29]. Most variables demonstrated good levels of reliability, ranging from a moderate α of 0.74 to exceeding a strong α of ≥0.8 [30]. Six variables were below the accepted minimum threshold of α≥0.67 [30]. Two of these were adjusted ("negative emotional appeal" was combined with the other two types of emotional appeals into a new variable "emotional appeal") or removed ("background music valence"). For the remaining 4 with an α of 0.65 ("presentation style: other," "nondiagnostic symptom: physical signs," "mention of daily occurrence," and "emotional appeal"), agreement among coders was 95%. Because of this high level of coder agreement but low Krippendorff α levels, Gwet AC1 (agreement coefficient 1) was calculated as an additional reliability coefficient to estimate expected agreement for these variables, as suggested by Riffe and colleagues [23]. All variables reached Gwet AC1 of 0.95, indicating excellent reliability. Consequently, these variables were retained. Table A1 in Open Science Framework [25] details all reliability statistics.

Statistical Analyses

To examine whether TikTok videos created by medical professionals scored higher on diagnostic accuracy than those created by nonprofessionals (H1a), a nonparametric Mann-Whitney *U* test was conducted with creator type as the independent variable and the diagnostic accuracy score (ranging from 0 to 4) as the dependent variable. A nonparametric test was chosen instead of a *t* test because Shapiro-Wilk tests revealed that the diagnostic accuracy score was not normally distributed (medical professionals: $W(72)=0.886$, $P<.001$; nonprofessionals: $W(138)=0.857$, $P<.001$). Since the a priori power analysis was based on a *t* test, a post hoc power analysis using G*Power 3.1 was conducted for nonparametric testing, revealing 99.2% of power to detect medium effects ($\alpha=0.05$, $w=0.30$, n=210; [31]). Because Mann-Whitney *U* tests compare the distribution of ranks rather than means, results are reported in terms of the median diagnostic accuracy score.

To test whether a lower diagnostic accuracy score predicted higher user engagement (H1b), a hierarchical regression analysis was conducted. The independent variable was the diagnostic accuracy score, and the dependent variable was the engagement score. The engagement score was log-transformed to correct extreme positive skew and reduce the influence of outliers, a recommended practice for highly skewed data [32]. All assumptions for hierarchical regression (normality, homoscedasticity, and absence of multicollinearity) were checked and met. In the first step, we entered the control variables creator type (medical professional vs nonprofessional) and the 4 most common presentation styles (direct-to-camera formats, text-based videos, acted-out scenes, and animations). In the second step, the diagnostic accuracy score was added to the model to examine its unique association with engagement above and beyond these controls.

To assess whether TikTok videos created by medical professionals feature fewer personal experiences (H2a), emotional appeals (H2b), and background music (H3a) than those created by nonprofessionals, 3 chi-square tests of independence were conducted. To examine whether the presence of personal experiences (H2c), emotional appeals (H2d), and background music (H3b) leads to higher engagement, 3 hierarchical regression analyses were conducted. The independent variables were the presence (vs absence) of (1) personal experiences, (2) emotional appeals, and (3) background music. The dependent variable was the log-transformed engagement score. Creator type and presentation styles were again entered as control variables as a first step, followed by the focal stylistic predictor in the second step.

Results

Descriptive Statistics

Depression Symptom Portrayals

Each of the 210 TikTok videos was coded for the presence of diagnostic and nondiagnostic depression symptoms. The most common diagnostic symptoms were "depressed mood" (n=131, 62.4%), "reduced energy or fatigue" (n=82, 39%), and "changes in sleep" (n=66, 31.4%). Nondiagnostic symptoms—those not included in the *ICD-11*—were also common, with 134 (63.8%) videos referencing at least one. The 3 most mentioned nondiagnostic symptoms were "putting on a front" (n=45, 21.4%), "social isolation" (n=30, 14.3%), and "irritability" (n=22, 10.5%). Distributions of diagnostic and nondiagnostic symptoms are presented in Tables 1 and 2, respectively.

XSL•FO
RenderX

Table . Distribution of *ICD-11*[a] diagnostic symptoms in TikTok videos about depression (N=210)[b].

ICD-11 symptoms	Videos, n (%)
Depressed mood	131 (62.4)
Reduced energy or fatigue	82 (39.0)
Changes in sleep	66 (31.4)
Hopelessness	61 (29.0)
Feelings of worthlessness	58 (27.6)
Changes in appetite	53 (25.2)
Loss of interest	53 (25.2)
Difficulty concentrating	37 (17.6)
Suicidal thoughts	33 (15.7)
Psychomotor agitation	11 (5.2)

[a]*ICD-11: International Classification of Diseases, 11th Revision*.

[b]The table shows the number and proportion of TikTok videos in which each *ICD-11* diagnostic symptom of depression was portrayed, based on a content analysis of English-language TikTok videos.

Table . Distribution of nondiagnostic symptoms in TikTok videos about depression (N=210)[a].

Symptom	Videos, n (%)
Putting on a front	45 (21.4)
(Social) isolation	30 (14.3)
Irritability	22 (10.5)
Lack of self-care	19 (9.0)
Loneliness	19 (9.0)
Media distractions	14 (6.7)
Low motivation	13 (6.2)
Body pains	13 (6.2)
Drug (ab)use	12 (5.7)
Messy room	11 (5.2)

[a]The table shows the number and proportion of TikTok videos in which each nondiagnostic symptom (ie, symptoms not recognized in *International Classification of Diseases, 11th Revision* (*ICD-11*) criteria for depressive episodes) was portrayed, based on a content analysis of English-language TikTok videos. The table lists the 10 most prevalent nondiagnostic symptoms; other nondiagnostic symptoms coded for occurred in <5% of videos.

Diagnostic Accuracy of TikTok Videos About Depression

Each video was evaluated against 4 diagnostic accuracy criteria. To meet the first criterion, a video had to depict at least one core depression symptom and at least one additional symptom. In total, 111 (52.9%) TikTok videos met this criterion. The second criterion, referring to the daily occurrence of symptoms, was met in 42 (20%) videos. The third criterion, requiring symptoms to persist for at least 2 weeks, held for 25 (11.9%) videos. The fourth criterion, absence of nondiagnostic symptom mentions, was met in 76 (36.2%) videos. Overall, 59 (28.1%) videos met none of the criteria, 78 (37.1%) met exactly 1 criterion, 50 (23.8%) met 2 criteria, 16 (7.6%) met 3 criteria, and only 7 (3.3%) met all 4 criteria. The total diagnostic accuracy score ranged from 0 to 4 with a mean of 1.21 (SD 1.04), indicating that, on average, videos met just over one diagnostic criterion.

Creator and Stylistic Features of Depression-Related TikTok Videos

In terms of creator type, 138 (65.7%) TikTok videos were posted by creators without any indication of a professional medical background, while 72 (34.3%) were created by medical professionals. With respect to the key emotional stylistic features examined, 174 (82.9%) videos included background music, 164 (78.1%) featured emotional appeals, and 80 (38.1%) included personal experiences. The most common presentation style was in the form of creators speaking directly to the camera (n=66, 31.4%), followed by text-based presentations (n=62, 29.5%), acted-out scenes (n=41, 19.5%), and animation (n=19, 9%).

Engagement Metrics of TikTok Videos About Depression

All individual and total engagement metrics are visualized in Table 3. On average, the total engagement score of a TikTok video about depression in our sample was 260,880 (SD 609,338). All engagement variables showed substantial positive

XSL•FO
RenderX

skewness, indicating that a few outliers achieved disproportionately high numbers of engagement.

Table . Engagement metrics of TikTok videos about depression (N=210)[a].

	Descriptive statistics			
Engagement metrics	Min	Max	Mean (SD)	Skewness
Likes	2	4,200,000	220,431 (500,476)	4.19
Comments	0	31,900	1726 (3941)	4.62
Saves	0	967,100	29,556 (88,337)	7.36
Shares	0	430,500	9167 (33,978)	9.79
Engagement	2	5,622,300	260,880 (609,338)	4.84

[a]Descriptive statistics are based on a content analysis of English-language TikTok videos about depression. Engagement is operationalized as the sum of likes, comments, saves, and shares. Skewness indicates the asymmetry of the distribution of each engagement metric. A skewness value above 1 suggests a highly right-skewed distribution.

Hypothesis Testing

H1a: Creator Type and Diagnostic Accuracy

The median diagnostic accuracy score for medical professionals was higher (median 1.40, IQR 1-2) than for nonprofessionals (median 1.11, IQR 0-2); however, the difference was not statistically significant (U=4208, z=−1.90; P=.06). Given the overall low diagnostic accuracy score across the TikTok videos (mean 1.21, SD 1.04), an exploratory analysis was conducted to examine whether medical professionals (n=72) and nonprofessionals (n=138) significantly differed in how they fulfilled each of the 4 individual diagnostic criteria. The first criterion—mentioning at least one core and one additional symptom—was met in 49 (68.1%) videos created by medical professionals, and in 62 (44.9%) videos created by nonprofessionals, revealing a significant difference (U=3819; z=−3.18; P=.001). The second criterion, mentioning daily occurrence of symptoms, was mentioned in 14 (19.4%) professional and 28 (20.3%) nonprofessional videos, revealing no significant difference (U=4926; z=−0.15; P=.89). The third criterion—mentioning symptom duration—was proportionally slightly more often fulfilled by medical professionals (n=12, 16.7%) than by nonprofessionals (n=13, 9.4%), but this difference was not significant (U=4608; z=−1.54; P=.13). Lastly, the fourth criterion, not mentioning nondiagnostic symptoms, was met in 26 (36.1%) medical professional videos and in 50 (36.2%) nonprofessional videos (U=4962; z=−0.02; P=.99).

H1b: Influence of Diagnostic Accuracy on Engagement

To examine whether diagnostic accuracy predicted engagement, we conducted a hierarchical linear regression analysis. The first model—with only the control variables creator type and presentation style—was statistically significant ($F_{5,204}$=4.04; P=.002), explaining 9% of the variance in engagement (R^2=0.09). Adding diagnostic accuracy in the second model did not significantly improve model fit (ΔR^2=0.008; $\Delta F_{1,203}$=1.77; P=.19), and diagnostic accuracy was not a significant predictor of engagement (B=−0.10, SE 0.08; P=.19). However, 2 control variables were associated with significantly lower engagement: videos featuring creators speaking directly to the camera (B=−0.62, SE 0.23; P=.008) and videos centered around text (B=−0.65, SE 0.21; P=.003). This highlights that presentation style is a distinct factor influencing engagement.

H2a, H2b, and H3a: Creator Type and Use of Stylistic Features

The use of personal experiences differed by creator type (χ^2_1[n=210]=41.15, P<.001). TikTok videos created by medical professionals contained significantly fewer personal experiences (n=6, 8.3%) compared to nonprofessionals (n=74, 53.6%). The difference was moderate in strength (ϕ=0.44) [31]. The use of emotional appeals also differed across creator types (χ^2_1[n=210]=25.01, P<.001). Emotional appeals appeared in 42 (58.3%) TikTok videos created by medical professionals and in 122 (88.4%) nonprofessional TikTok videos, another moderate difference (ϕ=0.35) [31]. Finally, the third chi-square revealed a significant difference in the use of background music (χ^2_1[n=210]=11.15, P<.001), with a small effect size (ϕ=0.23) [31]. Background music was less often used in medical professional TikTok videos (n=51, 70.8%) compared to nonprofessional TikTok videos (n=123, 89.1%).

H2c, H2d, H3b: Influence of Stylistic Features on Engagement

To test whether personal experiences (H2c), emotional appeals (H2d), or background music (H3b) were associated with engagement, 3 hierarchical regression analyses were conducted. Adding personal experiences to the model resulted in a significant increase in explained variance (ΔR^2=0.024, P=.02), indicating that videos with personal experiences achieved higher engagement (B=0.41, SE 0.18; P=.02). Similarly, adding emotional appeals significantly improved the model fit (ΔR^2=0.059, P<.001), also positively predicting engagement (B=0.73, SE 0.19, P<.001). Finally, the inclusion of background music also resulted in a significant increase in explained variance (ΔR^2=0.027, P=.01), with background music positively predicting engagement (B=0.54, SE 0.22; P=.01). Across all models, videos featuring creators speaking directly to the camera (Bs ranging from −0.49 to −0.69; .003≤P≤.04) and videos centered around text (Bs ranging from −0.56 to −0.64; .002≤P<.01) consistently predicted lower engagement.

XSL•FO
RenderX

Discussion

Summary

This study analyzed 210 TikTok videos about depression and found that, overall, diagnostic accuracy was low, with no significant differences between content created by medical professionals and nonprofessionals. Lower diagnostic accuracy itself was not associated with higher engagement. Rather, videos featuring personal experiences, emotional appeals, and background music—stylistic elements more commonly found in nonprofessional content—were found to be the drivers behind higher engagement. By contrast, more static formats such as direct-to-camera explanations and text-centered videos were associated with lower engagement. Altogether, these findings indicate that engagement with depression-related TikTok videos is dependent on stylistic elements and emotional appeals, rather than diagnostic accuracy.

Diagnostic Accuracy of Depression Videos on TikTok

Contrary to our hypothesis, lower diagnostic accuracy was not a significant predictor of higher engagement, challenging prior research on mental health content on TikTok [6,9,10]. While this at first presents a promising finding, it is important to highlight that the diagnostic accuracy was low across the entire sample. Specifically, although around half of the TikTok videos met the first diagnostic accuracy criterion by mentioning at least one core and one additional diagnostic symptom, the other 3 criteria—referencing daily occurrence, a duration of at least 2 weeks, and avoiding nondiagnostic symptoms—were rarely fulfilled. Importantly, diagnostic accuracy in this study reflects alignment with established clinical diagnostic criteria rather than an evaluation of the authenticity or validity of lived experiences shared on the platform. As such, low diagnostic accuracy should be interpreted as limited diagnostic completeness, not as an invalidation of personal experiences with distress or depression.

At the same time, limited diagnostic completeness may contribute to concept creep [11], whereby the meaning of clinical terms such as “depression” gradually expands to encompass a broader range of experiences. When diagnostic boundaries become diffuse, this may increase the likelihood of self-diagnosing based on incomplete information. Moreover, an experimental study in the context of attention-deficit/hyperactivity disorder (ADHD) misinformation on TikTok found that misinformed emerging adults had higher confidence in their ADHD knowledge and higher treatment-seeking intentions than emerging adults who were not exposed to such misinformation [33]. This highlights that inaccurate content can create a false sense of understanding and may add pressure on already overstretched mental health services. It is crucial for future research to assess how exposure to varying levels of diagnostic accuracy in depression portrayals on TikTok affects self-diagnosing behaviors and concept creep.

Diagnostic Accuracy of Medical Professionals vs Nonprofessional Creators

Strikingly, the low diagnostic accuracy also characterized depression portrayals by medical professionals, whose videos were not significantly more accurate than those of nonprofessionals. One possible explanation for this finding is that TikTok’s fast-paced and short-form video format encourages even professionals to simplify their content and incorporate nondiagnostic, more relatable symptom portrayals. In doing so, they may prioritize accessibility over diagnostic depth, omitting key criteria for clarity or conciseness. At the same time, however, our findings show that medical professionals relied less on implementing personal experiences, emotional appeals, and background music in their content compared to nonprofessionals. This aligns with previous findings in the mental health space [5,7] and suggests that medical professionals may simplify *what* they communicate without fully adapting *how* they communicate to the platform’s engagement norms. Future research should specifically examine why so many medical professionals choose to add nonclinical symptom portrayals and omit essential *ICD-11* criteria when adapting their messages for TikTok, and how they balance clinical responsibility with platform-specific communication norms. In-depth interviews or focus groups with medical professionals can add to a better understanding of how they adapt to the platform.

While the fact that medical professionals did not produce more diagnostically accurate content may be understandable in light of their efforts to reduce cognitive load, it raises serious concerns from the perspective of the ELM of persuasion. Medical professionals hold source credibility, a powerful peripheral cue in the ELM that can have a strong influence on persuasion and trust [34]. As such, the disconnect between diagnostic accuracy and medical professional status found in this study is problematic, as users may be persuaded by the authority of the source, even when the message lacks diagnostic accuracy. In a platform environment where users rely heavily on heuristics [20], the persuasive power of credibility without accuracy may contribute to the spread of misleading or oversimplified information.

How Stylistic Features Drive Engagement

Emotional appeals, personal experiences, and background music were all common features of the TikTok videos in our sample. In line with our hypotheses, the presence of each of these features individually predicted significantly higher engagement, supporting earlier studies about mental health content on TikTok [2,5-7,10]. These findings are also in line with identification theory [21], which suggests that users may become more engaged when they can emotionally and psychologically align with a media persona.

This study’s findings further align with notions from transportation theory [22]. Not only did the presence of background music increase engagement, but among the 4 most common presentation styles that were tested across our analyses, direct-to-camera and text-based formats consistently led to lower engagement. This suggests that purely informational or static formats may not be sufficient to capture users’ attention on a highly dynamic platform like TikTok, regardless of whether the creator is a medical professional or not. Future research could explore these dynamics more directly by using experimental designs to assess whether variations in background

XSL•FO
RenderX

music and presentation style influence users' sense of transportation and therefore engagement with the content.

While diagnostic accuracy did not predict engagement patterns and was low overall, it remains unclear whether higher diagnostic accuracy would undermine or enhance the effectiveness of stylistic features. Importantly, the danger of heightened identification and transportation lies in the possibility that users may connect more deeply with diagnostically inaccurate content. However, as young people increasingly rely on online resources and information as part of their mental health help-seeking process [35], high diagnostic accuracy of such resources could serve a more constructive role and act as a stepping stone toward seeking offline help [36]. In this context, TikTok's engaging and accessible format positions it as a particularly promising platform for mental health communication [37]. Future experimental research should examine whether stylistic features still lead to increased engagement when combined with accurate symptom portrayals of depression.

Practical Implications

The majority of TikTok videos about depression in our sample were posted by creators without any indication of a professional medical background and showed very low diagnostic accuracy. Although such videos may not be intended to provide diagnostic guidance, they are frequently encountered by users seeking information about depression and may shape how viewers interpret their own experiences, draw conclusions about whether their symptoms "count" as depression, and influence help-seeking decisions. This underscores the need for clearer contextual cues to support the interpretation of mental health–related content. Platforms such as TikTok could consider introducing labels or informational prompts that clarify whether content reflects professional medical information and direct users to authoritative resources. Such measures could help reduce misinterpretation without restricting creators' ability to share lived experiences.

However, directing users toward professional content is only beneficial if such content is diagnostically accurate. Yet, in our sample, medical professionals frequently mentioned nondiagnostic symptoms and omitted the daily occurrence criterion at rates similar to nonprofessionals. To improve the quality of their TikTok content, medical professionals should avoid vague symptom portrayals and place greater emphasis on clearly communicating key clinical criteria. At the same time, medical professionals can enhance engagement by incorporating stylistic features such as background music, emotionally appealing framing, or patient-informed narratives that include personal experiences, all without compromising clinical integrity. Importantly, they should also avoid overly static formats and information-only presentation styles, as these were associated with lower engagement. Future research should experimentally test how medical professional content that integrates both high diagnostic accuracy and stylistically engaging features performs in terms of engagement on TikTok. Such work could guide the development of evidence-based best practices for professional mental health communication on TikTok.

Limitations

Although this study offers novel insights into the diagnostic accuracy and stylistic features of TikTok videos about depression by developing an objective and replicable diagnostic coding tool for evaluating depression symptom portrayals, several limitations should be noted. First, regarding the sample, while the included videos were posted between February 2020 and April 2025, most dated to the period between 2022 and 2024. As TikTok's platform affordances, content norms, and mental health discourse evolve rapidly, the findings should be interpreted in light of this temporal context. In addition, the sample size was necessarily limited given the labor-intensive nature of manual content analysis. Future research could build on the diagnostic accuracy coding framework developed in this study by applying automated content analysis approaches to examine depression portrayals at a much larger scale.

Second, the *ICD-11* criteria may have been overly strict given the nature of TikTok as an entertainment-oriented platform. Creators, whether medical professionals or not, often aim to simplify complex topics to fit the platform's short-form video format, thereby omitting certain *ICD-11* criteria. A more flexible or scaled application of *ICD-11* criteria might better capture how diagnostic elements are adapted for such contexts in future studies. Third, while the inclusion of more detailed coding in terms of valence (ie, distinguishing between a positive and negative emotional appeal, coding the emotional valence of the background music) was intended, low ICR prevented its implementation. Future research should leverage automated content ‐ analysis methods, such as audio valence detection models that can reliably classify the emotional tone of background music, and natural language processing algorithms to more precisely identify and quantify specific emotional appeals.

Lastly, engagement was measured only in terms of active interactions (likes, comments, shares, and saves), not passive reach (views). Although reach data was available and could have been added to the engagement score, it was excluded because, in line with prior research, engagement was conceptualized as a form of user interaction that reflects an active decision to respond to content, rather than mere exposure [7,10]. Nevertheless, including reach metrics in future research could offer a more comprehensive understanding of content visibility and its potential influence.

Conclusions

Our study shows that depression content on TikTok is often low in diagnostic accuracy, both among medical professional creators and nonprofessional creators. Although this low diagnostic accuracy itself did not significantly drive engagement, stylistic features did: the presence of personal experiences, emotional appeals, and background music was each associated with higher engagement. By contrast, we found that videos from medical professionals and more static, information-heavy formats drew less engagement. These patterns raise concerns about concept creep and potential premature self-diagnosis, but they also point to a clear opportunity: clinically accurate information can be made more compelling if professionals adopt platform-native content styles. Practically, this means

XSL•FO
RenderX

foregrounding clear symptom criteria alongside human-centered, emotionally resonant presentation and sensory cues. Future research has the opportunity to develop evidence-based best practices for professional mental health communication on short-form video platforms. TikTok, if used thoughtfully, can be part of—and not a barrier to—credible, engaging depression education for young audiences.

Acknowledgments

We are grateful to Lucia Gheorghita for her assistance with the intercoder reliability testing.

Funding

No external financial support or grants were received for this project.

Data Availability

All relevant data and syntax can be found on the Open Science Framework [25].

Authors' Contributions

Formal analyses: ER, AvdW
Project administration: AvdW
Supervision: AvdW
Writing – original draft: ER
Writing – review & editing: ER, AvdW, IB

Conflicts of Interest

None declared.

References

1. Faverio M, Anderson M, Park E. Teens, social media and mental health. Pew Research Center. 2025 Apr 22. URL: https://www.pewresearch.org/internet/2025/04/22/teens-social-media-and-mental-health/ [accessed 2026-03-27]
2. Basch CH, Donelle L, Fera J, Jaime C. Deconstructing TikTok videos on mental health: cross-sectional, descriptive content analysis. JMIR Form Res 2022 May 19;6(5):e38340. [doi: 10.2196/38340] [Medline: 35588057]
3. Tudehope L, Sofija E, Harris N. VentTok: exploring the mental health narrative on TikTok. Stigma Health 2024. [doi: 10.1037/sah0000577]
4. Claps EV, Li A, Conrow A, Schiess N, Mullins R, Swift JK. TikTok style videos and mental health: effects on stigma, self-pathologizing, and help-seeking. Curr Psychol 2025 Apr;44(8):7082-7094. [doi: 10.1007/s12144-025-07716-z]
5. Jerin SI, O'Donnell N, Mu D. Mental health messages on TikTok: analysing the use of emotional appeals in health-related #EduTok videos. Health Educ J 2024 Jun;83(4):395-408. [doi: 10.1177/00178969241235528]
6. Turuba R, Zenone M, Srivastava R, et al. Do you have depression? A summative content analysis of mental health-related content on TikTok. Digit Health 2025;11. [doi: 10.1177/20552076241297062] [Medline: 39839955]
7. Lau N, Srinakarin K, Aalfs H, Zhao X, Palermo TM. TikTok and teen mental health: an analysis of user-generated content and engagement. J Pediatr Psychol 2025 Jan 1;50(1):63-75. [doi: 10.1093/jpepsy/jsae039] [Medline: 38950415]
8. Samuel L, Kuijpers K, Bleakley A. TherapyTok for depression and anxiety: a quantitative content analysis of high engagement TikTok videos. J Adolesc Health 2024 Jun;74(6):1184-1190. [doi: 10.1016/j.jadohealth.2024.02.002] [Medline: 38493396]
9. Verma S, Sinha SK. How evidence-based is the "hashtag ADHD test" (#adhdtest). A cross-sectional content analysis of TikTok videos on attention-deficit/hyperactivity disorder (ADHD) screening. Australas Psychiatry 2025 Feb;33(1):82-88. [doi: 10.1177/10398562241291956] [Medline: 39422639]
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]
11. Haslam N, Dakin BC, Fabiano F, et al. Harm inflation: making sense of concept creep. Eur Rev Soc Psychol 2020 Jan 1;31(1):254-286. [doi: 10.1080/10463283.2020.1796080]
12. Speerforck S, Jürgensen V, Göbel M, Meyer N, Schomerus G. "Concept creep" in perceptions of mental illness - an experimental examination of prevalence-induced concept change. Eur Arch Psychiatry Clin Neurosci 2026 Feb;276(1):33-37. [doi: 10.1007/s00406-023-01737-0] [Medline: 38233638]
13. Foster A, Ellis N. TikTok-inspired self-diagnosis and its implications for educational psychology practice. Educ Psychol Pract 2024 Oct;40(4):491-508. [doi: 10.1080/02667363.2024.2409451]
14. Hall R, Keenan R. More than half of top 100 mental health TikToks contain misinformation, study finds. The Guardian. 2025 May 31. URL: https://www.theguardian.com/society/2025/may/31/more-than-half-of-top-100-mental-health-tiktoks-contain-misinformation-study-finds [accessed 2026-04-10]

XSL•FO
RenderX

15. Thapar A, Eyre O, Patel V, Brent D. Depression in young people. Lancet 2022 Aug 20;400(10352):617-631. [doi: 10.1016/S0140-6736(22)01012-1] [Medline: 35940184]
16. Li J, Tang L, Pu Y. My story of depression: a content analysis of autobiographic videos on Douyin. Health Commun 2024 Apr 15;39(5):906-914. [doi: 10.1080/10410236.2023.2191887]
17. Mittmann G, Schrank B, Steiner-Hofbauer V. A scoping review about the portrayal of depression and anxiety in mainstream and social media. Int J Psychol 2024 Dec;59(6):1075-1083. [doi: 10.1002/ijop.13232] [Medline: 39164881]
18. Aragon-Guevara D, Castle G, Sheridan E, Vivanti G. The reach and accuracy of information on autism on TikTok. J Autism Dev Disord 2025 Jun;55(6):1953-1958. [doi: 10.1007/s10803-023-06084-6] [Medline: 37544970]
19. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. In: Berkowitz L, editor. Advances in Experimental Social Psychology: Academic Press; 1986:123-205. [doi: 10.1016/S0065-2601(08)60214-2]
20. Barta S, Belanche D, Fernández A, Flavián M. Influencer marketing on TikTok: the effectiveness of humor and followers' hedonic experience. J Retail Consum Serv 2023 Jan;70:103149. [doi: 10.1016/j.jretconser.2022.103149]
21. Cohen J. Defining identification: a theoretical look at the identification of audiences with media characters. Mass Commun Soc 2001 Aug;4(3):245-264. [doi: 10.1207/S15327825MCS0403_01]
22. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 2000;79(5):701-721. [doi: 10.1037/0022-3514.79.5.701]
23. Riffe D, Lacy S, Watson BR, Lovejoy J. Analyzing Media Messages: Using Quantitative Content Analysis in Research, 5th edition: Routledge. [doi: 10.4324/9781003288428]
24. Li B, Liu M, Liu J, Zhang Y, Yang W, Xie L. Quality assessment of health science-related short videos on TikTok: a scoping review. Int J Med Inform 2024 Jun;186:105426. [doi: 10.1016/j.ijmedinf.2024.105426]
25. Rainer E, Wal A, Beyens I. Depression portrayals on TikTok: engagement patterns as a function of diagnostic accuracy, stylistic features and creator type. Open Science Framework. 2026 Mar 1. URL: https://osf.io/ycjvh/overview [accessed 2026-03-27]
26. 6A70.0 Single episode depressive disorder, mild. ICD-11 for Mortality and Morbidity Statistics. 2024. URL: https://icd.who.int/browse/2024-01/mms/en#1922675618 [accessed 2026-03-27]
27. Cuncic A. What is the ICD-11? Verywell Mind. 2026 Feb 15. URL: https://www.verywellmind.com/overview-of-the-icd-11-4589392 [accessed 2026-03-27]
28. De Swert K. Calculating inter-coder reliability in media content analysis using Krippendorff's alpha. : Center for Politics and Communication, University of Amsterdam; 2012 Feb 1 URL: https://www.polcomm.org/wp-content/uploads/ICR01022012.pdf [accessed 2026-03-27]
29. Lacy S, Watson BR, Riffe D, Lovejoy J. Issues and best practices in content analysis. J Mass Commun Q 2015 Dec;92(4):791-811. [doi: 10.1177/1077699015607338]
30. Krippendorff K. Content Analysis: An Introduction to Its Methodology, 3rd edition: SAGE Publications; 2019. [doi: 10.4135/9781071878781]
31. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 2013. [doi: 10.4324/9780203771587]
32. West RM. Best practice in statistics: the use of log transformation. Ann Clin Biochem 2022 May;59(3):162-165. [doi: 10.1177/00045632211050531] [Medline: 34666549]
33. Schiros A, Bowman N, Antshel K. Misinformation mayhem: the effects of TikTok content on ADHD knowledge, stigma, and treatment-seeking intentions. Eur Child Adolesc Psychiatry 2025 Nov;34(11):3521-3533. [doi: 10.1007/s00787-025-02769-8] [Medline: 40471415]
34. Hong Y, Lee N, Kirkpatrick CE, Hu S, Lee S, Hinnant A. "Trust me, I'm a doctor." How TikTok videos from different sources influence clinical trial participation. Health Commun 2025 Feb 23;40(3):417-428. [doi: 10.1080/10410236.2024.2346680]
35. Pretorius C, Chambers D, Cowan B, Coyle D. Young people seeking help online for mental health: cross-sectional survey study. JMIR Ment Health 2019 Aug 26;6(8):e13524. [doi: 10.2196/13524] [Medline: 31452519]
36. Rutter LA, Howard J, Lakhan P, Valdez D, Bollen J, Lorenzo-Luaces L. "I haven't been diagnosed, but I should be"—insight into self-diagnoses of common mental health disorders: cross-sectional study. JMIR Form Res 2023 Jan 13;7(1):e39206. [doi: 10.2196/39206] [Medline: 36637885]
37. Comp G, Dyer S, Gottlieb M. Is TikTok the next social media frontier for medicine? AEM Educ Train 2021 Jul;5(3). [doi: 10.1002/aet2.10532] [Medline: 34095694]

Abbreviations

AC1: agreement coefficient 1
ADHD: attention-deficit/hyperactivity disorder
ELM: elaboration likelihood model
ICD-11: International Classification of Diseases, 11th Revision
ICR: intercoder reliability

XSL•FO
RenderX

Edited by T Mackey; submitted 05.Oct.2025; peer-reviewed by GA Alvares, L Samuel; accepted 02.Mar.2026; published 13.Apr.2026.

Please cite as:
Rainer E, van der Wal A, Beyens I
Portrayals of Depression on TikTok: Content Analysis of Diagnostic Accuracy, Creator Type, and Stylistic Features
JMIR Infodemiology 2026;6:e85323
URL: https://infodemiology.jmir.org/2026/1/e85323
doi:10.2196/85323

© Elena Rainer, Amber van der Wal, Ine Beyens. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 13.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Nonnegotiable Symbolic Value and Sugar-Driven Food Habits in Indonesia: Mixed Methods Study Using a Digital Sociological Approach

Ewina Efriani Manik*, MSc; Sudarsono Hardjosoekarto*, Prof Dr; Ricardi Adnan*, Prof Dr; Radhiatmoko Radhiatmoko*, PhD; One Herwantoko*, PhD; Hakiki Nurmajesty, MSc; Darwan Darwan, MSc; Farrah Eriska Putri, MArch; Astuti Sri Pawening, MSc

Department of Sociology, Faculty of Social and Political Science, University of Indonesia, Pondok Cina, Depok, Indonesia

*these authors contributed equally

Corresponding Author:
Sudarsono Hardjosoekarto, Prof Dr
Department of Sociology, Faculty of Social and Political Science, University of Indonesia, Pondok Cina, Depok, Indonesia

Abstract

Background: The sugar market in Indonesia reflects the distinct consumer behavior shaped by economic and deeply rooted cultural factors. This study explores how symbolic values attached to sugar sustain persistent, often irrational or uncontrollable consumption, highlighting the need for a demand-side perspective in the economic sociology of sugar markets.

Objective: This study analyzes the nonnegotiable symbolic value of sugar and its implication to uncontrollable consumption in Indonesia. Referring to the framework of product valuation in the social order of markets by Beckert, it offers insights into both the symbolic and material values of sugar.

Methods: The applied method complements digital mixed method approaches used in prior research. Digital data from online news and YouTube were visualized through textual network analysis and social network analysis to describe the symbolic and material values of sugar. In-depth interviews with key actors and limited field observations on food and beverage labels were also conducted.

Results: Findings reveal that the symbolic value of sugar increases significantly when processed into food or beverages, shaping food habits and habitus across diverse ethnic groups in Indonesia and reinforcing early dependence on sugar. Weak enforcement of labeling regulations on food and beverage packages further impedes shifts in consumer perceptions of the risks of excessive sugar consumption.

Conclusions: This study contributes a demand-side perspective to the economic sociology of the sugar market, proposing strategies to address the sugar-driven food habits and habitus from the perspective of consumer behavior. Simultaneously, it assesses producer compliance with regulations on the sweetness level to reduce sugar consumption and the prevalence of noncommunicable diseases.

(JMIR Infodemiology 2026;6:e77261) doi:10.2196/77261

KEYWORDS

economic sociology; sugar market; symbolic value; uncontrollable consumption; digital research

Introduction

Background

This study applies an economic sociology perspective to public health, particularly in controlling sugar consumption in Indonesia. In both traditional and modern contexts, sugar is not merely a sweetener for foods and beverages but also a cultural symbol of harmony, happiness, and social status, reproduced through food habits and habitus from an early age [1]. Economic growth, urbanization, lifestyle changes, and the promotion of sugar-based products have intensified sugar consumption, leading to adverse health outcomes. This prompts questions about the effect of uncontrolled sugar consumption on public health and the symbolic value of sugar on market demand in Indonesia.

The surge is particularly evident among Generation Z, for whom sweetened beverages (ie, bubble tea, milk coffee, and soft drinks) serve not merely to quench thirst but also to express pleasure, social identity, and self-expression as part of urban lifestyle. Sugar thus embodies a dual role: materially as a sweetener and symbolically as a powerful cultural element with a nonnegotiable symbolic value.

XSL•FO
RenderX

This symbolic value is reinforced through everyday practices and popular culture promoted by the media. Exchanging chocolates on Valentine's Day or advertisements declaring "three Moo-Moo candies equal one glass of milk" and "there is a hint of sweetness to it" associate sweetness with affection and pleasure [2]. These messages sustain the position of sugar as a symbol of intimacy and happiness, fueling the demand for sugary and processed foods and beverages heavily promoted through print and television media [3].

From a public health perspective, excessive sugar consumption, aggravated by sedentary lifestyles, raises the risk of type 2 diabetes at a young age. Symptoms such as fatigue, excessive thirst, and slow-healing wounds often go unnoticed due to low public awareness of blood sugar monitoring [4]. Ironically, the cultural association of sweetness with enjoyment obscures these risks and burdens the national health care system.

From an economic sociology perspective, the sugar market is shaped not only by prices and regulations but also by social and cultural constructions of value. As Beckert [5] argues, product valuation in markets involves symbolic dimensions constructed through social and cognitive processes among market actors. While the World Health Organization (WHO) recommends limiting sugar to 25 grams per person (no more than 5 teaspoons) [6], Indonesia remains setting its national limit at 50 grams, with imports calculated on an assumption of 62 grams per person per day. This indicates a high dependency on sugar in both households and the food and beverage industry. Between 2020 and 2025, the World Bank discloses a 3.7% rise in sugar consumption in Indonesia, maintaining its position as the primary sugar importer since the 1980s [7].

Globally, such consumption patterns fuel noncommunicable diseases (NCDs), as observed in the United Kingdom. Since food consumption behavior is influenced by purchasing decisions, regulatory interventions should target not only nutritional and economic dimensions but also the symbolic and cultural factors shaping consumption patterns [8]. This study thus examines how the symbolic value of sugar influences market demand and public health in Indonesia through textual network analysis (TNA) and social network analysis (SNA).

Objectives

This study explores 2 key aspects of the sugar market in Indonesia. First, it examines consumer behavior in reducing food habits and habitus shaped by sugar consumption. Second, it investigates producer compliance with regulations controlling sugar levels in foods and beverages.

To analyze the nonnegotiable symbolic value of sugar, this study adopted a network-based analytical approach integrating TNA and SNA. This approach is consistent with Hong and Lee [9], who used TNA to map long-term discourse structures and knowledge trends, and with Leem et al [10], who conceptualized TNA as a data-driven method for systematically uncovering conceptual structures and semantic relationships within text corpora. The application of SNA follows the framework proposed by Paterson et al [11], which emphasizes its role in identifying dominant actors, power relations, and policy dynamics within organizational networks.

The analysis was conducted in 5 stages. First, TNA was used to visualize the material value of sugar as reflected in discourses of standardization and regulation. Second, TNA captured the symbolic value of sugar by classifying patterns of cognitive anchoring and social positioning in public discourse. Third, SNA mapped dominant actors and network relations involved in policymaking on sweetness levels in packaged foods and beverages. Fourth, market observations were taken to assess compliance with sugar-content labeling. Fifth, all findings were synthesized to explain the persistence of sugar consumption, driven not only by material considerations or health information, but also by the socially and culturally institutionalized symbolic value of sugar.

Overall, this study addresses 3 main questions: First, how is the material value of sugar perceived, and how does it affect public health negatively? Second, how is the symbolic value of sugar socially and culturally constructed, leading to excessive consumption? Third, what sociological interventions can reduce sugar consumption while addressing its entrenched nonnegotiable symbolic value?

Literature Review

The Construction of Symbolic and Material Value of Goods in Economic Sociology

Market sociology, as introduced by Beckert [5,12,13], mainly focuses on how product value and quality are created. Beckert [5] argues that "the more the value of products becomes detached from the fulfillment of purely functional needs, the more they depend upon symbolic assignments of value that must be constructed by market actors." Consequently, understanding both symbolic and material qualities is important for meaningful market exchanges.

Several studies emphasize that material value remains relevant for predicting consumer preferences and brand choices. Similarly, other studies note that businesses enhance customer experiences by aligning products with consumer expectations [14] .

Quoting Durkheim, Beckert [15] posits that "value emerges from the symbolic connections made between goods and the socially rooted values." Beckert further explains that "symbolic value is value from the symbolic meaning of objects." It suggests that numerous goods exchanged in the market derive their primary value from symbolic meanings rather than physical qualities. Consumers consume products for their symbolic meanings rather than material use [14].

In this study, the symbolic value of sugar can be understood through diverse cultural contexts. Among Javanese communities, sweet foods signify harmony, gratitude, and positive social relations [1]. These values are reflected in everyday consumption practices and transmitted through food habitus from an early age. However, public health studies indicate that excessive sugar intake is a major risk factor for NCDs, including diabetes and obesity [16,17].

Beckert [5] presents the social order of markets framework, which theorizes that product valuation is socially and culturally patterned through 4 aspects: standardization, cognitive

XSL•FO
RenderX

anchoring, normative legitimacy, and social positioning. Standardization appears in formal regulations such as the Regulation of the Ministry of Health (Permenkes) No. 30/2013 on daily sugar intake limits. Cognitive anchoring captures the cognitive and cultural associations between sweetness and togetherness in Javanese culinary traditions. Social positioning highlights the roles of key actors such as the sugar industry, government, and consumers in legitimizing and distributing the market value of sugar. Collectively, this framework enables a comprehensive analysis of the nonnegotiable symbolic value of sugar in Indonesia.

Sugar Market Valuation: The Nonnegotiable Symbolic Value

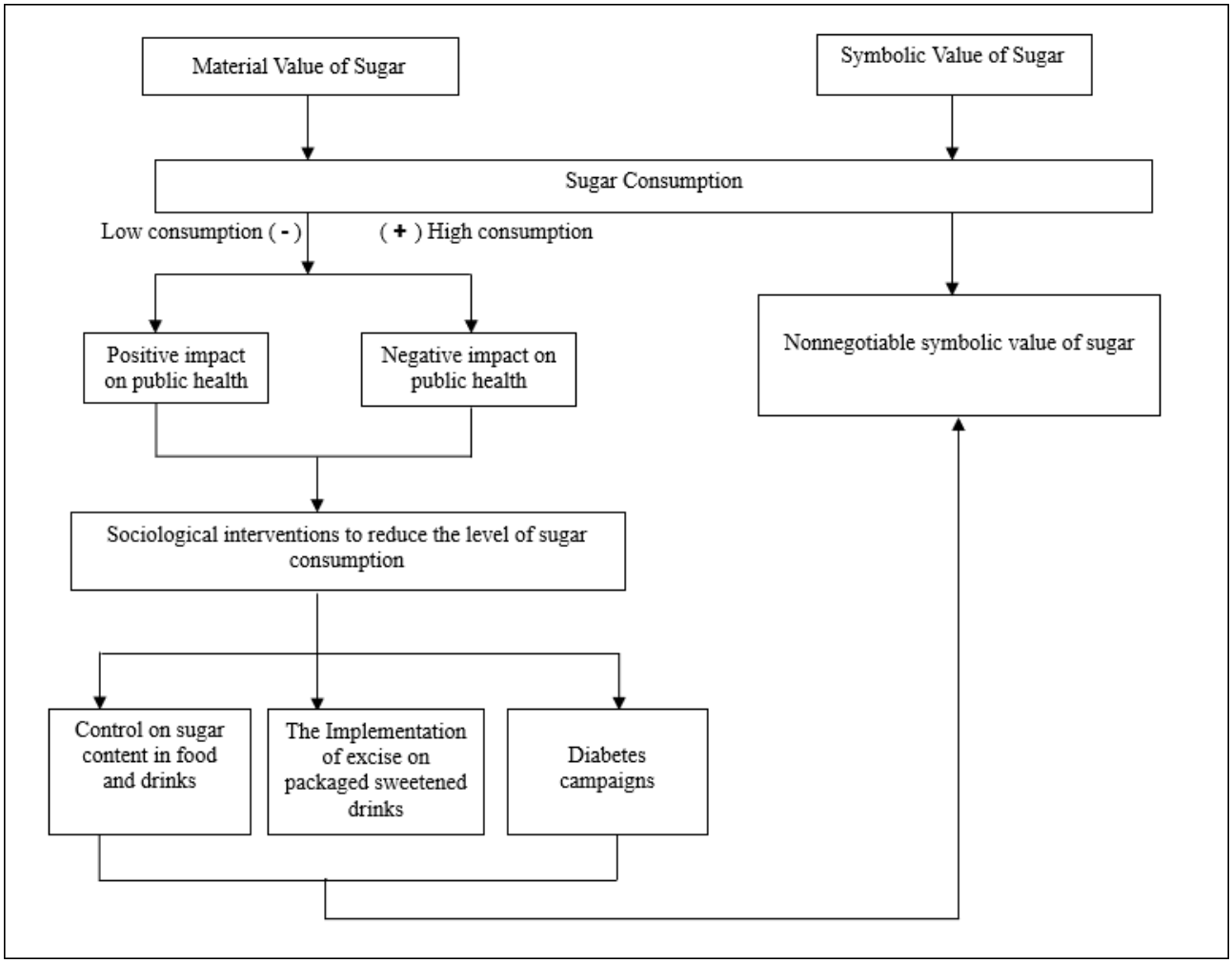
This study examines the nonnegotiable symbolic value in sugar consumption. Sugar plays a pivotal role in cultural communication as well as in economic and political contexts. Eating patterns are inseparable from cultural context, while food culture shapes individual identity [3].

Previous studies highlight the important symbolic value of food. Similarly, research linking socioeconomic status and consumption patterns discloses that food functions not only as nourishment but also as a bearer of social, cultural, and emotional meanings [18].

The nonnegotiable symbolic value of sugar appears in 2 stages. First, in material terms, low sugar consumption correlates with positive impacts on public health. On the other hand, high sugar consumption, as indicated by high sugar content (Brix value), leads to negative impacts on public health. Overcoming these impacts requires regulating sugar content in foods and beverages, imposing excise taxes on packaged foods and beverages, and promoting diabetes awareness.

Second, the symbolic value of sugar outweighs its material value, leading to its distinctive status as an immutable, culturally or religiously significant food item. This nonnegotiable symbolic value drives high sugar consumption and contributes to negative impacts on public health. Figure 1 summarizes this phenomenon.

Figure 1. Nonnegotiable symbolic value framework.



The Sugar Market: Cognitive Anchoring, Social Positioning, and Standardization

This study analyzes the sugar market through 3 aspects of market valuation by Beckert [5]: cognitive anchoring, social positioning, and standardization. Cognitive anchoring pertains to individual and collective comprehension of markets and products. In the sugar market, it includes knowledge of the market dynamics, pricing mechanisms, and supply-demand factors that shape decisions on purchasing, selling, and investment.

XSL•FO
RenderX

Social positioning concerns the role of social recognition in assigning value to goods. Value is linked to the status the goods confer [19]. For example, wine signifies symbolic prestige beyond material quality. In the sugar market, social positioning involves key stakeholders such as major sugar companies, government entities, and farmer groups in the social and economic hierarchy, which dictates access to resources, information, and influence. Notably, large sugar companies may wield greater influence over pricing compared to small farmers.

Standardization establishes shared criteria for assessing product quality [5]. It is defined as the basis of the physical or material quality of goods, supported by market expansion and digital technology [20] . In the sugar industry, standardization governs how sugar quality and composition are evaluated, shaping consumer perception and global trade valuation.

Theoretical Framework

Within economic sociology, commodity value derives from both material attributes and embedded symbolic meanings. Beckert [5] argues that markets operate within a social order in which product valuation encompasses material value, referring to the use and tangible benefits of products, and symbolic value, which includes social meanings, status, and identity associated with consumption. Thus, valuation is not merely functional but socially constructed through cultural relations and symbolic meanings.

Symbolic value emerges from interactions among consumers, producers, and prevailing cultural and moral structures. Consumption in this regard is viewed as a symbolic communication system, wherein goods are used to express identity and social position [21]. Meanwhile, the concept of cultural capital [22] suggests that consumption preferences indicate the habitus and class structures that influence consumer behavior. These perspectives demonstrate that symbolic value significantly shapes consumption behavior, often beyond economic rationality.

This study applied the framework of economic sociology by Beckert [5,13] on symbolic value and market valuation, positing that product value emerges from social construction rather than material properties alone. In the context of sugar consumption in Indonesia, this framework elucidates how the emotionally, culturally, and socially embedded nonnegotiable symbolic value of sugar shapes consumption practices. Prior supporting studies on the traditional herbal medicine (jamu) market [20], digital group solidarity [23], and moral embeddedness in the labor market [24] conceptually reinforce the interrelations among social norms, digital interactions, and consumption behavior.

To analyze these dynamics, this study adopted 3 key dimensions of Beckert [5] (standardization, cognitive anchoring, and social positioning) to explain how collective perceptions, social norms, and individual positions within social structures reproduce the symbolic value of sugar. Integrating digital network analyses, specifically TNA and SNA, this study explores how the symbolic value of sugar is constructed, maintained, and reinforced within online public spaces and contemporary digital culture.

Methods

Research Methodology

This study used a mixed methods approach with a qualitative digital orientation, integrating TNA and SNA to examine nonnegotiable symbolic value and uncontrollable sugar consumption in Indonesia. This approach explored how symbolic value, public communication, and consumption practices are shaped through digital interactions. Purposive sampling was guided by 3 theoretical dimensions of Beckert [5] (standardization, cognitive anchoring, and social positioning) to capture relevant social contexts and public narratives. The analysis drew from both primary and secondary data. Primary data included web-scraped digital texts and regulatory documents, while secondary data encompassed market observations and interviews with 3 key informants. Data processing and interpretation combined the quantitative rigor of network analysis with qualitative interpretive depth, enabling a comprehensive understanding of the complexity of symbolic value and sugar consumption practices in Indonesia. Validation was achieved through source and method triangulation, member checking, and peer debriefing, ensuring consistency and reliability.

Data Collection

Data collection began with identifying keywords related to the material and symbolic values of sugar (Table 1), followed by web scraping using NCapture (Lumivero). Primary data were collected from online news articles, YouTube videos uploaded between January 2018 and July 2022 (see Multimedia Appendix 1), and government policy documents. They were cleaned, coded, and categorized in NVivo (Lumivero) based on material value, symbolic value, and social positioning. They were then processed through stop-word filtering in WORDij (Amsterdam School of Communication Research) and imported into Gephi (Gephi Consortium) to generate TNA visualizations of the material and symbolic values of sugar. Concurrently, the SNA visualization mapped dominant actor networks regulating sugar levels in packaged foods and beverages.

XSL•FO
RenderX

Table . Keywords for web scraping from online media articles.

Criteria	Keywords
Standardization	Standardized sugar in food and beverages, sugar consumption policy standardization, SNI[a] for sugar, and sweetness level.
Cognitive anchoring	Sugar for diet, low-calorie sugar, experiences with sugar and artificial sweeteners, experiences with sugar and without sugar, reasons for using sugar in foods and beverages, sugar content, reasons for having a fondness for sweet foods, and the symbolic meaning of sugar.
Social positioning	The history and culture of sugar consumption, the royal symbolism of sugar, and the political economy of sugar.

[a]SNI: Indonesian National Standard.

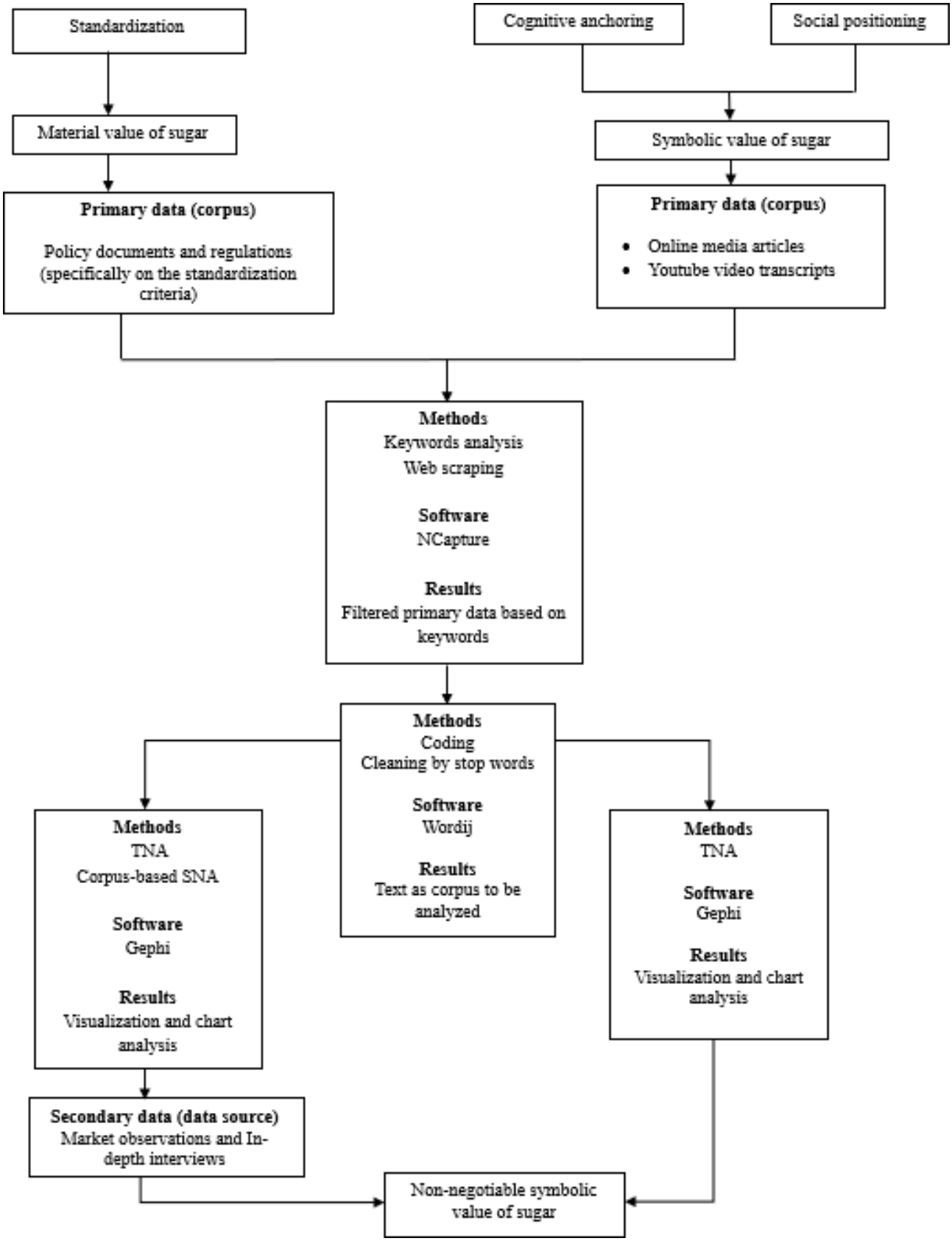
The material value of sugar, analyzed through standardization, was drawn from 13 online articles and 3 regulations: Permenkes No. 30/2013 on daily sugar intake limits and nutritional information labeling for packaged products, Permendag No. 14/2020 on sugar import provisions, and the Regulation of the National Agency of Drug and Food Control (BPOM) No. 4/2014 on artificial sweetener limits in foods and beverages. The symbolic value of sugar was explored through 106 online articles, 76 on cognitive anchoring and 30 on social positioning, and 7 YouTube seminar transcripts.

Secondary data were obtained from market observations examining labeling compliance in packaged foods and beverages. In addition, online interviews were held between January 2023 and March 2023 with 3 purposively selected key informants: a senior researcher from the National Research and Innovation Agency, a research project officer from the Center for Indonesia's Strategic Development Initiatives, and a clinical physician who is also a person with diabetes. Discussions addressed the effectiveness of the sugar-sweetened beverage (SSB) excise policy, the implementation of Permenkes No. 30 of 2013 on sugar, salt, and fat (SSF) labeling, consumer literacy, and responses of the food and beverage industry to existing regulations. Informants also highlighted the contribution of SSB to the rising NCD prevalence in Indonesia, assessed current regulatory effectiveness, described market conditions and consumer preferences for sweetness, and suggested effective intervention strategies such as public education, product reformulation, and cross-sectoral policy approaches.

These methods enabled the study to map how the nonnegotiable symbolic value of sugar is reproduced in both digital discourse and everyday practices. These findings elucidate its role in sustaining sugar-based dietary habits in Indonesia and inform more comprehensive sociological interventions to control sugar consumption by considering cultural and symbolic dimensions. The methodology framework is presented in Figure 2.

XSL•FO
RenderX

Figure 2. Methodology framework. SNA: social network analysis; TNA: textual network analysis.



Ethical Considerations

The study was reviewed and approved by the research ethics committee of the Faculty of Social and Political Sciences, Universitas Indonesia (reference: KET-12/UN2.F9.KEP/PPM.00.02/2025; certificate: SER-12/UN2.F9.KEP/PPM.00.02/2025). The research used secondary and observational data and did not involve direct interaction or intervention with human participants. Therefore, the requirement for informed consent was waived by the ethics

XSL•FO
RenderX

committee. Participants did not receive any financial or material compensation. All data were analyzed in an aggregated and anonymized form to ensure confidentiality and prevent individual identification.

Results

TNA and SNA Visualizations of the Material Value of Sugar and its Negative Impacts on Public Health

TNA of Material Valuation Based on Standardization

The TNA visualization of the sugar and artificial sweetener market, focusing on standardization, reveals 27 nodes and 58 edges with 6 clusters (Figure 3). These clusters include: purple (38/626, 6.07%), green (33/626, 5.27%), yellow (26/626, 4.15%), orange (21/626, 3.35%), blue (20/626, 3.19%), and pink (17/626, 2.72%).

Figure 3. Textual network analysis (TNA) visualization of sugar based on standardization.

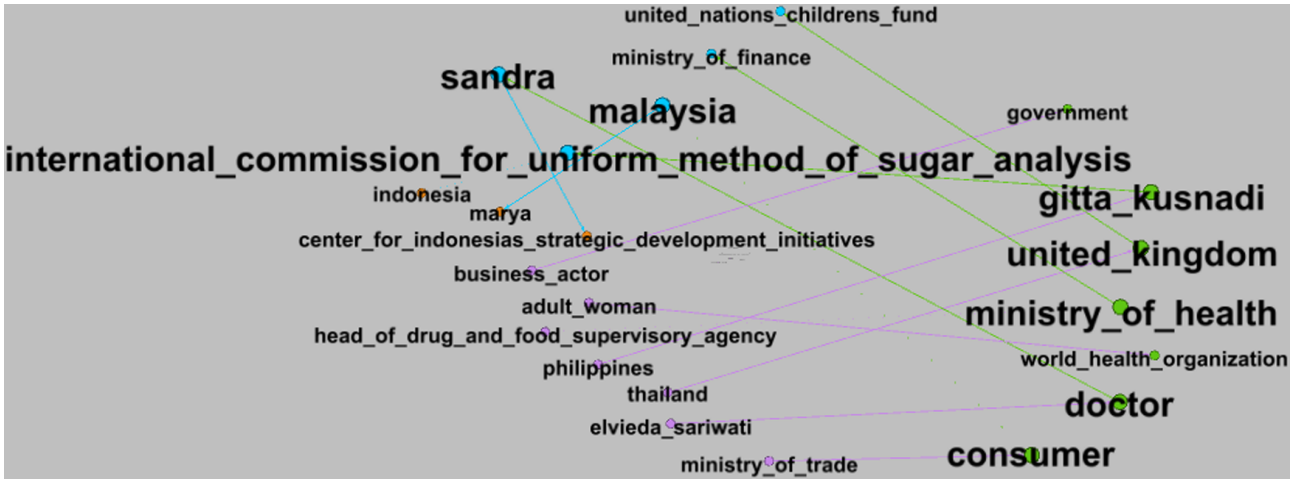


Figure 3 illustrates the structure of public discourse on sugar, food, beverages, and health through several color-coded thematic clusters. The purple cluster centers on beverages and their association with "excise," "government," "industry," "impact," and "packaging," framing sweetened beverages as objects of fiscal policy and industrial regulation. The green cluster groups "sugar," "white," "crystal," "sweet," and "yellow," reflecting the physical attributes and sensory perceptions of sugar as a consumer product.

Figure 4 presents an SNA visualization of sugar-related discourse based on standardization, highlighting key regulatory and industry actors. The yellow cluster comprises "health," "regulation," and "message," highlighting the role of public health as an intermediary between food and policy as well as a channel for communicating regulatory messages. The orange cluster focuses on "food," "ingredient," "additional," "maximum," and "usage," depicting technical discourses on food composition and limits on added sugar. The blue cluster connects "diabetes," "obesity," "disease," and "child," representing narratives of health risks and NDCs associated with sugar consumption, including among children. Meanwhile, the pink cluster links "gram," "spoon," and "eat," indicating measurement and everyday practices of consuming sugar. Overall, this mapping demonstrates that discourse on sugar is segmented into policy, public health, technical food considerations, disease risk, product perception, and consumption practices, without implying a direct causal relationship with individual behavior.

XSL•FO
RenderX

Figure 4. The social network analysis (SNA) visualization of sugar based on standardization.



SNA of Influential Actors in the Policy on the Sweetness Levels in Packaged Foods and Beverages

The corpus-based SNA visualization focused on standardization uncovers dominant actors in the policy on the sweetness levels in packaged foods and beverages. This visualization comprises 29 nodes and 15 edges, with 4 clusters of actors in reference to eigenvector centrality.

Eigenvector centrality analysis further quantified the relative influence of these clusters (Table 2). The most dominant clusters (4/29, 13.79%) consist primarily of governmental and international health actors such as the Ministry of Health, BPOM, WHO, and ICUMSA, indicating their pivotal role in shaping narratives on sugar regulation and public health. The purple cluster holds the most substantial influence in formulating the sugar content policies for packaged foods and beverages, demonstrating strong interconnections among key stakeholders, including the Ministry of Health, the Ministry of Trade, and BPOM. In contrast, the public and business sectors are expected to exert greater influence following the implementation of these policies.

Table . Value of eigenvector centrality.

Cluster color	Eigenvector centrality	Value, n/N (%)
Purple	Elvieda Sariwati, MD, adult women, Ministry of Trade, Head of BPOM[a], Philippines, Thailand, and business actors	4/29 (13.79)
Light green	Ministry of Health, consumers, doctors, Gitta Kusnadi, United Kingdom, WHO[b], the government	4/29 (13.79)
Blue	ICUMSA[c], UNICEF[d], Malaysia, Ministry of Finance, Sandra	4/29 (13.79)
Orange	Marya, CISDI[e], Indonesia	3/29 (10.34)

[a]BPOM: National Agency of Drug and Food Control.

[b]WHO: World Health Organization.

[c]ICUMSA: International Commission for Uniform Methods of Sugar Analysis.

[d]UNICEF: United Nations Children's Fund.

[e]CISDI: Center for Indonesia's Strategic Development Initiatives.

TNA Visualizations of the Nonnegotiable Symbolic Value of Sugar in Processed Foods and Beverages Containing Refined Sugar

TNA of Symbolic Valuation Based on Cognitive Anchoring

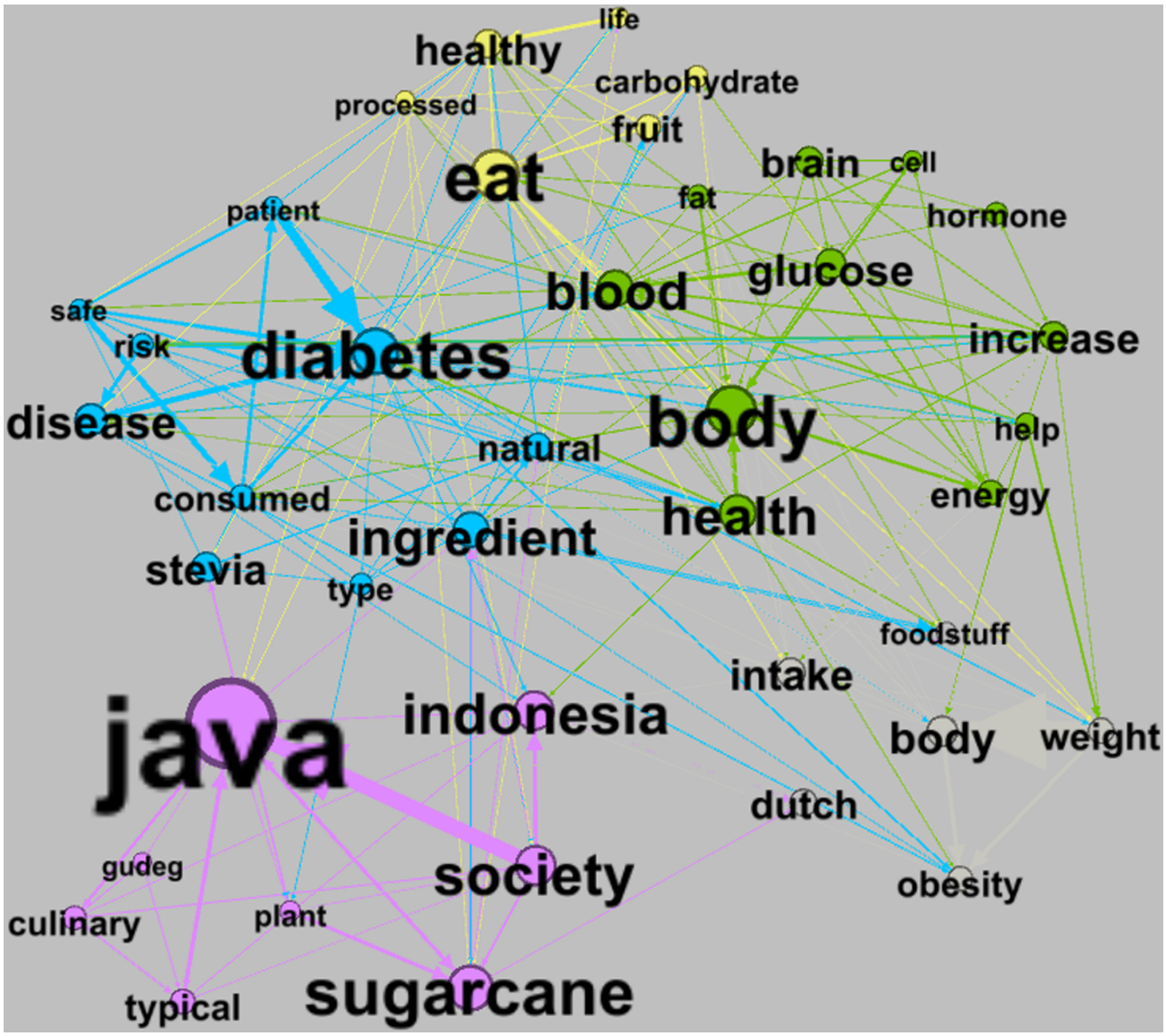
The TNA visualization of the sugar market based on cognitive anchoring reveals 56 nodes and 430 edges with 5 clusters. These clusters include: purple (171/2071, 8.26%), green (148/2071, 7.15%), blue (121/2071, 5.84%), yellow (95/2071, 4.59%), and gray (49/2071, 2.37%).

Figure 5 presents a TNA visualization mapping semantic linkages within discourses on sugar consumption in Indonesia. The word network reveals overlapping thematic clusters, illustrating sugar as not only a nutrient but also a health, cultural, and social issue. The purple cluster ("Java," "sugarcane," "society," and "Indonesia") represents the sociocultural dimensions of sugar consumption, linking geographical context, sugarcane production, culinary traditions, and the role of sugar in social life and local identity. The green cluster ("body," "glucose," "blood," "health," "energi," and "hormone") reflects the understanding of sugar as part of biological and metabolic

XSL•FO
RenderX

processes, with implications for health and body weight. The blue cluster ("diabetes," "disease," "risk," and "patient") captures medical and public health discourses emphasizing the risks of sugar consumption, particularly diabetes, and preventive approaches grounded in health and safety.

Figure 5. Textual network analysis (TNA) visualization of sugar based on cognitive anchoring.



The yellow cluster ("eat," "healthy," "carbohydrate," "fruit," and "fat") relates to everyday eating practices, food choices, and healthy lifestyle narratives that connect nutritional knowledge with consumption behavior. The gray cluster ("intake," "foodstuff," "weight," "obesity," and "body") reflects health-oriented rationalities and dietary intake control, focusing on measurable impacts on nutritional status and body weight. Overall, the interconnected clusters suggest that sugar consumption is constructed through interactions between material values (eg, "health," "the body," and "disease") and symbolic values rooted in the social and cultural context of Indonesia.

TNA of Symbolic Valuation Based on Social Positioning

The TNA visualization of the sugar and artificial sweetener market based on social positioning displays 33 nodes and 136 edges with 5 clusters. These clusters include: purple (83/752, 11.04%), green (48/752, 6.38%), blue (30/752, 3.99%), orange (26/752, 3.46%), and pink (23/752, 3.06%). The visualization is presented in Figure 6.

XSL•FO
RenderX

Figure 6. Textual network analysis (TNA) visualization of sugar based on social positioning.

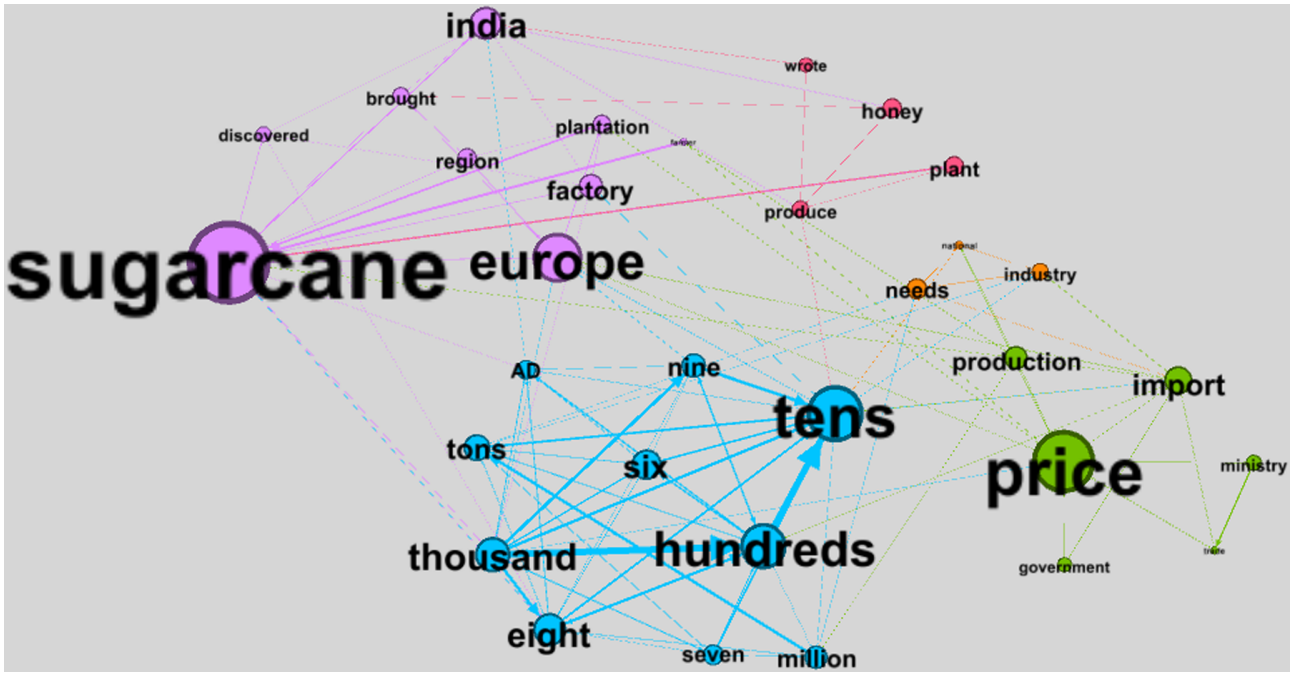


Figure 6 presents a TNA visualization of the discourse structure surrounding sugarcane and the sugar market through interconnected thematic clusters. The purple cluster centers on "sugarcane," "Europe," "India," "plantation," and "factory," representing the historical, geographical, and production bases of sugarcane, including colonial diffusion pathways and early production systems based on plantations and factories. The green cluster ("price," "import," "government," and "ministry") reflects sugar market mechanisms and regulatory frameworks, emphasizing pricing, import policies, and the role of the state and ministries in sugar market regulation and stabilization. The blue cluster is dominated by numerical terms ("tens," "hundreds," "thousand," and "million"), demonstrating quantification, scale, and volume, indicating how sugar-related discourse is frequently framed through measurements of production, distribution, and consumption.

The orange cluster, which includes "industry," "needs," and "national," frames sugar as a strategic commodity for meeting national needs, highlighting the role of industry in ensuring supply and production sustainability. This cluster situates sugar within a macroeconomic and food security framework, in which national consumption requirements underpin policy decisions and industrial activities. The pink cluster ("plant," "produce," "honey," and "wrote") represents discourses on sweeteners in natural and preindustrial contexts, reflecting early practices and understanding of sweetener sources from plants and honey and their documentation in historical narratives. Overall, the interconnections among clusters indicate that discourse on sugarcane and sugar is constructed through interactions among historical–productive dimensions, economic quantification, market mechanisms, and industrial and policy logics.

Discussion

Alternative Sociological Interventions to Reduce Sugar Consumption While Addressing its Nonnegotiable Symbolic Value to Improve Public Health

The Implementation of Permenkes No. 30 of 2013

Permenkes No. 30 of 2013, mandating the labeling of SSF content and health warnings on processed and ready-to-eat foods, represents a key strategy for NCD risk control. However, TNA focused on standardization (Figure 3) demonstrates that producers largely ignore the recommended daily sugar intake (maximum 50 g or 4 teaspoons). This implies weak regulatory enforcement and low producer awareness of the right of consumers to transparent information.

Market observations indicate widespread noncompliance among Indonesian food and beverage producers, reflecting weak accountability in protecting consumer health. The absence of restrictions on added or artificial sweeteners further worsens this issue. As presented in Figure 7, many packaged beverages lack clear sugar content labeling, highlighting the urgent need for stricter sanctions.

XSL•FO
RenderX

Figure 7. Packaged beverage produced in Indonesia with noncompliant labeling.



Several domestic producers partially comply with Permenkes No. 30/2013 by providing basic health information. Figure 8 presents an example of a locally manufactured beverage that includes sugar content, health information, and producer details on its label. However, such instances remain limited, and the quality and completeness of labeling information vary widely across brands.

XSL•FO
RenderX

Figure 8. A packaged beverage produced in Indonesia with incomplete labeling.



In contrast, Japanese food and beverage products consistently feature comprehensive labeling, including sugar content, manufacturer details, website, and 2D barcode (Figure 9), enabling consumers to make informed choices. This comparison reveals that Indonesia lags behind Japan in labeling transparency and consumer protection. Strengthening Permenkes No. 30/2013 thus requires not only administrative enforcement but also sociocultural alignment with Indonesian consumption behavior.

XSL•FO
RenderX

Figure 9. Packaged foods and beverages produced in Japan, including (A) vinegar-based beverage, (B) green tea bags, (C) chocolate biscuit, and (D) biscuit snack.



The TNA and SNA findings indicate that sugar consumption in Indonesia is shaped by the interrelation between material factors and symbolic meanings embedded in social and cultural contexts. Sugar is perceived not only as a food ingredient, but also as a symbol of pleasure, togetherness, and well-being that is continuously reproduced through public discourse. Within the market valuation framework of Beckert [5], these findings reflect the interplay between material and symbolic values in market practices, where government standardization grounded in health rationality often confronts cognitive associations of "sweetness" with affection, happiness, and social comfort.

Rather than establishing direct causal links to consumption behavior, TNA elucidates the cognitive and cultural context that sustains sugar consumption. The SNA findings complement this by highlighting the role of key actors within regulatory and market networks in maintaining and reinforcing these symbolic meanings. Consequently, policy approaches that rely solely on text-based nutritional labeling tend to be less effective, as they conflict with the deeply institutionalized symbolic understanding of sugar. Front-of-pack color-coded labeling may therefore serve as a symbolic intervention with the potential to challenge these meanings, while bridging public health policy with the social and cultural structures underlying excessive sugar consumption.

Revision of Permenkes No. 30/2013

The mandate under Permenkes No. 30 of 2013 to disclose SSF content and health warnings on processed and ready-to-eat foods has not effectively changed public consumption patterns. Interviews with 3 key informants from the National Research and Innovation Agency of Indonesia, Center for Indonesia's Strategic Development Initiatives, and the medical sector identified 3 main barriers: low nutrition literacy, noncommunicative label design, and weak regulatory enforcement.

An informant from the Public Health and Nutrition Research Center of National Research and Innovation Agency of Indonesia identified low nutrition literacy as the primary barrier. Although nutritional labeling meets global standards, public awareness remains limited. "Many people do not understand or pay attention to details. Even expiration dates are often ignored." The informant suggested a simplified color-coded system (green, yellow, and red) to communicate product risks more clearly.

Echoing this, an informant from Center for Indonesia's Strategic Development Initiatives noted that narrative text-based labels are ineffective in a visually driven digital era. Consumers respond more to visual promotions than textual warnings. "There are no regulations on playful characters for children, leaving gaps for improvisation. Health warnings remain non–eye-catching." The informant further emphasized the need for digitalized monitoring and cross-sector collaboration among the Ministry of Health, BPOM, and the Ministry of Industry to improve compliance.

XSL•FO
RenderX

A medical practitioner and diabetes survivor added that most consumers do not understand or ignore nutritional information. "If people never read the information, the regulation is useless." This underscores the gap between material value (regulations and factual data) and symbolic value (social meaning attached to sugar), as described by Beckert [5]. In this context, nutritional information lacks sufficient social impact to alter the symbolic perception of sugar as a source of pleasure and energy.

The implementation of article 4, paragraph (2), which warns that "consuming more than 50 g of sugar per day increases the risk of hypertension, stroke, diabetes, and heart attack," also remains ineffective. Many "healthy" products, such as yogurt, instant oatmeal, and low-fat salad dressing, contain 12 to 26 g of sugar per serving, reinforcing sweetness as a symbol of pleasure and premium value rather than a health risk.

These findings highlight the need to revise Permenkes No. 30 of 2013 to improve both the content and delivery of health warnings. A color-coded system, green for below the daily limit, yellow for 50%, and red for more than 60% to 75%, can enable faster risk interpretation and encourage producer compliance. Beyond administrative compliance, labeling should function as a social intervention to shift the symbolic meaning of sugar from pleasure toward health awareness.

This approach should be integrated with public education through social media, schools, and primary healthcare facilities. Incorporating standardization, cognitive anchoring, and social positioning dimensions from the theory of Beckert [5], the policy can serve as a multidimensional intervention uniting regulatory measures, social literacy, and symbolic transformation to promote healthier consumption behavior.

An Ideal Example of Sweetness Level Labels and Health Warnings on Processed Foods

Referring to the guidelines by BPOM in 2020, processed food labels must include: (1) product and trade names, (2) ingredient list, (3) net weight or volume, (4) producer or importer details, (5) halal certification as mandated, (6) production date and code, (7) expiry date, (8) distribution permit number or marketing authorization number, (9) ingredient origin, and (10) color code (green, yellow, and red), as proposed in this study.

Labels should also encompass nutritional values, a 2D barcode, and additional legally mandated information. Products containing artificial sweeteners must include health warnings such as "Intended for diabetics or those seeking low-calorie options. Not suitable for children under five or pregnant/lactating women. Overconsumption may cause a laxative effect." [25].

Since the mid-2000s, front-of-pack nutritional labeling has been applied in the United Kingdom. In October 2012, the UK government recommended a traffic-light system for front-of-pack, which is now widely adopted by producers. This system communicates total fat, saturated fat, sugar, and salt content: green for low, yellow for moderate, and red for high levels [15].

Controlling Sugar Consumption Through the Visualization of Cognitive Anchoring

TNA focusing on cognitive anchoring (Figure 5) reveals strong interactions between the symbolic and material values of sugar across narrative clusters. The purple cluster highlights cultural reasons behind Javanese preferences for sweet foods, framing sugar as a symbol of intimacy, happiness, and prosperity. Within the framework proposed by Beckert [5], this indicates how symbolic value is deeply embedded in cognitive anchoring, making sugar consumption a socially inherited habitus.

Conversely, the gray cluster emphasizes narratives on health risks, such as obesity and diabetes, exposing tension between the material value of sugar as energy and awareness of its health impacts. The blue cluster presents solutions, advocating natural sweeteners such as stevia to preserve the symbolic meaning of sweetness while promoting health.

Indonesian consumers exhibit a strong preference for high-sugar beverages such as packaged tea, with a 250 ml serving containing approximately 21 g of sugar (pink cluster). This suggests that processed sugar carries greater symbolic value than raw sugar. Sweet beverages are frequently associated with refreshment, social bonding, and modern lifestyle, forms of symbolic positioning consistent with Beckert [5]. The longstanding fondness for sweet foods in Java and Yogyakarta, established in childhood, reflects a culturally stable sugar-based habitus.

These findings align with international studies demonstrating that sugar consumption is influenced not solely by biological needs but also by social, cultural, and emotional factors. Sugar has been linked to social rituals such as communal meals and celebrations [26], and it carries a dual meaning, representing both pleasure and a health risk in France and Denmark [27]. In Indonesia, sugar also signifies luxury, happiness, and social status. Globalization has been associated with increasing sweet beverage consumption among adolescents, while family influence and digital education shape sugar consumption among children [28].

Research also underscores the symbolic dimension of sugar in public policy. Marketing strategies leveraging social symbolism significantly affect consumer choices toward low-sugar products, indicating that consumption decisions are often driven by social meaning rather than rational evaluation [29]. Health perceptions, social norms, and public policies such as nutrition labeling play critical roles in reducing sugar intake. Historically, sugar served as a marker of status, power, and social identity from colonial to industrial times [30].

Therefore, controlling sugar consumption in Indonesia requires more than regulatory measures or health campaigns. Effective interventions should integrate symbolic and cultural understanding with the dimensions of standardization (regulation), cognitive anchoring (cultural internalization), and social positioning (actor and market structures) by Beckert [5] to create sustainable social change.

XSL•FO
RenderX

Controlling Sugar Consumption Through the Visualization of Social Positioning

The visualization of social positioning in controlling sugar consumption reveals the dominance of the purple cluster, tracing the historical and social evolution of sugar from its early production in Indian sugarcane plantations to its status as a luxury commodity among European elites. Historically, the material and symbolic values of sugar were confined to privileged classes, positioning it as a symbol of luxury, status, and power [5]. This exclusivity established sugar not merely as an economic commodity but as a marker of class distinction.

In modern society, this symbolic legacy endures in a transformed manner. The rising demand for sugar in the food and beverage industry reflects its evolution from a luxury item to a mass symbol of pleasure and modern lifestyle. Despite broader economic accessibility, sugar continues to represent comfort, celebration, and social bonding that continue to drive overconsumption, particularly in urban settings. Consistent with the framework of Beckert [5], social positioning legitimizes the allure of sugar beyond its utilitarian function.

International trade dynamics further highlight the dual nature of sugar as a material necessity and a symbolic commodity. Global expansion and economic centrality demonstrate the persistence of structural power relations, where corporate actors and trade systems not only regulate sugar availability but also shape its social and cultural meanings. Consequently, controlling sugar consumption requires more than economic and regulatory measures. It demands sociological interventions to confront the deeply embedded symbolic value of sugar in everyday life.

Sociological Intervention in Sugar Market Valuation Through the Implementation of the Excise Policy on Packaged Sweetened Foods and Beverages

The rising consumption of SSBs is influenced by multiple factors, including their relatively low prices and widespread availability, highlighting the need for sociological interventions to reduce intake, particularly among children and youth [2]. Excise taxes have proven effective in reducing consumption by increasing retail prices. Since 2015, the United States has enforced excise taxes on sugary beverages, causing a 21% decline in consumption. The “Soft Beverages Levy” introduced in the United Kingdom in 2018 reduced sugar intake by 10%, while the excise tax policy applied in Mexico in 2014 achieved a 6% - 8% decrease. Similarly, South Africa adopted the “Health Promotion Levy” in 2018 and observed a considerable 29% reduction in sugar consumption [31].

Sociological Intervention in Sugar Market Valuation Through Diabetes Awareness Campaigns

The Health Minister of Indonesia emphasizes proactive strategies to prevent and control diabetes. These include regular risk monitoring and lifestyle modification, particularly in high-risk communities. The government advances this agenda through community empowerment programs under Integrated Development Posts for NCDs.

The Ministry of Health also promotes CERDIK, an acronym for check health regularly, avoid smoking, exercise regularly, eat a healthy diet, rest adequately, and keep balance between body and mind [32]. Effective diabetes awareness campaigns further necessitate community involvement, using testimonials and success stories to motivate healthier behaviors.

Dealing With the Nonnegotiable Symbolic Value

This economic sociology study using the framework of Beckert [5] offers key insights into sociological interventions. First, it reveals that consumer behavior is shaped by both the symbolic and material values of a commodity. The nonnegotiable symbolic value in sugar consumption is identified, in which meanings attached to sugar extend beyond rational economic considerations. This enriches the concept of value construction of Beckert [5] that symbolic meanings can persist even when confronted with health risks and economic rationality.

Second, the nonnegotiable symbolic value of sugar manifests through its association with emotional satisfaction, happiness, and social bonding. Sweet foods hold cultural significance in traditions and celebrations, symbolizing generosity, joy, and prosperity, elevating sugar from a mere commodity to a social symbol. Consequently, reducing sugar consumption requires sociological strategies that address its symbolic dimension alongside economic and health-based approaches. Integrating 3 analytical dimensions (ie, standardization, cognitive anchoring, and social positioning) of Beckert [5] clarifies the persistence of sugar consumption in Indonesia.

First, standardization is reflected in state regulations, such as Permenkes No. 30 of 2013, which sets daily sugar intake limits but suffers from weak enforcement and low public awareness. Second, cognitive anchoring is evident in cultural preferences for sweet foods and beverages, reinforced by digital advertising and everyday practices that normalize sugar as a symbol of comfort, happiness, and social acceptance. Third, social positioning involves dominant actors, including major sugar corporations, government agencies, and consumer groups, who shape policy direction, market access, and the symbolic legitimacy of sugar products.

Linking these 3 dimensions to empirical findings demonstrates that high sugar consumption in Indonesia is not merely economic but deeply rooted in social and cultural meanings. This study extends the theoretical framework of Beckert [5] by incorporating the cultural context of Indonesia, digital marketing narratives, and the paradox between high health-risk awareness and persistent overconsumption. It thus reinforces the theoretical relevance of the market valuation model proposed by Beckert in explaining the persistence and complexity of contemporary sugar consumption patterns.

Critical Discussion: Extending the Symbolic Value Theory in the Digital Era

The symbolic value theory by Beckert [5,15] interprets commodity consumption through social status, identity, and imagined futures. However, its application to sugar consumption in Indonesia remains limited. This study extends the framework by introducing 3 dimensions capturing contemporary social, cultural, and economic shifts.

XSL•FO
RenderX

First, the local cultural dimension. In Indonesia, sugar symbolizes togetherness and tradition beyond social status. Sweet tea and snacks are deemed integral to social interactions in West Java [33]. Unlike the distinction theory [22], sugar consumption is more inclusive and collective, deeply rooted in everyday communal practices.

Second, the digital narrative dimension. Digital promotion normalizes sugary beverage consumption, particularly among youth. Indonesian food and beverage companies are reported to use social media hashtags, promotional characters, and interactive games to target children and adolescents [34]. Similarly, brands such as Coca-Cola frame sugary beverages with happiness, togetherness, and active lifestyles [3]. Thus, the symbolic value of sugar now arises from interactive, cross-cultural, and emotionally charged digital narratives, not merely market mechanisms.

Third, the public health paradox dimension. Despite awareness of diabetes risks, sugar consumption remains high due to cravings, social norms, and digital marketing [33,34]. As observed in France and Denmark, sweetness functions as a socially accepted "guilty pleasure" [27].

Recent reports [17,35] confirm these trends among Generation Z, linking growing diabetes risks to heavy consumption of "Instagrammable" sugary beverages promoted as modern lifestyle symbols. Consequently, the theory of Beckert [5] requires expansion to encompass: (1) digitalization, where social media algorithms and visual content shape perceptions of value and pleasure; (2) globalization, where consumption narratives adapt to global trends; and (3) public health paradox, where heightened health awareness enhances rather than reduces the symbolic appeal of indulgence.

By integrating these dimensions, this study introduces the conceptual model of digital-symbolic value of consumption, where symbolic meanings emerge from interactions among digital capitalism, global aspirations, and youth self-expression. In this concept, sugar consumption in Indonesia transcends social status or collective culture, becoming part of digital identity and global lifestyle. Accordingly, this study advances economic sociology by highlighting how digitalization, globalization, and health ambivalence redefine the symbolic meanings of commodities in the modern era.

Limitations

The main limitation of qualitative digital research lies in sample determination. Data from online news media and YouTube are difficult to apply due to their uncertain nature. Although digital research may lack repeatability, it remains accountable and fulfills the criterion of recoverability with regard to data collected within a specific time period.

Conclusions

Demand-Side Perspective on Sugar Consumption and Regulatory Compliance

This study contributes to the demand side of economic sociology concerning the sugar market. It explores strategies to address or mitigate the sugar-driven food habits and habitus from the perspective of consumer behavior. Simultaneously, it examines producer compliance with regulations aimed at controlling sweetness levels in foods and beverages, thus supporting efforts to reduce sugar consumption and the prevalence of NDCs.

An Illustration of the Material Value of Sugar Which Has a Negative Impact on Public Health as Shown by the TNA Visualization

The TNA visualization of material value based on standardization (Figure 3) highlights that the widespread availability of affordable sweetened foods and beverages has increased the prevalence of NCDs, particularly diabetes. Measures to control sweetness levels are thus crucial to protect public health. Implementing excise taxes on SSB will incentivize producers to reformulate products with greater consideration for consumer health.

An Illustration of the Nonnegotiable Symbolic Value of Sugar in Processed Foods and Beverages Containing Refined Sugar as Shown by the TNA Visualization

The TNA visualization of cognitive anchoring (Figure 5) reveals that while raw sugar has limited symbolic value, its use in processed foods and beverages elevates its sociocultural significance, shaping consumer behavior. Meanwhile, the TNA visualization of social positioning discloses that sugar was historically reserved for the upper class (Figure 6), but its accessibility today demonstrates a notable evolution of its symbolic value into a staple commodity accessible to all social groups.

Alternative Sociological Interventions to Limit Sugar Consumption and Address Its Nonnegotiable Symbolic Value

Controlling sugar consumption habits and habitus and its embedded symbolic value requires a multifaceted intervention. First, the market valuation of sugar and sweet products should be reshaped through targeted consumer education and campaigns. Second, regulatory measures, including the adjustment of sweetness levels, refinement of Permenkes No. 30/2013, and consistent law enforcement, are essential to ensure compliance. Third, implementing excise taxes on sweetened foods and beverages can discourage overconsumption.

Strict enforcement of labeling regulations mandating the disclosure of SSF content remains crucial, as many packaged products still fail to comply. Addressing policy gaps is necessary to effectively curb sugar consumption and challenge its symbolic value. Moreover, empowering women, particularly housewives, to shape consumer cognition and promote healthy lifestyles highlights the importance of active involvement from the Ministry of Women Empowerment and Child Protection [35].

Sociological Intervention in Sugar Market Valuation Through the Imposition of Excise on Packaged Sweetened Beverages

By implementing excise on packaged sweetened beverages, the government can better understand the social ramifications of the policy and the necessary measures to ensure that it benefits public health. Prior to the enforcement, it is imperative to conduct a comprehensive analysis of potential social impacts

XSL•FO
RenderX

and establish mechanisms for ongoing monitoring and evaluation. This proactive approach enables the government to promptly identify any emerging issues and make requisite adjustments.

Sociological Intervention in Sugar Market Valuation Through Diabetes Awareness Campaigns

Diabetes awareness campaigns aim to change public perceptions of sugar, curb excessive consumption, and foster healthier lifestyles. This involves educating the public about the adverse effects of excessive sugar intake through various channels such as social media, seminars, and educational programs in schools, as well as facilitating access to healthier food and beverage alternatives.

The Nonnegotiable Symbolic Value of Sugar in the Community

Sugar holds enduring cultural, economic, and social meanings that resist change. Beyond its material function, sugar embodies symbolic significance that shapes social life. Addressing NCDs thus requires transforming these deeply rooted values which often drive excessive consumption.

This study reveals that sugar consumption in Indonesia cannot be fully explained by conventional theories of symbolic value in economic sociology. While the theory of social distinction [22] links symbolic consumption to hierarchies, among Generation Z in Indonesia, sugar and sweet products signify collectivity and shared experience through celebrations, digital interactions, and everyday rituals. Consequently, symbolic value has shifted from exclusivity to inclusivity, becoming expressions of affection and lifestyle identity.

Existing theories overlook how the normalization of sugar in popular culture fosters collective overconsumption, reinforced by media, advertising, and digital algorithms. In the sugar market of Indonesia, symbolic value arises not only from cultural meanings but also from technological infrastructures that accelerate symbolic diffusion and strengthen emotional attachment. Accordingly, this study extends symbolic consumption theory by emphasizing digital, affective, and habitual dimensions. It proposes the concept of digital-symbolic value of consumption, where symbolic value is mediated by social media, digital marketing algorithms, and globalized consumption trends. Evidence demonstrates that young Indonesians perceive sugar not merely as sustenance but as an expression of digital identity and participation in the global emotional economy [34,36].

Funding

This study was funded by Universitas Indonesia under the Universitas Indonesia International Indexed Publication Grant of 2022. The opinions expressed are solely those of the authors and do not necessarily represent those of the funding institution.

Conflicts of Interest

None declared.

Multimedia Appendix 1
YouTube videos and their transcripts as primary data sources.
[DOCX File, 14 KB - infodemiology_v6i1e77261_app1.docx]

References

1. Nabil A. The philosophy of sweet cuisine in Javanese society [Article in Indonesian]. Atmago. 2023. URL: https://www.atmago.com/berita-warga/filosofi-masakan-manis-dalam-masyarakat-jawa_0cbd8c04-ae47-40b6-9442-a421f1b261c9 [accessed 2023-06-09]
2. The sugar dilemma: a brief review of the existence of sweeteners in Indonesia [Article in Indonesian]. Validnews. 2024. URL: https://validnews.id/opini/dilema-gula-telaah-singkat-eksistensi-pemanis-di-indonesia [accessed 2024-10-8]
3. Brownbill AL, Miller CL, Braunack-Mayer AJ. The marketing of sugar-sweetened beverages to young people on Facebook. Aust N Z J Public Health 2018 Aug;42(4):354-360. [doi: 10.1111/1753-6405.12801] [Medline: 29972262]
4. Zoellner J, Reid A, Porter K, Frederick C, Hilgart M, Ritterband L. Development of a digital behavioral intervention to reduce the consumption of sugar-sweetened beverages among rural Appalachian adults: multiphased, human-centered design approach. JMIR Hum Factors 2023 Feb 1;10:e41262. [doi: 10.2196/41262] [Medline: 36724036]
5. Beckert J. The social order of markets. Theor Soc 2009 May;38(3):245-269. [doi: 10.1007/s11186-008-9082-0]
6. Virdita R. The recommended daily sugar intake limits by WHO and the ministry of health, categorized by gender. [Article in Indonesian]. Kontan.co.id. 2022. URL: https://kesehatan.kontan.co.id/news/ini-batas-konsumsi-gula-per-hari-menurut-who-dan-kemenkes-berdasarkan-jenis-kelamin [accessed 2025-09-23]
7. Sutanto RA, Muljaningsih S. Analysis of factors influencing sugar imports in Indonesia [Article in Indonesian]. Kinerja J Ekon Manaj 2022;19(1):29-36. [doi: 10.30872/jkin.v19i1.10880]
8. Rook DW. The ritual dimension of consumer behavior. J Consum Res 1985 Dec;12(3):251. [doi: 10.1086/208514]
9. Hong HS, Lee SK. Text network analysis of research topics and trends on global health nursing literature from 1974 -2017. J Adv Nurs 2021 Mar;77(3):1325-1334. [doi: 10.1111/jan.14685] [Medline: 33617029]

XSL•FO
RenderX

10. Leem S, Shin J, Kim JY, Shim SR. Text network analysis to develop a search strategy for a systematic review. Appl Sci (Basel) 2024;14(19):8909. [doi: 10.3390/app14198909]
11. Paterson S, McInerney EEW, Evans SD. Using social network analysis as a tool in action research. J Community Psychol 2022 Sep;50(8):3421-3437. [doi: 10.1002/jcop.22844] [Medline: 35322891]
12. Beckert J. The moral embeddedness of markets. : Max Planck Institute for the Study of Societies; 2005 May 6 URL: https://www.econstor.eu/bitstream/10419/19920/1/dp05-6.pdf [accessed 2026-02-05]
13. Beckert J. Markets from meaning: quality uncertainty and the intersubjective construction of value. Cambridge J Econ 2020;44(2):285-301. [doi: 10.1093/cje/bez035]
14. Islam T, Attiq S, Hameed Z, Khokhar MN, Sheikh Z. The impact of self-congruity (symbolic and functional) on the brand hate: a study based on self-congruity theory. Br. Food J. 2019 May 1;121(1):71-88. [doi: 10.1108/BFJ-03-2018-0206]
15. Beckert J. Imagined Futures: Fictional Expectations and Capitalist Dynamics: Harvard University Press; 2016.
16. Harrington RA, Scarborough P, Hodgkins C, et al. A pilot randomized controlled trial of a digital intervention aimed at improving food purchasing behavior: the front-of-pack food labels impact on consumer choice study. JMIR Form Res 2019 Apr 8;3(2):e9910. [doi: 10.2196/formative.9910] [Medline: 30958277]
17. 49% of Generation Z continue to consume sugary foods and beverages despite concerns about developing diabetes [Article in Indonesian]. Rakyat Benteng. 2025 Feb 1. URL: https://rakyatbenteng.bacakoran.co/read/10769/49-gen-z-tetap-makan-dan-minum-minuman-manis-meski-khawatir-kena-diabetes [accessed 2025-02-06]
18. Fielding-Singh P. A taste of inequality: food's symbolic value across the socioeconomic spectrum. SocScience 2017;4:424-448. [doi: 10.15195/v4.a17]
19. Aspers P. Knowledge and valuation in markets. Theor Soc 2009 Mar;38(2):111-131. [doi: 10.1007/s11186-008-9078-9]
20. Nurmajesty H, Hardjosoekarto S, Herwantoko O, Ramadhani DC, Salsabila SA. Symbolic and material valuation of jamu: economic sociology of Indonesian jamu market. Asian J Bus Res 2022 Apr 1;12(1):99-123. [doi: 10.14707/ajbr.220122]
21. Klassen KM, Douglass CH, Brennan L, Truby H, Lim MSC. Social media use for nutrition outcomes in young adults: a mixed-methods systematic review. Int J Behav Nutr Phys Act 2018 Jul 24;15(1):70. [doi: 10.1186/s12966-018-0696-y] [Medline: 30041699]
22. Bourdieu P. Distinction: A Social Critique of the Judgement of Taste: Harvard University Press; 1984.
23. Luayya L, Hardjosoekarto S, Herwantoko O, Muhammaditya N. Digital group solidarity: rethinking the typology of Beckert's moral embeddedness in the contemporary market arena. Asian J Bus Res 2022 Dec 31;12(3):124-143. [doi: 10.14707/ajbr.220138]
24. Sitorus AMH, Hardjosoekarto S, Jahja RS, Herwantoko O, Anam FK. The moral embeddedness of labor market transactions in Indonesia: a study of Chinese migrant workers. Int J Soc Econ 2023 Nov 17;50(11):1485-1500. [doi: 10.1108/IJSE-11-2022-0737]
25. BPOM regulation No. 31 of 2018 on labeling of processed food [Article in Indonesian]. National Agency of Drug and Food Control. 2018. URL: https://peraturan.bpk.go.id/Details/219910/peraturan-bpom-no-31-tahun-2018 [accessed 2025-09-23]
26. Ratcliffe E, Baxter WL, Martin N. Consumption rituals relating to food and drink: a review and research agenda. Appetite 2019 Mar 1;134:86-93. [doi: 10.1016/j.appet.2018.12.021] [Medline: 30572007]
27. Moura AF, Aschemann-Witzel J. A downturn or a window of opportunity? How Danish and French parents perceive changes in healthy eating in the transition to parenthood. Appetite 2020 Jul 1;150:104658. [doi: 10.1016/j.appet.2020.104658] [Medline: 32171779]
28. Christensen BJ, Bestle SMS, Trolle E, et al. Evaluation of parental acceptability and use of intervention components to reduce pre-school children's intake of sugar-rich food and drinks. Int J Environ Res Public Health 2022 Jun 29;19(13):7967. [doi: 10.3390/ijerph19137967] [Medline: 35805623]
29. Pitt H, McCarthy S, Keric D, et al. The symbolic consumption processes associated with "low-calorie" and "low-sugar" alcohol products and Australian women. Health Promot Int 2023 Dec 1;38(6):daad184. [doi: 10.1093/heapro/daad184] [Medline: 38158741]
30. Mintz SW. Sweetness and Power: The Place of Sugar in Modern History: Viking; 1985.
31. Five countries that have implemented taxes on sugar sweetened beverages [Article in Indonesian]. CNN Indonesia. 2022 Sep 27. URL: https://www.cnnindonesia.com/ekonomi/20220927110704-92-853100/5-negara-yang-pajaki-minuman-bergula-tinggi [accessed 2023-01-08]
32. Prevent, prevent, and prevent: the world's voice against diabetes [Article in Indonesian]. Kementerian Kesehatan Republik Indonesia. 2018. URL: https://kemkes.go.id/id/cegah-cegah-dan-cegah-suara-dunia-perangi-diabetes [accessed 2025-09-23]
33. Nuraeni R, Simanjuntak SM. Sugar consumption of the adult age group: phenomenological study. J Berita Ilmu Keperawatan 2023;16(2):176-183. [doi: 10.23917/bik.v16i2.1892]
34. Guo IJ, Padmita AC, Matsuzaki M, et al. The use of social media to promote unhealthy food and beverage consumption among Indonesian children. BMC Nutr 2025 Mar 21;11(1):57. [doi: 10.1186/s40795-025-01040-2] [Medline: 40119452]
35. Hardjosoekarto S. Cabinet reshuffle [Article in Indonesian]. DetikNews. 2018. URL: https://news.detik.com/kolom/d-4187984/tata-ulang-kabinet [accessed 2022-09-23]
36. Generation Z and changes in culinary preferences [Article in Indonesian]. Kumparan. 2024. URL: https://kumparan.com/pengetahuan-umum/generasi-z-dan-perubahan-dalam-preferensi-kuliner-22CerdurENB [accessed 2026-02-05]

XSL•FO
RenderX

Abbreviations

BPOM: National Agency of Drug and Food Control
CERDIK: check health regularly, avoid smoking, exercise regularly, eat a healthy diet, rest adequately, and keep balance between body and mind
NCD: noncommunicable disease
SNA: social network analysis
SSB: sugar-sweetened beverage
SSF: sugar, salt, and fat
TNA: textual network analysis
WHO: World Health Organization

Edited by M Haupt; submitted 10.May.2025; peer-reviewed by A Baviera-Puig, D Ajayi; revised version received 09.Jan.2026; accepted 09.Jan.2026; published 27.Feb.2026.

Please cite as:
Manik EE, Hardjosoekarto S, Adnan R, Radhiatmoko R, Herwantoko O, Nurmajesty H, Darwan D, Putri FE, Pawening AS
Nonnegotiable Symbolic Value and Sugar-Driven Food Habits in Indonesia: Mixed Methods Study Using a Digital Sociological Approach
JMIR Infodemiology 2026;6:e77261
URL: https://infodemiology.jmir.org/2026/1/e77261
doi:10.2196/77261

© Ewina Efriani Manik, Sudarsono Hardjosoekarto, Ricardi Adnan, Radhiatmoko Radhiatmoko, One Herwantoko, Hakiki Nurmajesty, Darwan Darwan, Farrah Eriska Putri, Astuti Sri Pawening. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 27.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Leveraging AI for Analysis of Digital Health Information on Cancer Prevention Among Arab Youth and Adults: Content Analysis

Alia Komsany[1], PhD; Obada Al Zoubi[2], PhD; Laetitia Sebaaly[3], MA; Gabrielle Harrison[3], BA; Orysya Soroka[1], MS; Safa ElKefi[4], PhD; David Scales[1], MD, PhD; Erica Phillips[1], MS, MD; Laura C Pinheiro[1], MPH, PhD; Israa Ismail[1], BA; Perla Chebli[5], MPH, PhD

[1]Division of General Internal Medicine, Weill Cornell Medicine, New York City, NY, United States

[2]Independent Researcher, Boston, MA, United States

[3]Independent Consultant, New York, NY, United States

[4]School of System Sciences and Industrial Engineering, Watson College of Engineering, Binghamton University, New York, NY, United States

[5]NYU Langone Health, New York City, NY, United States

Corresponding Author:
Alia Komsany, PhD
Division of General Internal Medicine
Weill Cornell Medicine
Division of General Internal Medicine, 530 East 70th Street
New York City, NY, 10021
United States
Phone: 1 6469625036
Email: alk4019@med.cornell.edu

Abstract

Background: As TikTok (ByteDance) grows as a major platform for health information, the quality and accuracy of Arabic-language cancer prevention content remain unknown. Limited access to culturally relevant and evidence-based information may exacerbate disparities in cancer knowledge and prevention behaviors. Although large language models offer scalable approaches for analyzing online health content, their utility for short-form video data, especially in underrepresented languages, has not been well established.

Objective: We aimed to characterize and evaluate the quality of Arabic-language TikTok videos on cancer prevention and explore the use of large language models for scalable content analysis.

Methods: We used the TikTok research application programming interface and a GPT-assisted keyword strategy to collect Arabic-language TikTok videos (2021-2024). From an initial collection of 1800 TikTok videos, 320 were eligible after preprocessing. Of these, the top 25% (N=30) most-viewed were analyzed and manually coded for content type, cancer type, uploader identity, tone and register, scientific citation, and disclaimers. Video quality was assessed using the Patient Education Materials Assessment Tool for Audiovisual Materials for understandability and actionability, and the Global Quality Scale (GQS). GPT-4 was used to generate artificial intelligence annotations, which were compared to human coding for select variables.

Results: The top 25% (N=30) most-viewed videos amassed a total of 21.6 million views. Diet and alternative therapies were most common (15/30, 50%), which included recommendations to reduce hydrogenated oils, increase fruit and vegetable intake, and the use of traditional remedies such as garlic and black seed. Only 6.6% (2/30) of videos cited scientific literature. General cancer (15/30, 53%), breast (5/30, 17%), and cervical (4/30, 13%) cancers were most frequently mentioned. Doctors led 30% (9/30) of videos and were more likely to produce higher quality content, including significantly higher global quality scores (GQS=4, median 4, IQR 4-4 vs 3, median 3, IQR 2-3, P=.06). Over half of the videos had low understandability (16/30, 53%) and actionability (18/30, 60%). Emotionally framed content had the highest engagement across likes and shares, although this did not reach statistical significance (P=.08 and P=.05, respectively). However, emotional tone was significantly associated with lower GQS scores (P=.01). GPT-4 showed high agreement with human coders for cancer type (Cohen κ=1.0), strong agreement for GQS (κ=0.94), but low agreement for tone classification (κ=0.15), due to misclassification of emotional delivery from text-only input.

Conclusions: Arabic-language TikTok cancer prevention content is highly engaging but variable in quality, with emotionally framed videos attracting substantial attention despite lower informational value. Artificial intelligence-assisted tools show strong

XSL•FO
RenderX

potential for scalable, multilingual health content analysis, but multimodal approaches are needed to accurately interpret tonal and audiovisual features.

(JMIR Infodemiology 2026;6:e77888) doi:10.2196/77888

KEYWORDS

AI-driven content analysis; cancer prevention; engagement; digital health communication; TikTok

Introduction

Global Burden of Cancer

Worldwide, the cancer burden continues to rise, with an estimated 20 million new cases and 9.7 million deaths reported in 2022, projected to reach 35 million cases by 2050 [1]. The Arab world, which includes 22 countries, faces a particularly rapid increase in cancer incidence, with rates expected to rise 1.8-fold by 2030 [2]. Cancer ranks as a leading cause of death in many Arab nations, with Lebanon reporting the highest incidence of bladder cancer worldwide and Egypt contributing significantly to global liver cancer mortality [3,4]. Despite these rising trends, cancer prevention awareness remains limited, with low participation in screening programs and persistent misconceptions about cancer causes and treatment [5-8].

TikTok as a Source of Health Information

Social media platforms, such as TikTok, an emerging short-video app, have become major sources of health information worldwide [9]. During the COVID-19 pandemic, the platform saw a surge in health professionals and organizations using it to share medical knowledge and public health messages [10]. This shift highlighted the growing need for health care professionals to integrate video-based social media platforms, such as TikTok, into digital health communication strategies [11]. However, TikTok's global reach comes with region-specific challenges. Unlike other US-based platforms that apply universal moderation policies, TikTok uses localized moderation, tailoring its policies by region. This has raised concerns among Arabic-speaking users, particularly in North Africa, where dialect-specific moderation tools are lacking. Users often resort to strategies such as "algospeak" to avoid perceived censorship, and content moderation algorithms developed with limited dialect training data and nonnative annotators may misclassify or fail to flag harmful health misinformation. These dynamics highlight the urgent need to ensure equitable, culturally sensitive content governance as platforms such as TikTok become central to health communication ecosystems [12].

Arabic-Speaking Populations, an Understudied Demographic

Arabic-speaking populations, both in Arab countries and in diaspora communities, represent an understudied demographic in cancer prevention research [13,14]. Cultural beliefs, religious considerations, and misinformation often influence health behaviors, contributing to lower participation in preventive measures [15,16]. Language barriers further restrict access to reliable health information, with the Arabic language notably underrepresented in digital health research, along with the scarcity of validated Arabic-language health literacy tools and medical datasets [17-19]. This lack of research makes it challenging to assess the accuracy and effectiveness of Arabic-language health content, particularly on social media [20-22]. Understanding how cancer prevention messages are framed, their alignment with evidence-based guidelines, and their audience engagement is crucial for improving digital health communication among Arab nations and diaspora populations [23,24].

Large Language Models for Analyzing Digital Health Communication

Advancements in artificial intelligence (AI), particularly large language models (LLMs) such as GPT, offer new opportunities for analyzing digital health communication in understudied languages [25,26]. GPT-4 has demonstrated high accuracy in detecting sentiment, misinformation, and medical accuracy across multiple languages [26]. Unlike traditional natural language processing tools, GPT does not require extensive language-specific training, making it a scalable tool for content analysis [26-28]. This study sought to examine TikTok videos on cancer prevention in Arabic, assess the content quality of the videos, and explore the role of LLMs such as GPT-4 in evaluating digital health content. By identifying gaps in digital health communication, this research seeks to inform strategies for improving cancer prevention awareness among Arabic-speaking communities.

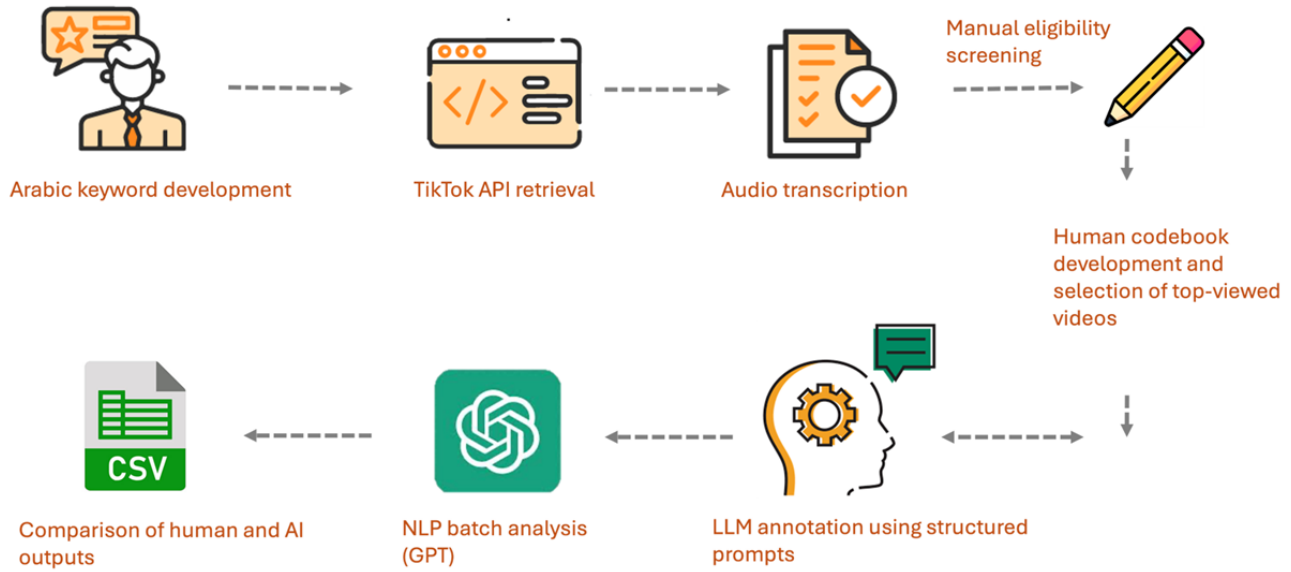
Methods

Data Source and Retrieval

We used a multistep analytic workflow to identify, process, and analyze Arabic-language TikTok videos related to cancer prevention, integrating human coding with AI-assisted annotation. Figure 1 provides an overview of the full workflow, including keyword development, video retrieval, transcription, eligibility screening, manual coding, AI-based annotation, and assessment of agreement between human and AI outputs.

XSL•FO
RenderX

Figure 1. Overview of the analytic workflow: Arabic-language TikTok videos were retrieved using an iterative keyword strategy, transcribed, and screened for eligibility. A subset was used to develop the coding framework, after which the top 25% (N=30) most-viewed videos were manually coded and annotated using a large language model. Agreement between human and AI classifications was assessed using Cohen κ coefficient. Full methodological details are provided in the Methods section. AI: artificial intelligence; API: application programming interface; LLM: large language model; NLP: natural language processing.



Using the TikTok application programming interface (API), we retrieved Arabic-language TikTok videos related to cancer prevention and the HPV vaccine from 2021 to 2024. This time frame was selected based on data from the Arab Youth Survey, which indicated an increasing trend in TikTok market penetration among young Arabs aged 18 to 24 years during this period. Specifically, daily TikTok usage more than doubled from 21% in 2020 to 50% in 2022, highlighting the platform's growing influence during this period [29]. Given that younger generations often play a key role in disseminating health information within their families, this period was considered optimal for capturing relevant content.

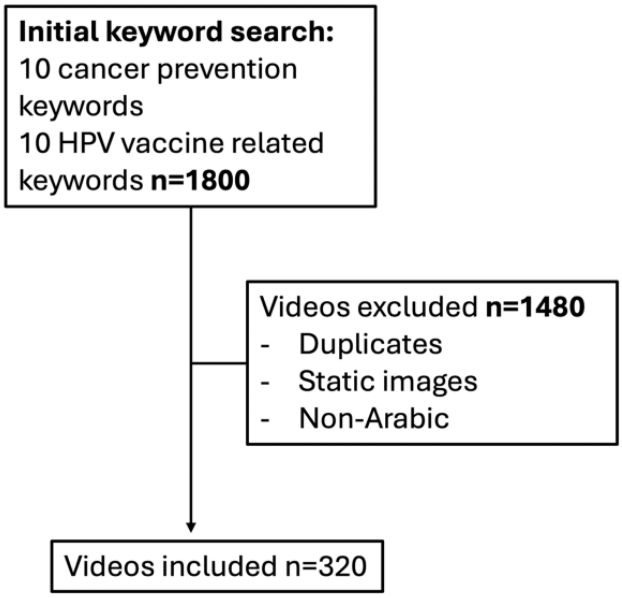
Search Strategy and Transcription

The research team developed an Arabic keyword (20 keywords) list focused on cancer prevention and HPV vaccination (Figure 1), which was iteratively refined and expanded using GPT-4 to ensure broad topical coverage (eg, "cancer prevention," "HPV vaccine," "tumor prevention," "healthy nutrition to prevent cancer," and "vaccination to protect against cervical cancer"). We then collected 1800 videos as above, excluding captions and comments (Figure 2). Videos were retrieved over several days because we were limited to 10,000 requests every 24 hours. Videos were transcribed using Sonix AI, selecting the Arabic transcription option. This Arabic transcription service is designed to handle a wide range of accents and regional dialects using advanced speech-to-text technology, though the model cannot be fine-tuned by users. All transcripts were manually reviewed for accuracy by the first author. No videos were truncated or cut during transcription, and the full audio of each TikTok was captured by Sonix's Arabic transcription. Sonix transcriptions focused on the lexical content of speech, the words and their semantic meaning, rather than acoustic properties such as pitch, pauses, or emphasis. As a result, nonverbal or visual information was not represented.

As part of our initial data collection using the TikTok API, we also retrieved publicly available aggregate engagement metrics for each video by country. These data, though not directly linked to the final analytic sample, provided insight into the geographic reach and broader engagement with Arabic-language cancer prevention content across global audiences. This included data on total views, likes, and shares by country, allowing us to assess which regions exhibited the highest levels of user interaction.

XSL•FO
RenderX

Figure 2. The filtering process used to identify top-performing Arabic-language TikTok videos related to cancer prevention. The final dataset included 30 videos in the top 25% most-viewed videos.



Codebook Development Using a Random 50-Video Subset

A random 16% (n=50) subset of the 320 videos was manually coded to create and refine the codebook. The research team reviewed coded data, discussed emerging categories, and reached consensus on definitions. The final codebook included classifications for content type, cancer type mentioned, tone and register, uploader type, presence of religious references, cautionary messages or disclaimers, target demographic, content quality indicators (global quality score and the Patient Education Materials Assessment Tool for Audiovisual Material [PEMAT AV]). Videos were coded as referencing scientific literature if they included direct citations or explicit mentions of scientific guidelines. This was used as a proxy for transparency, not as a definitive assessment of evidence-based accuracy. Tone and register were coded and simplified to 2 dimensions: emotional tone (eg, personal storytelling or emotional appeal, inclusive of expressive vocal delivery such as raised voice and dramatic pauses) and linguistic register (eg, casual or serious). This enabled consistency in both manual coding and AI-based classification and prioritized standardization over nuanced qualitative distinctions. If multiple tones were present within a video, the dominant style was recorded based on overall delivery. Videos were also coded for the specific types of cancer mentioned, including breast, colorectal, liver, pancreatic, brain, lung, cervical, oral, bladder, lymphoma, prostate, other cancers, and general cancer content (Textbox 1).

XSL•FO
RenderX

Textbox 1. Descriptions of content features and coding categories applied to Arabic TikTok videos.

Evidence-based

- Reference to a scientific study or cites scientific literature.

Emotional tone and linguistic register

- Describes how information is delivered:
  - Emotional tone: expresses personal experiences, urgency, or affective delivery. For example, “Dealing with this person causes cancer… it’s not food or drink that harms you; it’s people.”
  - Linguistic register: Casual: uses everyday, friendly language. For example, “I wish we could change this behavior, because it's literally a fountain of cancerous tumors.” Serious: uses formal or urgent phrasing. For example, “You have to come to the clinic. We have to do the mammogram. We have to take a sample. We have to..”

Content types

- Cancer prevention topics mentioned:
  - Diet alternative therapies: content promoting natural or nonclinical cancer prevention strategies, including the use of raw garlic, black seed, and dietary recommendations such as reducing hydrogenated oils.
  - Screening and early detection: mammograms and Papanicolaou test smears.
  - Vaccination: HPV vaccine.
  - Self-examination and symptom awareness: breast or testicular checks.
  - Smoking cessation: avoiding tobacco.
  - Stress or negativity: links between stress and cancer.
  - Survivor experience: sharing stories of cancer survival.
  - Chemical carcinogens: mention of environmental or food-based chemicals.

Cancers mentioned

- Specific and nonspecific cancer types cited, including:
  - General, breast, colorectal, liver, pancreatic, brain, lung, cervical, oral, bladder, lymphoma, prostate, other, and no cancer mentioned.

Speaker (doctor, self ID, or layperson)

- Whether the speaker self-identifies as a doctor (based on credentials in profile or linked accounts), self-identifies without affiliation, or does not claim any medical background.

Religious reference

- Mentions religious texts or spiritual framing of health advice.

Cautionary message or disclaimer

- Debunks a myth, adds a disclaimer, or highlights risks.

Target demographic

- Intended audience includes women, men, both genders, and youth.

Patient Education Materials Assessment Tool (PEMAT) understandability

- Percentage score based on clarity and ease of understanding.
- Coded as: high: 67%-100%, medium: 34%-66%, and low: 0%-33%.

PEMAT actionability

- Percentage score based on clear steps for action. Same categorization as understandability.

Global Quality Scale (GQS)

- 1=very poor (not useful) to 5=excellent (highly useful and comprehensive).

XSL•FO
RenderX

The coding framework included an assessment of uploader type which was classified into three categories: (1) doctors, whose credentials were corroborated through profile information (eg, "Dr" in username or bio); (2) self-identified doctors, who claimed a medical background without confirmable credentials; and (3) laypersons, with no stated or apparent medical affiliation. When the uploader status was ambiguous, the research team reviewed TikTok biographies, posted video content, and any linked social media profiles to determine affiliation. Clinic-affiliated accounts with a medical focus were also coded as doctor-led. Influencer status was assessed separately based on follower count, due to limited public information or unidentifiable handles. Creators with 100,000 or more followers were classified as influencers, regardless of medical background or professional identity. Videos were also coded for references to religious texts or beliefs, including phrases that framed health outcomes as divinely guided (*qadr*). Cautionary messages were defined as explicit statements that debunk myths, provide disclaimers, or warn against specific risks, such as ensuring that individuals with specific conditions avoid potential harms associated with alternative therapies or clarifying that the HPV vaccine is not exclusively for girls.

The target demographic of each TikTok video was categorized based on direct mention of the audience in the video, such as if the creator explicitly addressed a specific group (women or men), the reference to a specific age group (young people), or broad health advice, which was categorized as both genders.

Sample Selection Based on the 75th Percentile Cutoff

To focus on the most visible content, we selected the top 25% (N=30) most-viewed videos for detailed manual and AI analysis. To identify the most engaged content, we applied a 75th percentile cutoff. The decision to use the 75th percentile was based on established statistical principles for performance classification, a method that aligns with industry standards and prior studies that have used percentile-based cutoffs (75th percentile) to differentiate high-performing content from lower-engagement material [30]. This method helps mitigate the influence of outliers while allowing for the analysis of content that drives interaction and user engagement. By applying this approach, we ensured that this subset of videos reflected the most influential cancer prevention messaging on TikTok, aligning with established research methodologies in social media health communication [30].

Manual Coding and Interreliability Testing

The coding of the top 25% (N=30) of most-viewed videos was independently conducted by 2 study team members (AK and LS). The 2 coders met to finalize the codebook and resolve any coding discrepancies. To ensure coding reliability, interrater agreement was assessed using Cohen κ. Cohen κ was calculated for each coding category before reconciliation. Cohen κ values were interpreted using the following standard: values below 0.20 indicated slight agreement; 0.21-0.40, fair agreement; 0.41-0.60, moderate agreement; 0.61-0.80, substantial agreement; and 0.81-1.00, almost perfect agreement.

Assessment of Understandability, Actionability, and Quality

The PEMAT AV was used to assess the understandability and actionability of the videos [31]. The understandability section contains 13 items, and the actionability section includes 4 items, which can each be scored as 0 ("disagree"), 1 ("agree"), or "not applicable." For each section, PEMAT AV scores are calculated as percentages by dividing the points achieved by the items evaluated for the video. Therefore, higher values are indicative of higher understandability and actionability. The PEMAT AV has been widely used to evaluate health communication materials across formats, including videos, animations, and patient education modules for a range of topics, including chronic disease management, vaccine education, cancer prevention, and health literacy interventions [32,33]. We dichotomized PEMAT AV scores with 0%-66% considered as "low understandability," and 67%-100% as "high understandability." This threshold was based on the original guidance provided by Shoemaker et al [31], who recommended 70% as a benchmark for acceptable educational materials. The Global Quality Scale (GQS) was used to evaluate the overall quality, flow, and usefulness of each video's health information. This 5-point Likert scale has been widely used to evaluate the reliability and educational value of online medical and public health content [34]. The score represents the perception of the trained coder (in our case, 2 Arabic-speaking coders with experience evaluating health communication content). The GQS is scored based on the following scale 1="very poor quality, missing information, not useful"; 2="generally poor quality, some missing information, very limited use"; 3="moderate quality, some information adequately discussed, somewhat useful"; 4="good quality, most relevant information discussed, useful"; and 5="excellent quality, all relevant information discussed, very useful" [35].

AI-Based Annotation

To generate AI-based annotations, we used a 1-shot prompting approach, in which a single, structured prompt was provided to the model to classify each video based on predefined categories from our manually generated codebook [36]. The prompt included clear definitions to guide the model's interpretation. For instance, the model was instructed as follows: "Answer the questions as precisely and faithfully as possible using the provided context. The provided text is in Arabic with various dialects. Provide the answers in JSON format. Ensure that all responses are directly based on the provided text without assumptions or external information."

Questions included items such as "List any cancers mentioned in the text: Options: General, Breast, Colorectal, Liver, Pancreatic, Brain, Lung, Cervical, Oral, Bladder, Lymphoma, Prostate, Other and No cancer mentioned." For GQS scoring: "A Global Quality Score (GQS): Options: a score from 1 to 5 with:1: Poor quality, poor flow, and not useful 2: Generally poor quality and flow, but some information is listed 3: Moderate quality and flow, but some important information is poorly discussed, 4: Good quality and flow, but some topics are not covered 5: Excellent quality and flow, and very useful."

XSL•FO
RenderX

For each coding category, such as cancer type or GQS score, the model was instructed to return 1 label per video, such as output the answers in the following JSON format: “cancers_mentioned”: [<list from cancers list>], “GQS_Score”: “<one score>”

AI outputs were generated in Python (Python Software Foundation) through a batch analysis pipeline [37]. The GPT model was optimized using iterative prompt engineering, refining the input structure to improve consistency in classification and fidelity to the codebook. This enabled efficient, scalable annotation of video characteristics while minimizing ambiguity, and facilitated direct comparison between AI-generated and human-coded classifications. AI-generated outputs were reviewed by human coders and systematically compared to manual annotations for the top-viewed videos. Cohen κ was used to evaluate interrater reliability between human and AI classifications. This analysis was conducted using the Cohen kappa score function in Python, which measures agreement beyond chance for categorical variables.

Statistical Analysis

As this was an exploratory analysis, GPT-generated annotations were limited to 3 key categories: cancer type, tone and register, and GQS score. Descriptive statistics were used to summarize video characteristics, and inferential analyses were conducted using nonparametric tests due to high variability in the data. The Wilcoxon Rank Sum test was used to compare median engagement metrics (likes and shares) across groups, as engagement data were highly skewed. Fisher Exact Tests were used for categorical comparisons where sample sizes were small or expected cell counts were low. These statistical methods were selected to ensure robustness despite nonnormal distributions and heterogeneous group sizes.

Ethical Considerations

This study analyzed publicly available TikTok videos related to cancer prevention using the TikTok Research API and did not involve direct interaction with human participants. No private, identifiable, or nonpublic user data were collected. All data were accessed and analyzed in accordance with TikTok’s terms of service and research data use policies.

Results

Overview

The final analytic dataset included 30 TikTok videos, representing the top 25% most-viewed content from an initial pool of 320 Arabic-language videos related to cancer prevention (cutoff at 59,640 views). These 30 videos collectively amassed 21.6 million views, 445,000 likes, and 146,000 shares (see Tables 1 and 2).

XSL•FO
RenderX

Table 1. Characteristics of TikTok videos (N=30).

Characteristic	Values, n (%)
Content type	
Diet and alternative therapies	15 (50)
Screening and early detection	6 (20)
HPV[a] vaccination	3 (10)
Self-examination and symptoms to look out for	1 (3)
Smoking cessation	1 (3)
Stress and negativity	1 (3)
Survivor experience	1 (3)
Chemical carcinogens	1 (3)
Cancers mentioned	
General cancer	15 (50)
Breast cancer	6 (20)
Cervical cancer	4 (13)
Colon cancer	2 (7)
Bladder cancer	1 (3)
Multiple cancers	1 (3)
Testicular cancer	1 (3)
Emotional tone and register	
Casual	16 (53)
Serious	8 (27)
Emotional	6 (20)
Target demographic	
Both genders	19 (63)
Women	8 (27)
Men	1 (3)
Young people	2 (7)
Led by doctors (corroborated or self-identified)	
Yes (credentials corroborated)	9 (30)
Yes (self-identified or no confirmable credentials)	6 (20)
No (layperson did not state medical affiliation)	13 (43)
Medical clinic affiliated	2 (7)
Evidence-based	
Yes	2 (7)
No	28 (93)
Cautionary message or disclaimer	
No	16 (53)
Yes	14 (47)
PEMAT[b] understandability	
High (≥67%)	14 (47)
Low (≤66%)	16 (53)
PEMAT actionability	

XSL•FO
RenderX

Characteristic	Values, n (%)
High (≥67%)	15 (50)
Low (≤66%)	15 (50)
Religious reference	
Yes	6 (20)
No	24 (80)
GQS[c]	
1 (very poor)	1 (3)
2 (poor)	5 (17)
3 (moderate)	7 (20)
4 (good)	17 (60)
5 (excellent)	0 (0)

[a]HPV: human papillomavirus.

[b]PEMAT: Patient Education Materials Assessment Tool.

[c]GQS: Global Quality Scale.

Table 2. Other characteristics of TikTok videos (N=30).

Engagement	Minimum-maximum	Median (IQR)
Like count	524-116,493	3062 (1370-18,629)
Share count	37-36,403	751.5 (199-4019)
View count	59,116-8,490,149	176,391 (92,166-592,176)

Emotional Tone and Linguistic Register

Casual was the most common (16/30, 53%) code, followed by serious (8/30, 27%) and emotional (6/30, 20%). Emotional videos were more engaging than others, receiving the highest median likes and share counts. However, emotional videos were associated with lower global quality scores (median 2, IQR 2-3), while serious and casual videos received higher scores (median 4, IQR 4-4). The difference in GQS across tones was statistically significant (P=.01), indicating that higher engagement did not correspond with higher content quality.

Content Types

The most common content was diet and alternative therapies (15/30, 50%), including content promoting the use of raw garlic, black seeds, or reducing hydrogenated oils. This was followed by screening and early detection (6/30, 20%) and HPV vaccination (3/30, 10%). Other content types, including self-examination, stress, smoking cessation, and survivor stories, each appeared in only 1 of 30 (3%) videos. The videos in the top 25% (N=30) most-viewed videos focused on HPV vaccination, and those that mentioned cervical cancer were created by self-identified doctors, including 2 verified and 1 unverified account. Two used a serious tone, and 1 used a casual tone. All 3 videos received a global quality score (GQS) of 4.

Cancers Mentioned

The most frequently mentioned cancers were general cancer (16/30, 53%), with breast cancer (5/30, 17%) and cervical cancer (4/30, 13%) commonly referenced. Less frequently mentioned were colon (2/30, 7%), bladder (1/30, 3%), multiple cancers (1/30, 3%), and testicular cancer (1/30, 3%).

Speaker (Doctor, Self-Identified, or Layperson)

Among the top 25% (N=30) most-viewed videos, 17 (57%) were led by individuals identifying as doctors, including 9 (30%) verified doctors, 6 (20%) unverified, and 2 (7%) accounts affiliated with medical clinics. The remaining 13 of 30 (43%) accounts were led by laypeople. Of the 17 doctor-led accounts, 8 (47%) verified and 2 (12%) unverified accounts met the threshold for influencer status. Among the 13 laypeople accounts, 4 (31%) accounts met the influencer criteria. While only 4 of the 13 nondoctor-led accounts met the influencer threshold (≥100,000 followers), 3 additional laypeople's accounts had a substantial following between 20,000 and 60,000 followers.

Target Demographics, Religious Reference, and Presence of Cautionary Message

Religious framing appeared in 6 of 30 (20%) videos, with references to divine (*qadr*) will or spiritual health advice. Further, 14 of the 30 (47%) videos included a cautionary message or disclaimer, such as warnings about misinformation or clarification on cancer risk factors. Most videos targeted both genders (19/30, 63%), followed by women (8/30, 27%), young people (2/30, 7%), and men (1/30, 3%).

Evidence-Based, Patient Education Materials Assessment Tool: Understandability and Actionability

Only 2 (7%) of the top 25% (N=30) most-viewed videos explicitly cited scientific literature or guidelines. A total of 53%

XSL•FO
RenderX

(16/30) of videos were rated low on understandability (score ≤66%). Notably, all 6 doctor-led videos promoting diet and alternative therapies scored high (≥67%) for understandability, while all 9 layperson videos in that category scored low. Furthermore, 50% (15/30) of videos were rated low on actionability. Among diet and alternative health-related videos, those led by doctors were more likely to be actionable by 83% (25/30) compared to those led by laypeople (17/30, 56.6%), though this difference was not statistically significant (*P*=.58).

About GQS

A total of 60% (18/30) of videos were rated as good (score of 4), 23% (7/30) as moderate (score of 3), and 17% (5/30) as poor (score of 2). Only 1 (3%) video was rated very poor (score of 1), and none were rated excellent. Videos led by doctors promoting diet and alternative therapies had significantly higher GQS scores than those led by laypeople (*P*=.06).

Human and AI Agreement

Agreement between human coders was high across all domains (κ=0.84). There was perfect agreement between human and AI annotations for cancer type (κ=1.0) and strong agreement for GQS scoring (κ=0.94), though most discrepancies occurred between scores of 3 (moderate) and 4 (good), indicating difficulty distinguishing between mid and high-quality content. Agreement was lower for tone classification (κ=0.15), with AI misclassifying emotional delivery when relying on text-based input alone.

Geographic Reach of Arabic-Language Cancer Prevention Content on TikTok

TikTok platform data indicated that Arabic-language cancer prevention content generated substantial engagement from users in both Arab-majority countries (eg, Egypt, Jordan, and Saudi Arabia) and diaspora contexts such as the United States, France, and Germany. The United States ranked in the top 10 for total views, highlighting the global reach of Arabic-language cancer messaging.

There was high agreement in cancer type between human and AI annotations (κ=1.0), and similarly high agreement in GQS scoring (κ=0.94). Tone classification showed lower concordance. While the AI model correctly identified many casual and serious tones, it misclassified emotional content in several cases, resulting in slight overall agreement (κ=0.15). Manual review confirmed that GPT-based annotation performed reliably across dialectal variations from multiple Arabic-speaking countries, indicating that cross-dialect consistency is achievable when coupled with human verification.

Discussion

Principal Findings

This study makes 3 contributions to the broader literature on online health videos and TikTok specifically. First, it provides the first systematic analysis of Arabic-language TikTok videos on cancer prevention. Second, it identifies an engagement quality gap in Arabic language cancer prevention content, extending prior English-language findings that emotionally charged posts often receive higher engagement [38]. Third, the present study advances methodological research by evaluating GPT-4's performance on Arabic transcript-only inputs from short-form videos: the model demonstrated high reliability for structured categorical variables but low reliability for tone classification, underscoring the need for multimodal approaches that incorporate audio and visual cues. Together, these contributions deepen the understanding of Arabic-language TikTok health communication and illustrate both the potential and current limitations of AI-assisted content analysis across global digital ecosystems. While prior work has shown that LLMs can support qualitative researchers by generating themes from social media corpora in a single prompt [39], their use for systematic content analysis of short-form video data has not, to our knowledge, been previously demonstrated.

Within our sample, videos promoting diet and alternative therapies were among the most viewed. Studies in other cultural contexts have similarly shown that traditional or community-based health guidance often thrives because it is relational, linguistically resonant, and perceived as more trustworthy than institutional messages [40]. Research on TikTok in English more broadly echoes this pattern: a content analysis of health-related "EduTok" videos found that audiences most frequently engaged with educational posts related to diet, exercise, and sexual health, suggesting consistent user interest in familiar, lifestyle-oriented themes [41]. Together, these parallels suggest that what circulates widely on Arabic TikTok may reflect a broader sociocultural logic in which familiarity and affective connection drive credibility and engagement. It is important to note, however, that patterns related to content type and cancer type in this study are driven primarily by a small number of highly represented categories (eg, diet and alternative therapies and screening and early detection), which reflects the distribution of high-engagement Arabic-language TikTok content rather than a comprehensive representation of cancer prevention topics.

Beyond these general patterns, our data illustrate how an engagement quality gap appears specifically within Arabic TikTok cancer prevention content. Emotional tone was associated with higher engagement, even when informational quality was low. One widely circulated video, for example, claimed that "toxic people, not food or genetics," cause cancer, an emotionally resonant but scientifically inaccurate claim that drew considerable engagement. These findings align with prior TikTok-specific studies showing that affectively charged content outperforms factual or instructional posts [41]. Importantly, our results do not imply that emotional tone alone determines virality; rather, they suggest that emotional framing, creator identity, and algorithmic amplification together create conditions in which lower-quality but more affectively engaging messages can spread widely.

Targeting of young people was limited despite TikTok's prominence among youth. Only a small proportion of high-engagement videos explicitly addressed adolescents or young adults, even though early-life behaviors, such as HPV vaccination, tobacco use, diet, and physical activity, are critical for cancer prevention. The absence of youth-directed content suggests a missed opportunity to leverage TikTok as a public health tool for early prevention messaging. Instead, content

XSL•FO
RenderX

often targeted adult women or general audiences, which may reflect creator demographics or cultural communication norms. However, it is important to recognize that reliance on TikTok for health information is not limited to adolescents. Many Arabic speakers in diaspora contexts, including Arab Americans, turn to social media due to linguistic and cultural barriers in traditional health care settings [17]. Consistent with this, our platform data showed high engagement from diaspora countries, including the United States, emphasizing TikTok's role as a transnational source of Arabic language health information. Furthermore, prior research has shown that immigrants frequently rely on online platforms for relatable and accessible health content, making the quality of digital communication a critical equity issue [42]. These patterns parallel findings from US studies showing that African American and Hispanic adults were more likely than White adults to seek health information through social media during the COVID-19 pandemic, underscoring how communication inequities can drive platform reliance among marginalized groups [42]. Future research should examine whether these same patterns extend to Arabic-language health content on other short-form video platforms such as Instagram Reels (Meta), YouTube Shorts (Google LLC), Facebook Watch (Meta), and Snapchat Spotlight (Snap Inc), which share similar algorithmic dynamics but may differ in moderation and audience reach.

Most of the analyzed videos lacked references to peer-reviewed literature or established clinical guidelines, and only 30% (9/30) were led by doctors (whose credentials could be corroborated). It is important to distinguish between being evidence-based and citing sources. While a video may communicate content that aligns with scientific consensus, the absence of explicit references may reduce credibility, especially in digital environments where users rely on transparency to assess trustworthiness. Although doctor-led videos produced higher quality content on average (eg, higher GQS scores), professional identity alone did not ensure high understandability or actionability (high understandability suggests that most viewers, including those with limited health literacy, can grasp the essential messages being communicated). This is particularly salient given that populations with lower health literacy are more likely to rely on TikTok for health information [43].

Interpreting PEMAT AV and GQS together provides important insight into the quality of doctor-led content. PEMAT AV, which is validated for audiovisual materials, assesses whether information is communicated clearly and whether viewers are given actionable guidance, whereas GQS reflects a broader, more subjective appraisal of overall informational quality and usefulness. These differences are therefore meaningful rather than contradictory and underscore the value of using PEMAT AV and GQS as complementary measures when evaluating short-form health content [33]. Future studies may benefit from incorporating additional quality frameworks to further capture dimensions of informational rigor and communicative nuance.

Our findings show that doctor-led videos achieved higher quality scores but did not generate comparable engagement. This aligns with prior research on Arabic language health content across other platforms. For instance, studies of Arabic breast cancer videos on YouTube have shown that videos produced by trusted institutions tend to be more accurate but far less popular than those by individual users. This recurring pattern across platforms suggests that credibility alone does not guarantee visibility, a consistent challenge in health communication on social media. Effective health communication may therefore require pairing evidence-based content with narrative appeal, cultural resonance, and accessible delivery formats to compete with misinformation and emotionally engaging but lower-quality material.

Content moderation practices also play a role in shaping what health information circulates across Arabic-speaking regions. Unlike platforms that apply uniform global policies, TikTok relies on region-specific moderation teams and language filters, which may inconsistently flag or downrank health misinformation. While this study focused on cancer prevention, the engagement and credibility patterns we observed echo those reported in Arabic language TikTok vaccine content, suggesting that visibility dynamics may be shaped more by platform design and algorithmic incentives than by the specific health topic itself [22].

Finally, our study underscores the promise of LLMs for scalable analysis of health-related content in underrepresented languages. GPT-based coding achieved high reliability in classifying categorical variables such as cancer type and video quality (κ=0.94 to κ=1.0). However, it performed less consistently in detecting tone and register, particularly emotional delivery.

This finding stands in partial contrast to prior work, which has reported strong LLM performance in multilingual sentiment analysis of social media content [26]. A key difference, however, lies in the approach, which used large-scale labeled data from high-resource, text-based platforms. In contrast, our 1-shot method relied on transcripts of Arabic TikTok videos, where much of the emotional tone is conveyed through audiovisual cues such as intonation and facial expression not captured in text alone.

Emerging multimodal AI systems such as Gemini (Google LLC) and Google Cloud Video Intelligence, which can jointly interpret audio, visual, and textual inputs, hold promise for overcoming these limitations and enabling more contextually accurate annotation of short-form health content across languages. Applied tools such as ScreenApp, which automatically integrates speech-to-text, speaker detection, and scene-level video analysis, further illustrate how multimodal pipelines are already being used to extract meaningful patterns from audiovisual content [44].

The observed pattern in our findings points to 2 directions for future work: (1) developing a labeled Arabic TikTok dataset to refine tone detection in LLMs, and (2) adopting multimodal pipelines that combine audio, visual, and text cues to better capture emotion and on-screen gestures. This hybrid approach, using LLMs for large-scale triage and multimodal or human review for nuanced interpretation, offers a scalable path for improving health content analysis across languages.

These findings have important implications for public health outreach in Arabic-speaking communities, both within the Arab world and across diaspora populations. Given the limited availability of culturally tailored, Arabic-language health

XSL•FO
RenderX

education materials and the growing reliance on social media for information, TikTok represents both a powerful tool and a potential vector for misinformation. While it can amplify accurate messaging, it also enables the rapid spread of misinformation. Addressing this will require multipronged strategies: empowering health care providers with the skills to create engaging content, leveraging AI for content monitoring, and partnering with trusted community figures to amplify reliable messages. Future interventions may leverage narrative and emotionally resonant formats to pair evidence-based content with styles that match user engagement preferences.

Limitations

This study has several limitations. First, it focused on videos in the top 25% (N=30) most-viewed, which introduces an engagement bias and thus may not fully represent the broader landscape of Arabic-language cancer prevention content on TikTok. In addition, AI comparison was applied only to these top-viewed videos, rather than the full set of 320 eligible videos, limiting our ability to fully assess the scalability and generalizability of AI-based annotation across the entire dataset.

Second, we did not analyze or control for video length, which may influence engagement metrics. However, because the sample was drawn from the top 25% (N=30) of most-viewed videos, it likely reflects videos optimized for typical TikTok viewing behavior, minimizing major variability in duration effects. Third, while the Patient Education Materials Assessment Tool and GQS frameworks are validated tools, they may not fully capture the stylistic and communicative nuances characteristic of short-form, audiovisual social media content. Fourth, to facilitate AI-based classification, tone and linguistic register were simplified into broad categories, which may have limited the detection of the more subtle or culturally embedded communication styles typically captured through qualitative analysis. Fifth, although use of the TikTok API helped reduce algorithmic sampling bias, platform-specific ranking mechanisms and personalization features may still have influenced which videos achieved high visibility. In addition, analyses of content type and cancer type were concentrated within a narrow subset of categories due to the limited representation of other topics, and findings should therefore not be generalized to less frequently represented cancers or prevention behaviors. Finally, the AI model operated solely on transcribed audio, analyzing text without access to visual or prosodic cues such as facial expressions, gestures, or intonation, elements that are often common in video-based communication.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this paper.

Conflicts of Interest

None declared.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Al-Muftah M, Al-Ejeh F. Cancer incidence and mortality estimates in Arab countries in 2018: a GLOBOCAN data analysis. Cancer Epidemiol Biomarkers Prev 2023;32(12):1738-1746 [FREE Full text] [doi: 10.1158/1055-9965.EPI-23-0520] [Medline: 37733340]
3. Arafa MA, Rabah DM, Farhat KH. Rising cancer rates in the Arab world: now is the time for action. East Mediterr Health J 2020;26(6):638-640 [FREE Full text] [doi: 10.26719/emhj.20.073] [Medline: 32621496]
4. Abbas NF, Aoude MR, Kourie HR, Al-Shamsi HO. Uncovering the epidemiology of bladder cancer in the Arab world: a review of risk factors, molecular mechanisms, and clinical features. Asian J Urol 2024 Jul;11(3):406-422 [FREE Full text] [doi: 10.1016/j.ajur.2023.10.001] [Medline: 39139531]
5. Sayan M, Eren AA, Tuac Y, Langoe A, Alali B, Aynaci O, et al. Prostate cancer awareness in the middle east: a cross-sectional international study. JCO Glob Oncol 2024;10:e2400171. [doi: 10.1200/GO.24.00171] [Medline: 38991182]
6. Ahmed HAA, Abbas MH, Hussein HA, Nasr RSF, Lashen AA, Khaled H, et al. Correction: cervical cancer screening uptake in Arab countries: a systematic review with meta-analysis. BMC Cancer 2024;24(1):1478 [FREE Full text] [doi: 10.1186/s12885-024-13233-2] [Medline: 39614258]
7. Marouf A, Tayeb R, Alshehri GD, Fatani HZ, Nassif MO, Farsi AH, et al. Public perception of common cancer misconceptions: a nationwide cross-sectional survey and analysis of over 3500 participants in Saudi Arabia. J Family Med Prim Care 2023;12(6):1125-1132. [doi: 10.4103/jfmpc.jfmpc_1753_22] [Medline: 37636192]
8. Elshami M, Naji SA, Dwikat MF, Al-Slaibi I, Alser M, Ayyad M, et al. Myths and common misbeliefs about colorectal cancer causation in Palestine: a national cross-sectional study. JCO Glob Oncol 2024;10:e2300295 [FREE Full text] [doi: 10.1200/GO.23.00295] [Medline: 38166235]
9. Kirkpatrick CE, Lawrie LL. TikTok as a source of health information and misinformation for young women in the United States: survey study. JMIR Infodemiology 2024;4:e54663 [FREE Full text] [doi: 10.2196/54663] [Medline: 38772020]

XSL•FO
RenderX

10. Basch CH, Hillyer GC, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 2022;34(5):367-369 [FREE Full text] [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
11. Song S, Xue X, Zhao YC, Li J, Zhu Q, Zhao M. Short-video apps as a health information source for chronic obstructive pulmonary disease: information quality assessment of TikTok videos. J Med Internet Res 2021;23(12):e28318 [FREE Full text] [doi: 10.2196/28318] [Medline: 34931996]
12. Elswah M. Moderating Maghrebi Arabic Content on Social Media. URL: https://cdt.org/wp-content/uploads/2024/09/2024-09-26-CDT-Research-Global-South-Moderating-Report-English-Arabic-final.pdf [accessed 2026-01-20]
13. Zaidi M, Fantasia HC, Ahmed R, Lee DN, Valdman O, Poghosyan H, et al. Experiences with cancer screenings among arabic-speaking refugee women. Nurs Womens Health 2025;29(2):109-119. [doi: 10.1016/j.nwh.2024.09.004] [Medline: 39947246]
14. Bhattacharya R, Chen N, Shim I, Kuwahara H, Gao X, Alkuraya FS, et al. Massive underrepresentation of Arabs in genomic studies of common disease. Genome Med 2023;15(1):99 [FREE Full text] [doi: 10.1186/s13073-023-01254-8] [Medline: 37993966]
15. Hammoud MM, White CB, Fetters MD. Opening cultural doors: providing culturally sensitive healthcare to Arab american and American muslim patients. Am J Obstet Gynecol 2005;193(4):1307-1311. [doi: 10.1016/j.ajog.2005.06.065] [Medline: 16202719]
16. Abboud S, De Penning E, Brawner BM, Menon U, Glanz K, Sommers MS. Cervical cancer screening among Arab women in the United States: an integrative review. Oncol Nurs Forum 2017;44(1):E20-E33 [FREE Full text] [doi: 10.1188/17.ONF.E20-E33] [Medline: 27991600]
17. Al-Jumaili AA, Ahmed KK, Koch D. Barriers to healthcare access for Arabic-speaking population in an English-speaking country. Pharm Pract (Granada) 2020;18(2):1809 [FREE Full text] [doi: 10.18549/PharmPract.2020.2.1809] [Medline: 32477432]
18. Bergman L, Nilsson U, Dahlberg K, Jaensson M, Wångdahl J. Health literacy and e-health literacy among Arabic-speaking migrants in Sweden: a cross-sectional study. BMC Public Health 2021;21(1):2165 [FREE Full text] [doi: 10.1186/s12889-021-12187-5] [Medline: 34823499]
19. Reich H, Hegerl U, Rosenthal A, Allenhof C. Arabic-language digital interventions for depression in German routine health care are acceptable, but intervention adoption remains a challenge. Sci Rep 2024;14(1):12097 [FREE Full text] [doi: 10.1038/s41598-024-62196-8] [Medline: 38866810]
20. Alfaris E, Alhazzani Y, Alkhenizan A, Irfan F, Almoneef N, Alyousefi N, et al. Assessing the validity of health messages used by the Saudi public in WhatsApp. Patient Prefer Adherence 2023;17:67-73 [FREE Full text] [doi: 10.2147/PPA.S397661] [Medline: 36632071]
21. Alammari OB, Odeh O, Alotaibi R, Zamzam M, Alghamdi AS. Assessment of the most popular educational content in Arabic available through four social media applications in Saudi Arabia that explains idiopathic clubfoot deformity. Cureus 2025;17(1):e76976. [doi: 10.7759/cureus.76976] [Medline: 39912023]
22. Sallam M, Al-Mahzoum K, Alkandari L, Shabakouh A, Shabakouh A, Ali A, et al. Descriptive analysis of TikTok content on vaccination in Arabic. AIMS Public Health 2025;12(1):137-161 [FREE Full text] [doi: 10.3934/publichealth.2025010] [Medline: 40248416]
23. Plackett R, Kaushal A, Kassianos AP, Cross A, Lewins D, Sheringham J, et al. Use of social media to promote cancer screening and early diagnosis: scoping review. J Med Internet Res 2020;22(11):e21582 [FREE Full text] [doi: 10.2196/21582] [Medline: 33164907]
24. Conley CC, Otto AK, McDonnell GA, Tercyak KP. Multiple approaches to enhancing cancer communication in the next decade: translating research into practice and policy. Transl Behav Med 2021;11(11):2018-2032 [FREE Full text] [doi: 10.1093/tbm/ibab089] [Medline: 34347872]
25. Lu Z, Peng Y, Cohen T, Ghassemi M, Weng C, Tian S. Large language models in biomedicine and health: current research landscape and future directions. J Am Med Inform Assoc 2024;31(9):1801-1811. [doi: 10.1093/jamia/ocae202] [Medline: 39169867]
26. Rathje S, Mirea D, Sucholutsky I, Marjieh R, Robertson CE, Van Bavel JJ. GPT is an effective tool for multilingual psychological text analysis. Proc Natl Acad Sci U S A 2024;121(34):e2308950121 [FREE Full text] [doi: 10.1073/pnas.2308950121] [Medline: 39133853]
27. Miao H, Li C, Wang J. A future of smarter digital health empowered by generative pretrained transformer. J Med Internet Res 2023;25:e49963 [FREE Full text] [doi: 10.2196/49963] [Medline: 37751243]
28. Garbarino S, Bragazzi NL. Evaluating the effectiveness of artificial intelligence-based tools in detecting and understanding sleep health misinformation: comparative analysis using Google Bard and OpenAI ChatGPT-4. J Sleep Res 2024;33(6):e14210. [doi: 10.1111/jsr.14210] [Medline: 38577714]
29. Survey TAY. URL: https://arabyouthsurvey.com/en/ [accessed 2026-01-20]
30. Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MS. What people "like": analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. J Med Internet Res 2018;20(6):e10227 [FREE Full text] [doi: 10.2196/10227] [Medline: 29903694]

XSL•FO
RenderX

31. 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;96(3):395-403 [FREE Full text] [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
32. Furukawa E, Okuhara T, Liu M, Okada H, Kiuchi T. Evaluating online and offline health information with the patient education materials assessment tool: protocol for a systematic review. JMIR Res Protoc 2025;14:e63489 [FREE Full text] [doi: 10.2196/63489] [Medline: 39813665]
33. Dimitroyannis R, Fenton D, Cho S, Nordgren R, Pinto JM, Roxbury CR. A social media quality review of popular sinusitis videos on TikTok. Otolaryngol Head Neck Surg 2024;170(5):1456-1466. [doi: 10.1002/ohn.688] [Medline: 38431902]
34. Gudapati JD, Franco AJ, Tamang S, Mikhael A, Hadi MA, Roy V, et al. A study of global quality scale and reliability scores for chest pain: an Instagram-post analysis. Cureus 2023;15(9):e45629 [FREE Full text] [doi: 10.7759/cureus.45629] [Medline: 37868472]
35. Li M, Yan S, Yang D, Li B, Cui W. YouTube as a source of information on food poisoning. BMC Public Health 2019;19(1):952 [FREE Full text] [doi: 10.1186/s12889-019-7297-9] [Medline: 31311523]
36. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023;25:e50638 [FREE Full text] [doi: 10.2196/50638] [Medline: 37792434]
37. Atkinson CF. AI-pocalypse now: automating the systematic literature review with SPARK (Systematic Processing and Automated Review Kit) - gathering, organising, filtering, and scaffolding. MethodsX 2025;14:103129 [FREE Full text] [doi: 10.1016/j.mex.2024.103129] [Medline: 39846014]
38. Abbas MJ, Khalil LS, Haikal A, Dash ME, Dongmo G, Okoroha KR. Eliciting emotion and action increases social media engagement: an analysis of influential orthopaedic surgeons. Arthrosc Sports Med Rehabil 2021;3(5):e1301-e1308 [FREE Full text] [doi: 10.1016/j.asmr.2021.05.011] [Medline: 34712967]
39. Deiner MS, Honcharov V, Li J, Mackey TK, Porco TC, Sarkar U. Large language models can enable inductive thematic analysis of a social media corpus in a single prompt: human validation study. JMIR Infodemiology 2024;4:e59641 [FREE Full text] [doi: 10.2196/59641] [Medline: 39207842]
40. Cruz ML, Christie S, Allen E, Meza E, Nápoles AM, Mehta KM. Traditional healers as health care providers for the Latine community in the United States, a systematic review. Health Equity 2022;6(1):412-426 [FREE Full text] [doi: 10.1089/heq.2021.0099] [Medline: 35801152]
41. O'Donnell N, Jerin SI, Mu D. Using TikTok to educate, influence, or inspire? A content analysis of health-related EduTok videos. J Health Commun 2023;28(8):539-551. [doi: 10.1080/10810730.2023.2234866] [Medline: 37434532]
42. Goldsmith LP, Rowland-Pomp M, Hanson K, Deal A, Crawshaw AF, Hayward SE, et al. Use of social media platforms by migrant and ethnic minority populations during the COVID-19 pandemic: a systematic review. BMJ Open 2022;12(11):e061896 [FREE Full text] [doi: 10.1136/bmjopen-2022-061896] [Medline: 36396309]
43. Patil U, Kostareva U, Hadley M, Manganello JA, Okan O, Dadaczynski K, et al. Health literacy, digital health literacy, and COVID-19 pandemic attitudes and behaviors in U.S. college students: implications for interventions. Int J Environ Res Public Health 2021;18(6):3301 [FREE Full text] [doi: 10.3390/ijerph18063301] [Medline: 33806763]
44. Cascella M, Semeraro F, Montomoli J, Bellini V, Piazza O, Bignami E. The breakthrough of large language models release for medical applications: 1-year timeline and perspectives. J Med Syst 2024;48(1):22 [FREE Full text] [doi: 10.1007/s10916-024-02045-3] [Medline: 38366043]

Abbreviations

AI: artificial intelligence
API: application programming interface
GQS: Global Quality Scale
LLM: large language model
PEMAT AV: Patient Education Materials Assessment Tool for Audiovisual Materials

Edited by E Lee; submitted 21.May.2025; peer-reviewed by KH Lin, C Baur, Y Zhang, J Antoun; comments to author 18.Aug.2025; revised version received 30.Dec.2025; accepted 09.Jan.2026; published 09.Feb.2026.

Please cite as:
Komsany A, Al Zoubi O, Sebaaly L, Harrison G, Soroka O, ElKefi S, Scales D, Phillips E, Pinheiro LC, Ismail I, Chebli P
Leveraging AI for Analysis of Digital Health Information on Cancer Prevention Among Arab Youth and Adults: Content Analysis
JMIR Infodemiology 2026;6:e77888
URL: https://infodemiology.jmir.org/2026/1/e77888
doi:10.2196/77888
PMID:

XSL•FO
RenderX

©Alia Komsany, Obada Al Zoubi, Laetitia Sebaaly, Gabrielle Harrison, Orysya Soroka, Safa ElKefi, David Scales, Erica Phillips, Laura C Pinheiro, Israa Ismail, Perla Chebli. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 09.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Online Discussions of Men's Mental Health on Reddit and YouTube: Cross-Sectional Mixed-Methods Infodemiological Study

Anurag Shekhar[1*], PhD; Musawenkosi Donia Saurombe[1*], PhD

Department of Industrial Psychology and People Management, College of Business and Economics, University of Johannesburg, Johannesburg, South Africa

[*]all authors contributed equally

Corresponding Author:
Musawenkosi Donia Saurombe, PhD
Department of Industrial Psychology and People Management
College of Business and Economics
University of Johannesburg
D-Ring 409, Cnr Auckland Park and University Roads
UJ Auckland Park Kingsway Campus
Johannesburg, 2006
South Africa
Phone: 27 0604031649
Email: donaynaysauz24@gmail.com

Abstract

Background: Male mental health remains a major global concern, with men underrepresented in mental health care and overrepresented in suicide statistics. Masculine norms that link emotional restraint with strength can discourage help-seeking and vulnerability. Anonymous digital spaces such as Reddit (Reddit Inc) and YouTube (Google LLC) have become informal support environments where men share experiences and emotions outside traditional constraints. Understanding these interactions offers insight into masculine identity and help-seeking behavior.

Objective: This study examines how men discuss and negotiate mental health within anonymous online communities. It explores whether these spaces support emotional openness, peer validation, and challenges to hegemonic masculinity norms. It triangulates digital discourse with survey and interview data to assess how these patterns align with men's lived experiences and perceived barriers to support.

Methods: This cross-sectional, exploratory mixed methods study analyzed publicly available online discourse from Reddit (n=740 posts) and YouTube (n=6287 comments). The qualitative component included 23 adult men (aged 18-55 years, predominantly Asian and employed) recruited via LinkedIn (Microsoft) who completed an anonymous online survey. Of these, 9 volunteered for follow-up semistructured interviews. Data underwent computational text mining using the Natural Language Toolkit and National Research Council Lexicon for word frequency and emotion analysis, followed by Braun and Clarke's 6-phase reflexive thematic analysis. Online discourse patterns were then compared with survey and interview data. Theoretical frameworks included hegemonic masculinity, toxic positivity, and peer-support theory.

Results: Four themes emerged across the datasets: (1) normalizing emotional expression, (2) mutual validation and peer support, (3) coping through humor and irony, and (4) pushback against toxic positivity and societal norms. Emotion analysis showed prominent expressions of sadness, fear, trust, and anger across the Reddit and YouTube corpus. Survey data showed that 20 of 23 (87%) respondents reported having no safe offline space to discuss mental health. Interview participants (n=9) largely confirmed digital discourse themes, though some divergence emerged regarding whether humor functioned as deflection or connection.

Conclusions: This study combines large-scale analysis of online discourse with qualitative triangulation across Reddit, YouTube, surveys, and interviews. Theoretically, it extends inclusive masculinity theory into anonymous online contexts, showing how digital platforms enable men to negotiate emotional expression outside traditional masculine constraints. It introduces the concept of "digitally mediated sanctuaries" to describe online spaces where men practice vulnerability and mutual support with less social risk. From an infodemiological perspective, the findings show how mental health information and peer-support narratives circulate and gain legitimacy within male-dominated online communities. Findings can inform gender-sensitive digital mental health interventions that build on features men already use, including humor, anonymity, and peer validation. Digital peer environments may complement formal mental health services by reducing help-seeking barriers for men hesitant to access traditional care.

XSL•FO
RenderX

(JMIR Infodemiology 2026;6:e81315) doi:10.2196/81315

KEYWORDS

men's health; masculinity; mental health; social media; online peer support; emotional expression; infodemiology

Introduction

Background

Mental health is one of the most urgent global challenges today. More than 300 million people live with depression, and anxiety disorders affect a similar number. Suicide remains a leading cause of death among young people worldwide [1-5]. The COVID-19 pandemic, ongoing wars, and climate-related uncertainty have added to the burden of psychological distress across populations [6]. While mental illness affects people of all genders, men face particular challenges in how they experience and seek help for mental health struggles.

Men are far less likely than women to seek help when distressed. At the same time, they are more likely to die by suicide or turn to substance use as a coping mechanism [7-11]. This mismatch is often linked to social expectations about how men should behave. Traits such as toughness, emotional control, and self-reliance are rewarded in many cultures, leading many men to feel ashamed to speak about their emotions or ask for help [7,9]. These expectations can become internalized and may prevent men from recognizing when they need support.

Beyond traditional masculine norms, the cultural phenomenon of "toxic positivity" makes it harder for men to speak honestly about their emotions. This refers to the belief that people should always appear upbeat and positive, even when struggling [12-14]. Social pressure to stay strong and look happy can discourage men from reaching out, especially if they already believe that vulnerability is a sign of weakness.

For men who belong to marginalized racial, ethnic, or sexual minority groups, the risks are even higher. According to minority stress theory, people who face discrimination carry an added burden that affects their mental health [15]. These men often find it even harder to speak openly or seek support in traditional settings such as the workplace, school, or community.

In response to these barriers, many men are now discussing mental health and social constraints on anonymous platforms such as Reddit. Subreddits such as *r/MaleMentalHealth* have become spaces where men share personal struggles, reflect on emotional pain, and seek advice from peers without fear of judgment. These forums allow men to express what they often suppress in real life: vulnerability, sadness, frustration, and confusion, while receiving validation from others who understand their experience. Men frequently share culturally resonant imagery exploring emotional neglect and its consequences (refer to "Findings" subheading for examples).

YouTube has also emerged as an influential platform for men's mental health dialogue. Videos such as TED Talks (TED Conferences, LLC) on male vulnerability, emotional well-being, and societal expectations around masculinity often attract thousands of comments. These comment sections reflect strong engagement from male viewers, many of whom share their own experiences, question stereotypes, or offer support to others. These digital interactions reveal how men navigate mental health struggles and challenge dominant narratives around emotional suppression.

Infodemiology refers to the science of the distribution and determinants of information in electronic media, especially the internet, to inform public health and policy [16]. It involves analyzing digital trace data (such as social media posts, search queries, and online discussions) to understand health information–seeking behaviors, disease surveillance, and public perceptions of health issues. From an infodemiological perspective, understanding how mental health information circulates and is negotiated in these digital spaces is critical for public health.

This study examines this discourse by analyzing 740 anonymized posts from Reddit's r/MaleMentalHealth, 6287 comments on 4 TEDx Talks related to men's mental health on YouTube, and qualitative data from 23 men who participated in surveys and 9 follow-up interviews. Using a mixed methods design and reflexive thematic analysis, the research examines how men express distress, negotiate masculine norms, and connect with others in online and conversational spaces. By triangulating these data sources, the study provides deeper insight into how anonymous digital platforms and lived experiences together shape contemporary understandings of men's mental health.

This study has 3 integrated objectives: a qualitative objective to identify and describe emergent themes in men's mental health discourse on Reddit and YouTube through reflexive thematic analysis; a quantitative objective to analyze linguistic patterns and emotional tone across the combined corpus using computational text mining and emotion lexicon analysis; and a mixed methods objective to triangulate digital discourse findings with survey and interview data, examining whether online expressions of vulnerability, peer support, and norm negotiation align with men's lived experiences and help-seeking behaviors.

Masculinity, Mental Health, and Help-Seeking

A growing body of evidence positions men's mental health as a public health concern, particularly due to high suicide rates and the persistent gap in help-seeking behavior [8-11]. Despite significant mental health challenges, men are considerably less likely than women to pursue professional psychological support [7,9,11]. One major explanation lies in sociocultural norms surrounding masculinity, which discourage emotional openness and reinforce the stigma around mental illness [11].

Hegemonic masculinity, a dominant cultural ideal, promotes traits such as stoicism, emotional control, and self-reliance, which discourage men from acknowledging or disclosing distress [17]. Stronger conformity to these norms, particularly among middle-aged and older men, has been linked to elevated rates of depression [18]. While traditional masculinity may grant

XSL•FO
RenderX

social status and power, it also contributes to poor mental health outcomes by framing vulnerability as weakness [8].

Nondisclosure of distress is a common consequence of these norms, with men often internalizing the belief that seeking help is a failure of masculinity [19]. This silence can compound feelings of isolation and shame. Reddit and similar forums have emerged as crucial alternative spaces where men can speak more freely [20]. Studies show that depressive symptoms in men are associated with emotional suppression, and peer environments encouraging vulnerability can help reduce such tendencies [19].

The theory of hegemonic masculinity explains how dominant ideals stigmatize traits such as fear, sadness, or emotional dependence as unmanly [21]. Men are often taught to suppress these feelings, leading to shame, substance abuse, and in some cases, suicide [7]. Toxic masculinity describes how these expectations can manifest in extreme forms, such as mocking emotional expression, which undermines men's mental well-being [7]. Scholars have argued for a redefinition of masculinity that includes emotional expression as a form of strength, not weakness [7].

Some men grow up believing their worth is tied to their productivity, independence, and emotional detachment rather than to who they are intrinsically. This belief is vividly reflected in digital spaces, where commentary and shared narratives capture men's sense of conditional value. A widely circulated quote expressing that men are only loved conditionally based on what they provide resonates deeply in Reddit and YouTube discussions, reinforcing the idea that men must earn their emotional worth and complicating their willingness to express vulnerability or seek help.

Anonymity and Computer-Mediated Communication

Anonymity and computer-mediated communication dynamics help explain why digital spaces enable vulnerability and norm-challenging discourse among men. According to Suler's [22] online disinhibition effect, features such as anonymity, invisibility, and asynchronicity reduce fear of judgment and lower self-presentation pressures, allowing individuals to disclose personal experiences more freely. Similarly, Joinson [23] found that perceived anonymity in computer-mediated communication increases self-disclosure by diminishing social desirability concerns. Within male peer communities, these affordances temporarily suspend the social costs associated with violating masculine norms, creating psychological safety to articulate emotions or seek help without reputational threat [24].

Empirical research supports this process of online identity demarginalization. McKenna and Bargh [25] demonstrated that individuals with marginalized or stigmatized identities used anonymous online groups as low-risk environments to explore and affirm aspects of self that were constrained offline—often preceding greater authenticity in face-to-face contexts [25]. Similarly, De Choudhury and De [26] found that Reddit's structural affordances, including the use of "throwaway accounts," enabled open discussions of highly stigmatized topics such as suicidality, depression, and anxiety [26]. These findings align with this study's observation that Reddit and YouTube comment threads act as psychological safety zones where men renegotiate emotional boundaries and experiment with alternative masculinities beyond offline constraints.

Intersectionality and Minority Stress

Masculinity does not operate in isolation. Intersectionality theory, as developed by Crenshaw [27,28], urges researchers to consider how gender norms intersect with race, sexuality, and other identity categories. For example, Black men may face heightened pressure to appear emotionally strong, perceiving vulnerability as doubly threatening in the context of racial stereotypes [11]. Similarly, sexual minority men face a unique combination of masculine role expectations and societal stigma, increasing their psychological burden.

Meyer's minority stress theory further explains how systemic discrimination contributes to chronic psychological strain among individuals who identify as Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, and more (LGBTQ+) [15]. Concealing one's identity or experiencing rejection from peers can result in internalized stigma and elevated risk for depression and anxiety [15,29]. Although Reddit posts are largely anonymous, themes of intersectional stress can still emerge subtly in discussions, particularly in references to race, sexuality, or social isolation. Several posts across subreddits, including r/Science, actively engaged with research-based evidence on LGBTQ+ mental health disparities, often receiving strong support from users with similar lived experiences. This demonstrates how online communities reflect and reinforce awareness of minority stress and intersectional vulnerabilities.

Peer Support and Inclusive Masculinities in Online Spaces

Anonymity and accessibility make online peer support forums an increasingly attractive alternative to formal therapy [20,30,31]. Platforms such as Reddit enable men to engage in open emotional dialogue without fear of judgment or social consequences [32]. Prior research shows that men often favor informal help from peers or anonymous online groups over traditional mental health services [11].

Online communities, particularly those with a clear ethos of respect and empathy, foster a sense of psychological safety, where men can disclose vulnerabilities and receive affirmation [20]. These environments can reduce self-stigma by normalizing mental health struggles and offering shared narratives of hope and recovery [20,32,33]. Peer interaction on such platforms also aligns with mechanisms of universality and altruism found in group therapy contexts, where offering support to others boosts one's own sense of efficacy and connection.

This shift is part of a broader rethinking of masculinity. Emerging research on "positive" or "inclusive" masculinities recognizes that masculinity is plural and malleable. Rather than adhering to rigid ideals, these evolving masculinities emphasize emotional literacy, empathy, and self-care. Campaigns such as Movember's [34] "man of more words" exemplify how men are beginning to reclaim emotional expression as a strength rather than a liability. These shifts echo the American Psychological Association's call to promote healthier models of masculinity grounded in psychological well-being [35].

XSL•FO
RenderX

However, online discourse frequently reveals the gap between these aspirational messages and lived reality, with users critiquing the performative contradiction in societal responses.

Within this framework, online peer support operates through 2 interrelated mechanisms. At the individual level, it functions as a buffering process that mitigates the emotional strain associated with restrictive masculine norms by offering validation, empathy, and normalization of emotional expression. At the collective level, it acts as a social-influence mechanism that can gradually modify normative boundaries through repeated exposure to alternative expressions of masculinity. As men encounter peers modeling vulnerability and mutual care, these interactions may incrementally foster norm flexibility, reducing the perceived social cost of emotional openness. While our cross-sectional design cannot infer causal transformation, these discursive dynamics illustrate how digital peer spaces create psychosocial conditions conducive to both immediate coping and gradual normative reorientation.

Toxic Positivity as a Moderating Influence

The growing cultural expectation to "stay positive" or "look on the bright side" has evolved into what commentators call toxic positivity, a mindset that discourages the open expression of distress. For men, this can reinforce long-standing social norms that equate emotional control with strength and vulnerability with weakness [24,36]. While maintaining optimism can sometimes support coping, constant pressure to appear strong or grateful often invalidates genuine pain and contributes to emotional isolation. Many men, having internalized these expectations, delay seeking help until distress becomes overwhelming [37].

In digital environments, these pressures can take new forms. Social media platforms often encourage simplified motivational interactions [38]. Yet the anonymity and reach of platforms such as Reddit and YouTube can also reduce social risk, allowing men to question or reject these pressures in ways that may not feel possible offline. In communities such as r/MaleMentalHealth, participants often encourage one another to be honest, practice self-compassion, and challenge the idea that men must always appear in control. Humor and irony frequently emerge as tools that make this disclosure feel safer and more socially acceptable, helping men connect without losing face [39].

Toxic positivity, therefore, both constrains and shapes men's emotional expression online. It moderates how peer support unfolds by influencing which forms of vulnerability are considered acceptable and how they are communicated. Recognizing this tension helps explain how men negotiate authenticity in online spaces that promote both connection and restraint.

Beyond Binary Masculinity Models: Toward Digital Pluralism

The relationship between masculinity and men's well-being is increasingly understood through frameworks that reject singular or static definitions of manhood. Contemporary scholarship recognizes plural masculinities—the idea that men express gender in multiple, shifting ways across different social settings and identities [21,40,41]. This pluralistic view contrasts with traditional binary models that separate "hegemonic" and "subordinate" masculinities, instead emphasizing fluidity and context. Scholars have also argued for building upon men's strengths to promote positive masculinities that encourage empathy, help-seeking, and emotional growth [42,43].

This theoretical perspective extends meaningfully into digital environments, where technological affordances such as anonymity and audience reach enable digital pluralism—the coexistence of multiple, flexible masculinity performances within the same interactional space. Because much existing research on masculinity is based on offline contexts, digital discourse reveals distinct mechanisms that support identity fluidity. The relative anonymity of online platforms reduces the perceived social risk of vulnerability [44], allowing men to articulate emotions or uncertainties without reputational threat.

Within anonymous forums and peer networks such as Reddit and YouTube, men shift seamlessly between vulnerability, humor, stoicism, and care, forming masculinities that are situational rather than hierarchically fixed [21,43]. Humor and irony often function as socially acceptable strategies for emotional expression, helping men discuss mental health without contravening masculine display rules [39,45]. In this sense, online spaces become sites of performative negotiation, where masculine identity is continually reshaped through interaction and mutual recognition [46,47].

Thus, masculinity in digital contexts can be understood as a dynamic interplay of diverse voices negotiating authenticity, belonging, and emotional legitimacy in real time. Analyzing these online exchanges moves beyond binary or linear models of masculinity change, positioning digital discourse as a relational arena where men experiment with and normalize alternative masculinities [24,48].

Recent experimental research further supports the transformative potential of exposure to counter-normative narratives. Borinca and Gkinopoulos [49] demonstrate that heterosexual men's endorsement of traditional masculinity norms can be experimentally reduced when exposed to minority perspectives (ie, narratives of gay men embracing nontraditional relational and domestic roles). Their findings suggest that contact with minority masculinities fosters emotional openness by lowering the perceived threat to masculine identity. Complementing this, Borinca et al [50] show that when men perceive broader societal shifts toward more inclusive gender norms, they experience less discomfort when engaging in counter-stereotypical behaviors. Together, these studies provide causal evidence that exposure to alternative or minority masculinities, precisely what occurs in online communities where diverse vulnerability narratives coexist, can facilitate gradual attitude change. This supports our interpretation that digital peer spaces are not merely reflective but also generative environments for masculine norm transformation.

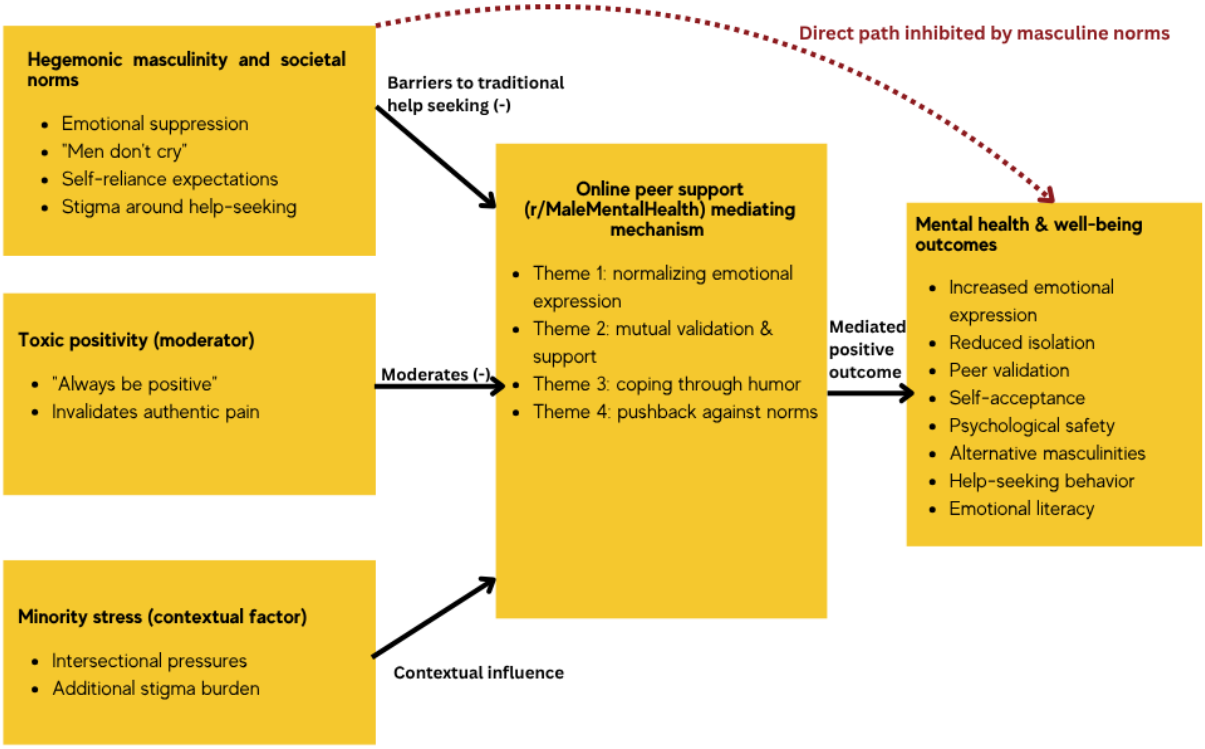
Theoretical Framework

The theoretical framework presented in Figure 1 synthesizes findings from our thematic analysis with established theoretical perspectives to explain how men navigate mental health

XSL•FO
RenderX

discourse in online peer support environments. This model positions online peer support as a critical mediating mechanism that transforms the traditionally inhibitory effects of masculine norms into positive mental health outcomes.

Figure 1. Theoretical framework: online peer support as a mediating mechanism between hegemonic masculine norms and mental health outcomes.



Framework Components and Relationships

This framework conceptualizes online peer support, such as Reddit's r/MaleMentalHealth and YouTube posts and comment sections, as a mediating mechanism between hegemonic masculine norms and men's mental health outcomes. Instead of a direct, inhibited pathway from masculine socialization to poor mental health, the framework identifies a mediated route where digital peer spaces enable men to engage in emotional expression, validation, and community support.

Key mediating mechanisms include the four themes: normalizing emotional expression, mutual validation, coping through humor, and pushback against norms. These are facilitated through what we term anonymous authenticity, where men feel safe expressing vulnerability online despite societal pressures to remain stoic.

Toxic positivity is positioned as a moderator, subtly shaping how masculine norms affect participation. While it allows some emotional expression, it restricts men to "positive" displays, often silencing authentic distress. Our analysis shows how men actively resist this pressure by cocreating norms that value genuine emotional honesty over superficial optimism.

Minority stress is framed as a contextual amplifier, not merely an individual-level variable. For men facing intersectional stigma—due to race, sexuality, or other factors—online platforms often provide the only accessible support space, intensifying their relevance.

This framework contributes to theory in 3 key ways: first, it moves beyond binary masculinity models by showing that masculine norms can either inhibit or enable help-seeking depending on context; second, it reinterprets peer support theory for the digital era by replacing identity disclosure with shared emotional experience; and third, it highlights online resistance, where men collectively challenge both traditional stoicism and toxic positivity through everyday digital interactions.

Overall, the framework reflects how online communities are reshaping men's mental health discourse, offering alternative pathways for healing, solidarity, and emotional growth. Within this framework, online peer support operates as a discursive environment that may enable norm flexibility and gradual reframing of masculine expression, rather than producing direct behavioral change. These digital interactions highlight how emotional openness is negotiated and tested within specific peer contexts, offering insight into potential pathways of masculine norm evolution.

Methods

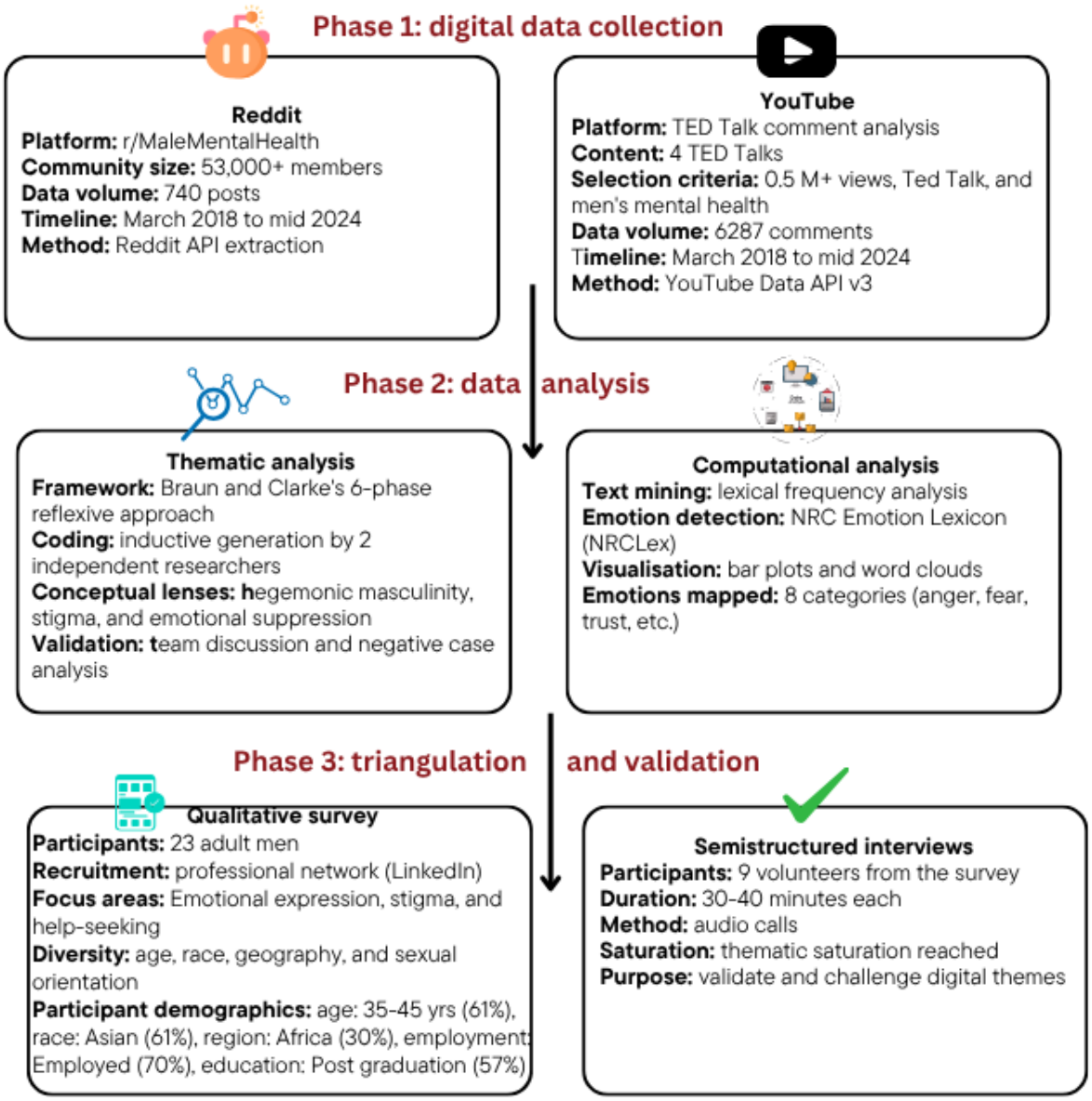
Study Design

This study used an exploratory mixed methods design using a convergent approach [51], in which qualitative and quantitative data were collected concurrently and integrated during interpretation to explore how men articulate and navigate mental health concerns in digital environments. The study size was determined by capturing all available posts from the selected Reddit thread during the collection period and all comments from the 4 YouTube videos. No additional sampling or truncation was applied. Primary data were drawn from 2 publicly accessible online platforms (Reddit and YouTube) to identify key themes and patterns in user-generated discourse. These

XSL•FO
RenderX

themes were then triangulated and validated through a follow-up qualitative survey and semistructured interviews with men from diverse backgrounds (refer to Figure 2). The design was informed by reflexive thematic analysis, supported by computational text mining and emotion analysis.

Figure 2. Convergent mixed methods design: data collection, parallel analysis, and triangulation phases. API: application programming interface; NRC: National Research Council.



Data Sources and Sampling

Reddit

A total of 740 posts, including original submissions and top-voted posts, were collected from the subreddit r/MaleMentalHealth, a community with over 53,000 members at the time of data collection. The posts spanned from the subreddit's creation in March 2018 through mid-2024. Data were collected using the Reddit application programming interface and included only publicly available, anonymized content. No personally identifiable information was recorded.

YouTube

We also extracted user comments from 4 TEDx Talks on YouTube that focused on men's mental health. Videos were selected using purposive sampling based on three criteria: (1) central relevance to men's mental health, (2) TEDx branding, and (3) a minimum of half a million views. Using approved access via the YouTube Researcher Program and the YouTube Data application programming interface (version 3), we extracted 6287 comments totaling over 343,060 likes in aggregate.

XSL•FO
RenderX

The selected videos were Why Big Boys Don't Cry by Gareth Griffith (TEDxUniversityofBristol), Mental Health Crisis in Men by Allan Kehler (TEDxDalhousieU), Suffering in Silence: The Emotional Abuse of Men by Dr Timothy Golden (TEDxWallaWallaUniversity), and We Need to Talk About Male Suicide by Steph Slack (TEDxFolkestone). These videos collectively reflect diverse perspectives on male mental and emotional health, making them highly relevant for analyzing public engagement and discourse in the comment sections.

Data Analysis

Qualitative Data Analysis

Primary outcomes included themes reflecting emotional expression, peer support, and attitudes toward masculinity. Secondary analytic focus examined discourse patterns and emotion lexicon categories (eg, anger, joy, and sadness) as indicators of affective tone. We used Braun and Clarke's [52,53] 6-phase reflexive thematic analysis framework to examine patterns in Reddit and YouTube discussions on men's mental health. The process began with data familiarization, where the primary researcher read and reread Reddit posts from r/MaleMentalHealth and high-engagement YouTube comments. Posts and comments were prioritized based on the number of likes to foreground socially endorsed discourse. In the initial coding phase, the first researcher applied open, inductive coding using Python libraries such as pandas, Natural Language Toolkit, and word cloud, allowing themes to emerge directly from the data rather than from pre-existing frameworks. These codes were then discussed with a second researcher in a process of reflexive dialogue to deepen interpretation and challenge assumptions. The goal was not to test coding consistency but to enhance analytical rigor and reflexive awareness. Collaborative discussions helped refine theme meanings and ensure sensitivity to multiple perspectives.

Codes were next grouped into preliminary themes, representing shared latent meanings within the data. In phase 4, both researchers reviewed and refined these candidate themes to ensure they were internally consistent, distinct, and sufficiently represented the dataset. Phase 5 involved clearly naming and defining the final themes, supported by the development of a thematic map to visualize intertheme relationships. Analytical credibility was strengthened through collaborative reflexivity, negative case analysis, and a shared interpretive process between researchers, ensuring that the final 4 themes captured both the emotional depth and social nuance of men's mental health discourse across platforms. As part of reflexive rigor, researchers actively sought negative cases, instances that challenged a candidate theme, and used them to refine theme boundaries and labels. Researchers monitored thematic saturation throughout the iterative coding process, considering it reached when no substantively new codes or insights emerged during the final rounds of analysis and when subsequent data confirmed the stability of the existing thematic structure.

Quantitative Data Analysis

To support and deepen qualitative interpretation, we conducted lexical frequency analysis and emotion detection using the National Research Council Emotion Lexicon via the National Research Council Lexicon Python library. The most frequently occurring words were visualized using bar plots and word clouds. Emotional tone was mapped, providing a computational snapshot of the affective landscape of digital discourse. These outputs were descriptive context for qualitative interpretation and were not used as inferential evidence.

Mixed Methods Integration

To complement the social media findings and assess whether digital themes aligned with lived experiences, the study used a sequential triangulation approach, comprising an online qualitative survey followed by in-depth interviews. A total of 23 adult men were recruited via professional networks such as LinkedIn to complete the survey, which included open-ended questions on emotional expression, stigma, coping strategies, and help-seeking behavior. Recruitment followed a purposive and convenience sampling approach through professional networks and online well-being groups. Interviewees were selected from survey respondents who consented to follow-up. This initial phase enabled a broader capture of individual reflections and language around mental health, with several posts and comments reinforcing or extending themes identified from Reddit and YouTube.

From this group, 9 participants volunteered for follow-up semistructured interviews, which were conducted via audio calls. Each interview lasted approximately 30-40 minutes and focused explicitly on validating or challenging the themes that emerged from the online data. Interview questions were framed to test the resonance of digital narratives, such as humor, emotional scarcity, peer validation, and pushback against toxic positivity, in real-life contexts.

Survey and interview participants were primarily recruited through LinkedIn posts shared by the first author. As a result, the participant pool largely comprised professionals with higher education backgrounds and access to online networks. While this recruitment approach facilitated voluntary participation and ensured ethical transparency, it likely introduced an occupational and educational bias, which should be considered when interpreting the findings.

Importantly, thematic saturation was reached by the 7th interview, with no substantially new codes or themes emerging thereafter. This aligns with Braun and Clarke's [54] large-scale saturation study, which found that 94% of frequent themes emerge within the first 6 interviews and 97% by the twelfth, supporting the adequacy of smaller, focused qualitative samples. Additionally, Sim et al [55] recommend a sample size of 6-15 participants for in-depth qualitative studies. Given that the primary aim was to test pre-existing digital themes rather than generate entirely new ones, the sample of 9 interviews was both methodologically sufficient and theoretically sound.

In summary, the survey and interviews provided meaningful triangulation of online discourse, offering depth, confirmation, and nuance to the digital findings, while remaining within accepted qualitative research standards for sample adequacy and saturation (Table 1).

XSL•FO
RenderX

Table 1. Demographic characteristics of survey participants (N=23).

Variable and category	Value, n (%)
Age group (years)	
35-45	14 (60.9)
45-55	4 (17.4)
25-35	4 (17.4)
18-25	1 (4.3)
Race	
Asian	14 (60.9)
White	4 (17.4)
Other	3 (13)
Mixed race	1 (4.3)
Black African	1 (4.3)
Region	
Africa	7 (30.4)
Asia	7 (30.4)
Europe	4 (17.4)
Other	2 (8.7)
Australia	1 (4.3)
North America	1 (4.3)
South America	1 (4.3)
Employment status	
Employed	16 (69.6)
Other (student, self-employed, etc)	4 (17.4)
Unemployed	3 (13)
Education	
Postgraduate	13 (56.5)
Graduate	10 (43.5)

Missing Data Handling

No missing data were present in the Reddit (n=740) or YouTube (n=6287) datasets, as all publicly accessible posts and comments were extracted by the software. For the survey component, all 23 respondents completed all mandatory fields, resulting in 0% missing data. Optional open-ended responses had a 100% completion rate. Interview data were complete for all 9 participants, with no session interruptions or incomplete transcripts. Therefore, no imputation procedures or missing data analyses were necessary.

Researcher Positionality and Reflexivity

The research team comprised a male primary researcher (first author) with 14 years of human resource management leadership experience and mental health expertise, and a female secondary researcher (second author) with an industrial psychology specialization. This gender-balanced approach provided complementary perspectives while mitigating analytical bias in examining masculine discourse dynamics. The primary researcher's positioning as male enabled insider understanding of masculine socialization experiences, while the secondary researcher provided external validation of interpretive processes. Researchers maintained reflexive memos documenting analytical decisions and potential biases throughout data collection and analysis.

Ethical Considerations

This study was approved by the University of Johannesburg Department of Industrial Psychology and People Management Research Ethics Committee (clearance no IPPM-2022-618[D]). All procedures were conducted in accordance with local legislation and institutional requirements. Ethical protocols were followed rigorously across all phases of data collection and analysis. Survey procedures adhered to the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines [56], with a detailed checklist provided in Multimedia Appendix 1. This study also analyzed publicly available Reddit comments, where users post with an expectation of broad visibility [57,58]. As the research focused solely on analyzing public text-based discourse (without directly engaging human participants), it was exempt from formal ethical review,

XSL•FO
RenderX

consistent with current regulatory interpretations [59,60]. However, following the principles of situated ethics, care was taken to protect participants from harm [26,61,62]. To minimize risks of reidentification, all usernames and personal identifiers were excluded, and findings are reported in aggregate form [58,63,64]. For the survey and interview components, all participants (N=23) provided their written informed consent via tick-box acknowledgment before accessing the questionnaire. The consent form detailed the study purpose, voluntary participation, anonymity protections, estimated time commitment, and the researcher's contact information. Participants were informed that they could withdraw at any time without penalty. All data collection and storage adhered to privacy-by-design principles. For survey and interview participants, no personally identifiable information (names, email addresses, IP addresses, geographic locations, or other identifiers) was collected at any stage. Survey responses were automatically anonymized through Google Forms' privacy settings, with restricted access limited to the research team. Interview notes were deidentified immediately upon completion, with no audio or video recordings made, to protect participant privacy and encourage candid discussion.

For Reddit and YouTube data, usernames and handles were systematically removed during data extraction and replaced with generic identifiers. All visual content from Reddit was recreated as schematic illustrations in Canva (Canva Pty Ltd) to prevent reverse image searching and protect user anonymity. No URLs, timestamps, or other metadata that could enable identification of original posts were retained. All data were stored on password-protected, encrypted devices, accessible only to the research team. No financial compensation, incentives, or rewards were offered to survey or interview participants. Participation was entirely voluntary and altruistic, and no employer or third-party organizations were involved in recruitment or data collection. No images in this paper contain identifiable individuals. All visual examples from Reddit were recreated as schematic illustrations in Canva, preserving thematic content while eliminating traceability to original users or posts. All narrative excerpts from online platforms were lightly paraphrased to prevent searchability while maintaining semantic meaning. This approach ensures that individual users cannot be identified despite the public nature of the original posts. Three recruitment posts were made by the first author on different days in July 2025 through his personal LinkedIn account (over 12,000 followers). The posts invited adult volunteers (aged ≥18 years) from any location to participate in an anonymous survey on men's mental health experiences. There was no paid promotion or employer involvement. Interested individuals accessed an information sheet and provided tick-box consent before proceeding to the survey. The survey form included the following statement:

"This survey is completely anonymous—we do not ask for your name, phone number, email, or any personally identifying information (such as social security or ID numbers). As a result, no one will be able to identify you from your responses, especially within a large sample."

At the end of the form, participants were invited to contact the researchers if they were willing to take part in a short follow-up interview about the topic. In total, 23 participants completed the survey, and 9 of them subsequently volunteered for interviews. To ensure participants felt comfortable speaking openly, no audio or video recordings were made. Instead, researchers took detailed written notes during each interview. At the start of every discussion, participants were reminded that no personal or identifying data would be shared with anyone and that their participation was voluntary.

Results

Overview

This section presents 4 key themes that emerged from the thematic analysis of Reddit posts, YouTube comments, and triangulated insights from survey and interview data. Collectively, these themes reveal how men use anonymous online spaces to express vulnerability, seek validation, cope through humor, and push back against societal norms of toxic positivity and stoicism. These digital communities offer men not only emotional refuge, but also opportunities to cocreate new narratives of masculinity grounded in honesty, empathy, and mutual support.

Before presenting thematic findings, we provide illustrative examples of the visual and narrative content that characterizes men's mental health discussions on Reddit and YouTube. These examples contextualize the tone, emotional depth, and cultural critique present in the analyzed discourse.

Figure 3 presents a widely shared African proverb addressing emotional neglect and its consequences, reflecting broader discussions about societal invalidation of men's emotional needs. Such culturally resonant imagery frequently generated extended discussion threads exploring men's experiences of being unseen or unsupported.

XSL•FO
RenderX

Figure 3. African proverb on male emotional suppression posted to r/MaleMentalHealth. Image recreated to protect user anonymity.

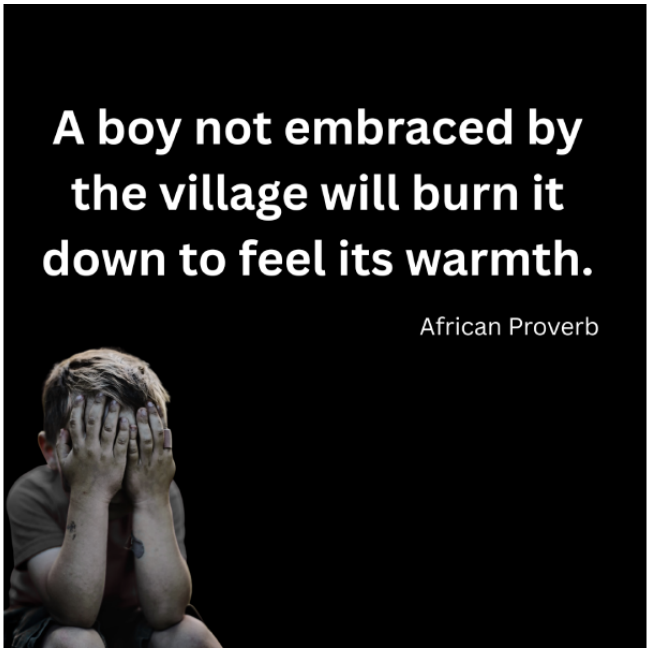


Figure 4 illustrates a recurring narrative about conditional male value, captured in a quote widely attributed to comedian Chris Rock. This perception, that men are valued only for what they provide rather than for their inherent worth, appeared repeatedly across platforms and aligned with survey participants' descriptions of feeling instrumentalized.

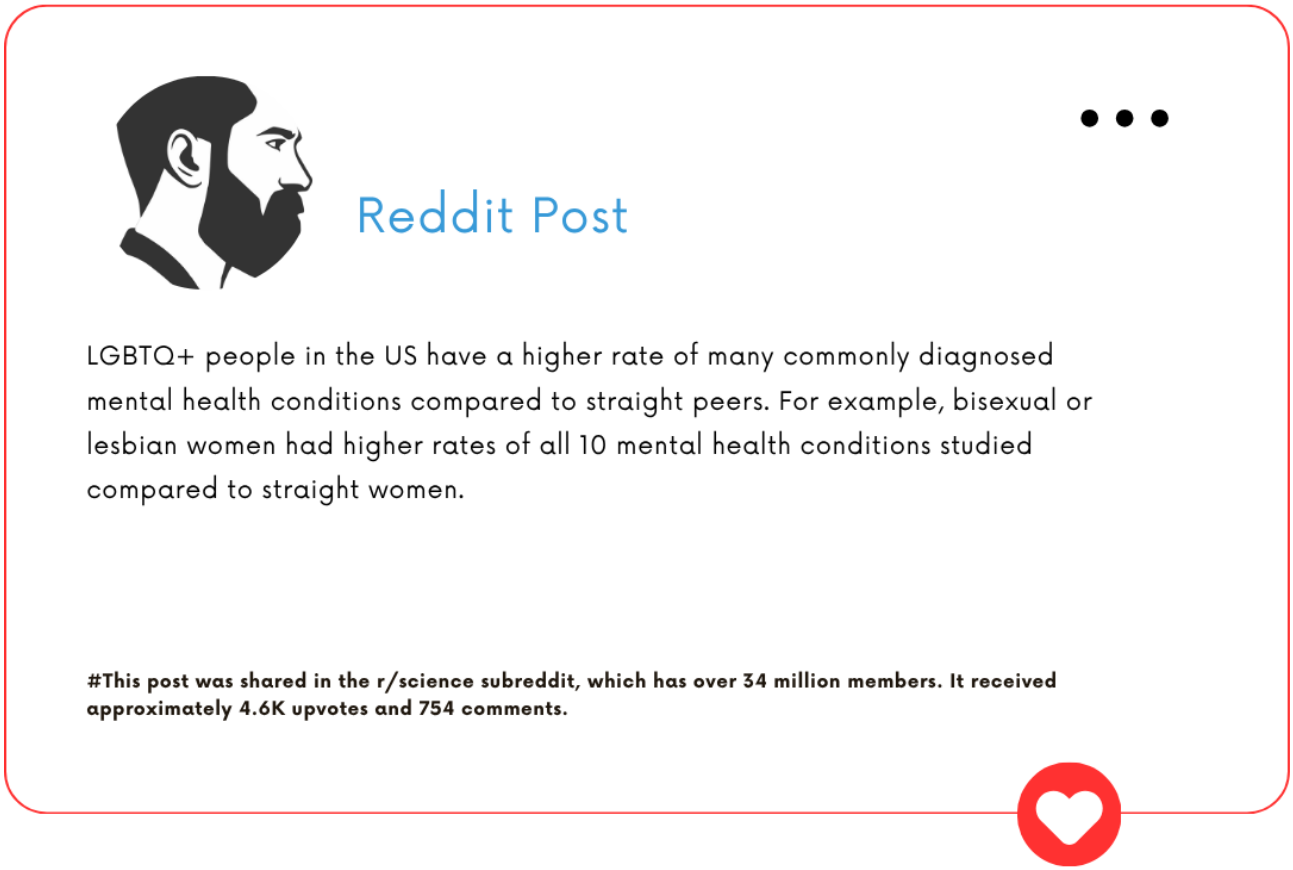
Figure 4. A quote attributed to comedian Chris Rock expressing cultural perception of conditional male value, widely circulated in online mental health discussions. Image recreated by authors.



Figure 5 demonstrates how online communities engage with empirical research on intersectional mental health issues, specifically LGBTQ+ mental health disparities. Posts sharing research findings often sparked discussions connecting minority stress theory to men's experiences of multiple marginalized identities.

XSL•FO
RenderX

Figure 5. Reddit post from r/Science discussing mental health disparities among individuals who identify as Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, and more (LGBTQ+). Image recreated to protect user anonymity.



Figures 6-7 exemplify the visual satire and personal testimony used to critique toxic positivity and mixed societal messages about male vulnerability. These formats allowed users to simultaneously express pain and critique cultural contradictions experienced when attempting emotional openness.

These representative examples illustrate the multimodal nature of digital mental health discourse, combining text, images, humor, and testimony to create spaces for emotional expression and social critique. The following sections present the 4 core themes that emerged from systematic analysis of this discourse.

As illustrated in Figure 2, the 4 themes were not derived from isolated data sources but through a process of systematic triangulation. Thematic patterns identified in Reddit and YouTube datasets were compared with survey and interview findings during Phase 3 of the study. This convergent approach enabled cross-validation across qualitative and quantitative strands: online expressions of emotion, humor, and peer support were examined for alignment or contrast with participants' self-reported experiences. Convergent patterns strengthened the credibility of themes, while divergent or contradictory accounts were treated as negative cases and discussed reflexively to refine thematic boundaries. Through this process, the analysis remained grounded across multiple data sources, ensuring that interpretations reflected a holistic view of men's mental-health discourse in digital environments.

XSL•FO
RenderX

Figure 6. A meme posted to r/MaleMentalHealth contrasts help-seeking messaging with social mockery of male vulnerability. Image recreated by authors.



Figure 7. YouTube comment critiquing toxic positivity through an account of emotional invalidation at a funeral. Image recreated by authors.



XSL•FO
RenderX

Theme 1: Normalizing Emotional Expression (Emotional Safety)

Participants across platforms are reclaiming emotional spaces where men can express vulnerability without judgment.

One of the most striking cross-platform insights, visible across Reddit, YouTube, and participant interviews, was the urgent need to normalize emotional expression among men. Digital spaces emerged as crucial venues where men could articulate their emotional struggles, often for the first time. These environments provided rare sanctuaries of psychological safety, where masculinity was not policed but redefined.

This theme stands in sharp contrast to hegemonic masculinity norms that discourage emotional vulnerability [17]. In many posts and comments, men directly challenged phrases such as "man up" or "be strong," revealing how such cultural expectations had silenced them throughout their lives.

YouTube comments further echo the emotional double bind men experience when navigating vulnerability. Many users highlighted the contradictory social expectations placed on men: "When a man opens up about his emotions, they say he should man up. When a man doesn't open up, they say it's 'toxic masculinity.'" This underscores the impossibility of conforming to prevailing norms without incurring judgment. Others reflected on direct personal invalidation, such as one man who shared, "I have a female friend and she called me weak for having depression and opening up my anxious thoughts." These narratives illustrate how emotional suppression becomes a survival strategy rather than a preference. As one commenter put it, "I only cry in private. Many women don't like sensitive men and see it as a sign of weakness." Another added with bitter irony, "Do you know who isn't allowed to talk about male suicide? Spoiler… Males." These voices reinforce the need for anonymous digital platforms where emotional honesty is possible, spaces where men can finally speak without fear of ridicule or erasure.

Another Reddit post (Figure 8) disrupted the popular myth that suicidal men show no visible signs before taking their lives. The user cited a study showing that 91% of middle-aged men who died by suicide had already sought professional help, arguing that cultural norms often ignore or downplay visible distress in men. They reframed this silence as a form of societal neglect, not individual failure.

Figure 8. Reddit post reframing male suicide as systemic neglect rather than individual silence. Image recreated by authors.



The emotional impact of such neglect was vividly captured in Multimedia Appendix 2, where a user described a partner's reaction to basic physical comfort: "My boyfriend nearly cried when I stroked his head. He told me no one ever did that for him. It broke my heart that something so simple could mean so much."

This idea, that even small gestures can hold emotional weight, was echoed in survey responses, where 20 of 23 men explicitly reported that they do not feel society provides a safe space to discuss mental health. This lack of openness was not abstract; it shaped lived experiences across age, profession, and background.

XSL•FO
RenderX

In one interview, an Indian senior manager in Human Resources recounted: "I didn't even know I was having a nervous breakdown. No one taught me in school, college, or even at work, and I work in HR, for God's sake! Thank God a friend insisted I see a doctor. You should've seen my face when the doctor told me it was a nervous breakdown. I was shell-shocked. I did not have anyone to share this with!"

YouTube comments reinforced this longing for validation. One widely liked post (6000+ likes) stated: "It's okay to feel strong negative emotions. It's okay to talk about it. This is the first time I don't feel broken. Reading these comments gives me hope."

Several men in the survey expressed a yearning for simple acts of care, such as a hug, a listening ear, or someone to validate their feelings, but noted that such support was rarely available to them. This emotional void was so pronounced that some respondents described turning to technology for comfort. One North American participant, a graduate-level man, remarked: "AI that validates your opinions and uses Socratic questioning to guide you to recovery." His comment highlights a growing trend: in the absence of safe interpersonal spaces, some men are turning to artificial intelligence tools in search of empathy, reflection, and emotional guidance. This underscores the urgency of creating more human-centered environments in which emotional expression is not only permitted but also actively nurtured.

Collectively, these sources demonstrate that men are investing in anonymous digital spaces, not for distraction but for connection and self-understanding. Platforms such as Reddit and YouTube are not merely hosting content; they are actively reshaping emotional norms through peer-to-peer empathy. What emerges is a collective rewriting of masculinity, one that prioritizes emotional literacy, mutual support, and the simple radical act of being heard.

Theoretical Interpretation

Theme 1 supports critiques of hegemonic masculinity, which frames emotional openness as weakness [21]. Platforms such as Reddit and YouTube serve as "digitally mediated sanctuaries," enabling men to express vulnerability without fear of judgment. These anonymous spaces challenge dominant norms by normalizing emotional need and fostering peer validation, echoing calls for healthier, psychologically supportive masculinities [35].

Theme 2: Mutual Validation and Peer Support (Belonging and Affirmation)

Anonymous online platforms serve as emotionally validating spaces where men's struggles are met with empathy, solidarity, and rare affirmations—often absent in offline life.

Anonymous digital communities such as Reddit and YouTube are emerging as vital ecosystems of emotional support, where men's disclosures are met with empathy, affirmation, and solidarity, which is often missing in their offline lives. In contrast to traditional environments such as workplaces, families, or male peer groups, where vulnerability may be mocked or dismissed, these online platforms provide men with a rare sense of being heard and validated.

Across Reddit threads, users frequently recounted how even the smallest compliments, often decades old, left lasting emotional imprints. One user wrote, "A girl telling me I have a nice smile when I was 14 still stands out as one of the nicest compliments I've ever received. I'm now 24." Another recalled, "I got a compliment from a coworker in 2010 about how I smelled really nice… I still buy the same deodorant and bath soap to this day and remember her compliment every time I use them." These memories reveal not only emotional scarcity but also the hunger for non–performance-based appreciation, a form of validation rarely offered to men in face-to-face settings.

This emotional deprivation was echoed in our primary data. A 28-year-old South African respondent shared in the survey, "I can't remember the last time anyone said something kind to me that wasn't about my work." Others spoke of longing for connection and affirmation that never materialized in their immediate circles. Of the 22 men surveyed, a majority answered "No" when asked if they had a safe space in society to talk about mental health. For many, online communities have become surrogate support systems, especially in the absence of affirming offline relationships.

Comments from the YouTube section of male mental health videos mirrored these concerns. One viewer reflected, "Recently, I told my mom how depressed I was every day. The first thing she asked was if I was still able to do my job properly… then she switched the subject to herself… not once acknowledging my depression." Another added, "I love how women are the ones advocating for men's rights, depression, suicide… because if men were doing it, we'd be labelled as sexist or misogynistic."

An employed African interviewee, a graduate, gave a poignant account of grief and peer dismissal: "Forget about peer validation and support from other men! The amount of toxic advice men get is unbelievable. In my circles, it's considered embarrassing if a young man is crying at the funeral of a loved one, rather than comforting the young one at his loss." This sense of isolation, minimization, and toxic resilience underscores why digital communities are so significant. They offer the validation that offline contexts frequently deny.

These dynamics are not only expressed in words but also visually represented through memes that resonate widely within these communities. As illustrated in Multimedia Appendix 3, one viral Reddit post noted: "The average guy is so starved for positive attention that a simple compliment is enough to get him interested in you." The meme attracted thousands of upvotes, reflecting its broad emotional resonance among men online.

In this environment, digital peer support becomes a mechanism of emotional reclamation. Common affirmations such as "We should complement each other more often" or "Everyone's life is precious" reflect a growing counter-narrative that values emotional expression, mutual care, and nontransactional affirmation. Rather than perform masculinity through stoicism or silence, men in these communities are rewriting the

XSL•FO
RenderX

script—collectively shaping new norms grounded in emotional openness.

As one participant explained, reflecting on the significance of online peer environments and artificial intelligence–based support systems: "What I get from social media is not therapy, but perspective. People here remind me that it's okay to feel, and that I'm not broken for wanting to be seen."

Interpretation in Theoretical Context

Theme 2 reflects challenges to hegemonic masculinity, which discourages emotional openness and mutual validation [21]. Reddit and YouTube function as "safe digital intimacy zones" where men exchange empathy and emotional support. This aligns with Sense of Community Theory [65], as these online spaces foster belonging, shared emotional connection, and nonstigmatizing expressions of vulnerability.

A small subset of posts framed humor as a form of avoidance rather than connection. These were treated as negative cases and used to refine the thematic boundaries. Accordingly, in Theme 3, we highlight how humor operates along a continuum, from emotional deflection to genuine bonding, depending on context and intent.

Theme 3: Coping Through Humor and Irony (Emotional Deflection and Connection)

Humor, irony, and self-deprecation emerge as culturally sanctioned tools that allow men to express psychological pain while preserving masculine norms of emotional restraint.

Across Reddit and YouTube, men often ridicule societal expectations such as "man up" and "don't be a pu**y," not to diminish their pain, but to reframe it in socially palatable ways. For instance, one Reddit user sarcastically asked, "Can we change 'Be a man' to mean it's okay to cry?" while another remarked, "I cried myself to sleep after a depression battle but refused to tell my girlfriend because I'd look weak." Such posts illuminate the tension between inner emotional needs and external masculine scripts. This emotional double-bind is often mediated through visual satire and dark memes, an aesthetic of shared struggle disguised as humor.

YouTube comments similarly reflect this mode of coping. One user noted, "Dead people receive way more flowers than the living because regret is stronger than gratitude," while another quipped, "If anything happens to a guy, it's just considered funny." These sharp observations lay bare the emotional neglect men experience, and the ironic laughter used to endure it.

This sentiment is reflected in interviews and survey responses as well. A South African participant stated, "I usually joke about my stress before I talk about it seriously… it makes it less awkward." Another shared, "Laughing at it feels safer than admitting how close I am to burning out." Humor, then, becomes a socially acceptable entry point for serious conversations that might otherwise be dismissed. A participant added in the interview: "It has been the best skill I have acquired. It helps me navigate social awkwardness and build relationships. It keeps me from shutting down." Others echoed this pattern of "laughing through the pain" in emotionally invalidating environments.

Visual Satire as Collective Commentary

This use of humor is also poignantly captured in popular memes. One widely upvoted post (Multimedia Appendix 4) shows a drowning figure labelled "Men's mental health" reaching for help, only to be slapped away by a hand labelled "Society" shouting "Be a man". The meme's caption simply reads: "Yikes. Accurate though."

Another meme (Multimedia Appendix 5) uses cartoon imagery to depict a man loading a revolver marked "Men having mental health issues" with bullets labelled "Be a man," a chilling metaphor of how societal messaging weaponizes masculinity against men themselves.

Interpretation in Theoretical Context

Theme 3 illustrates how humor acts as a culturally acceptable outlet for emotional pain under hegemonic masculinity, where vulnerability is often ridiculed [21]. In online spaces, humor becomes a tool of resistance, allowing men to share distress while preserving social acceptability. This aligns with the sense of community theory [65], as shared laughter fosters connection, emotional safety, and subtle defiance of rigid masculine norms.

Theme 4: Pushback Against Toxic Positivity and Societal Norms (Emotional Resistance)

This theme reflects how men are actively rejecting societal pressures that dismiss emotional pain and enforce performative strength, advocating instead for a more honest, compassionate masculinity.

This theme captures men's rejection of societal scripts that silence emotional expression and promote performative strength. Across platforms, participants challenge cultural expectations such as "man up" or "stay strong," which invalidate their distress and leave little room for vulnerability. In doing so, they advocate for a redefinition of masculinity rooted in emotional honesty.

A widely upvoted YouTube comment criticized the double standards faced by men attempting to open up: "Women: 'Men should talk about their feelings. It's okay to feel vulnerable.' Also, women: Sensitive men are needy and weak. I only date real men.'"

On Reddit, sarcastic memes and personal testimonies echoed this dissonance. One user posted an image highlighting that "Will Smith crying turned into a meme is proof that the mental health of men is not taken seriously" (Multimedia Appendix 6), while another shared a visual list reminding readers that "boys' feelings are valid too" (Multimedia Appendix 7). These posts reflect a broader cultural dismissal of male distress, often amplified by humor or invisibilization.

This rejection of emotional suppression was also voiced in interviews and open-ended survey responses. One South African respondent shared: "Yes. I now have a much more difficult time crying, even when alone, and I was forced to close off from my family, which resulted in me being kicked out of my home at 18." Another stated, "Therapy can't look at you when you're failing… men are never taught how to speak about their experiences in ways others can understand. It's a skill that has

XSL•FO
RenderX

to be learned." Several participants noted the internal conflict caused by having to appear strong while feeling shattered inside.

One Reddit post humorously critiqued performative masculinity: "I hate how, as a man, the solution to life's problems is inevitably 'lift weights.' Academic stress? Lift. Marital issues? Lift. Depression? Lift. At this point, it wouldn't surprise me if someone said lifting cures cancer." These ironic retorts offer a subversive form of resistance against toxic positivity.

Despite this pushback, the desire for compassion was clear. A simple yet powerful Reddit post read: "I don't want physical intimacy. I just want someone to hold me." Others shared artwork and messages such as "Men need to know it's okay to cry, to be vulnerable, to seek support" and "Your body is beautiful bro" (Multimedia Appendix 8), reinforcing the call for emotional affirmation rather than judgment or dismissal.

The cumulative effect of these expressions—memes, confessions, and affirmations—illustrates a growing resistance to outdated norms of stoicism. As captured in a Reddit quote, "Men are told to open up, and then to shut up. No wonder they don't feel safe talking." (Multimedia Appendix 9). These narratives point to a desire not only to speak but also to be heard, seen, and validated as emotionally complex human beings.

Interpretation in Theoretical Context

Theme 4 highlights resistance to hegemonic masculinity, where stoicism and strength are prioritized at the expense of emotional well-being [17]. By challenging tropes such as "man up," men expose the psychological cost of these norms. This resistance also aligns with inclusive masculinities theory [66], as men begin to legitimize emotional openness and redefine masculinity in more diverse and supportive ways.

Linguistic and Emotional Patterns in Online Discourse on Men's Mental Health

To complement the thematic findings, Python-based linguistic and emotion analyses were conducted on the Reddit and YouTube datasets. The word frequency analysis (refer to Multimedia Appendix 10) showed that frequently used terms—such as men, women, suicide, depression, and society—center discussions on gender, mental health struggles, and societal expectations.

The emotion analysis using the National Research Council Emotion lexicon (refer to Multimedia Appendix 11) revealed a mix of positive and negative emotions, with prominent expressions of sadness, fear, trust, and anger. This pattern suggests that while users often voice pain and vulnerability, the discourse also contains resilience, empathy, and occasional hope.

Together, these lexical and emotional trends underscore that online platforms function as spaces where men articulate distress and co-construct supportive narratives around mental health.

Discussion

Principal Findings

Four interrelated themes emerged: (1) normalizing emotional expression, (2) mutual validation and peer support, (3) coping through humor and irony, and (4) pushback against toxic positivity. Anonymous platforms functioned as "digitally mediated sanctuaries," enabling emotional disclosure with reduced social risk.

First, anonymous platforms such as Reddit and YouTube provided "safe digital intimacy zones" where men could express vulnerability without fear of judgment. This finding supports the notion of "anonymous authenticity," where the protective veil of anonymity fosters genuine emotional disclosure and enables men to renegotiate masculinity as open rather than stoic.

Second, mutual validation and peer support emerged as critical. Men consistently offered one another affirmation and understanding, which are emotional resources often unavailable offline. Online forums became surrogate support systems, echoing Sense of Community Theory [65], which emphasizes belonging, emotional connection, and fulfillment of needs. These peer-led spaces allowed men to construct alternative masculine identities grounded in openness and care.

Third, humor and irony served as culturally acceptable coping mechanisms. Men used dark jokes and memes to express distress while maintaining social cohesion. Rather than trivializing their experiences, shared humor acted as a collective pressure valve and reinforced emotional bonds within the community.

Fourth, participants explicitly pushed back against toxic positivity and rigid societal norms. They rejected platitudes such as "man up" or "just lift weights," critiquing the invalidation of men's emotional pain. These narratives marked a grassroots resistance to hegemonic masculinity through satire, storytelling, and emotional honesty.

Overall, our findings suggest that online communities enable men to reclaim their emotional voice and mitigate masculine stigma. These platforms serve both as safe spaces and informal "emotional classrooms," where men can develop relational skills, resist stigma, and foster more inclusive models of masculinity, transforming vulnerability into strength and connection.

Comparison With Prior Work

Our findings align with established literature on masculinity and mental health stigma, confirming the persistent influence of hegemonic norms that equate vulnerability with weakness [7,17]. Consistent with prior studies, participants described delaying help-seeking due to pressures to appear emotionally invulnerable [8,9]. This emotional suppression often led to shame, isolation, and maladaptive coping, reproducing a cycle noted in earlier research.

However, our study expands on this understanding by highlighting how digital peer spaces actively disrupt this cycle. While prior work has acknowledged men's preference for informal support [1], our analysis reveals how online forums, such as r/MaleMentalHealth posts and TEDx comment sections, serve as therapeutic spaces. These communities enable candid expression and validation, suggesting that masculine norms can be moderated by context. Wagner and Reifegerste [19], our observational data indicate men thriving in environments that legitimize emotional openness.

XSL•FO
RenderX

We also extend existing work on emotional deprivation. Past research has noted that men receive less social support [67], but our participants' deep emotional impact from small compliments reveals how acute that scarcity is. Similarly, our findings on humor build on Samson et al [68], showing how humor online acts not as avoidance but as a shared coping mechanism, fostering solidarity and facilitating deeper conversations.

Our theme on toxic positivity adds a gendered nuance to emerging critiques of this phenomenon [12,13]. Men in our study viewed forced cheerfulness as an extension of masculine pressure, invalidating pain under the guise of optimism. Unlike earlier pessimistic views, however, our data show men resisting this dynamic by affirming each other's right to struggle.

Theoretically, our work integrates hegemonic and inclusive masculinity perspectives [21,66], illustrating that traditional norms coexist with more emotionally expressive masculinities, especially among younger users. We also apply peer support theory [69,70] and introduce the idea of "collective authenticity," a community-based norm of truth-telling and support forged in anonymous online environments.

In summary, our study both confirms and extends prior scholarship. It demonstrates that while men remain constrained by cultural expectations, they are also influencing masculinity from the ground up through online peer support. This shift suggests future interventions may be more effective when they amplify organic, community-led approaches rather than solely relying on institutional efforts.

While the present findings highlight how online peer spaces can facilitate emotional openness and mutual validation, they should be interpreted as evidence of discursive potential rather than confirmed behavioral or cultural transformation. The study's cross-sectional and qualitative design does not allow for causal inference or longitudinal observation of change. Moreover, the men engaging in these online communities represent a self-selecting group already predisposed toward openness and help-seeking. As such, these findings illustrate how progressive masculinity discourses are articulated and negotiated in supportive digital environments, not how widespread they are across all male populations.

While online peer spaces can serve as supportive environments that validate men's emotional expression, they are not without risks. The same anonymity that enables openness may also facilitate misinformation, reinforce maladaptive norms, or create exclusionary dynamics that privilege certain masculine voices over others. For instance, echo-chamber effects can amplify oversimplified narratives about strength or self-reliance, while peer validation may occasionally normalize avoidance or emotional detachment. Recognizing these tensions situates our findings within a more nuanced understanding of digital peer support—one that acknowledges both its potential to foster norm flexibility and its susceptibility to reproducing aspects of hegemonic masculinity.

Implications for Practice

This study offers several actionable insights for improving mental health support for men. Recognize the Value of Online Peer Support

Anonymous online forums offer authentic spaces for emotional expression. Clinicians and organizations should consider these as legitimate supplementary supports and may guide hesitant clients to engage with them. Collaborating with community moderators can help disseminate trustworthy resources in user-friendly ways.

Encourage Positive Peer Modeling

The power of mutual validation suggests a need for peer-led support spaces. Workplace- or campus-based men's circles that promote compliments, empathy, and open sharing can counter emotional neglect [69,70]. Clinicians should also validate male clients' emotional disclosures, modeling empathy that many men rarely receive.

Address Toxic Positivity Directly

Campaigns must go beyond "it's okay to talk" and instead teach others how to respond constructively when men open up. Emotional first-aid training for teachers, coaches, and families can help shift from minimizing distress to holding space for it [71], normalizing sentiments such as "I'm here for you" over "stay strong."

Use Humor as a Gateway

Humor often serves as an emotional entry point. Mental health initiatives can adopt a lighter tone, through memes, jokes, or relatable content, to build rapport before addressing deeper issues. Validating this approach makes support more accessible and engaging.

Ultimately, interventions should be co-designed with men, building on grassroots peer support already thriving online. By bridging informal digital spaces and formal care systems, practitioners can help men feel equally safe opening up in therapy as they do on Reddit.

Strengths and Limitations

This study has several notable strengths. Chief among them is its triangulated, multimodal design, combining thousands of Reddit posts and YouTube comments with survey and 9 interview responses. This convergence allowed us to validate key themes across platforms and formats, enhancing credibility. The unusually large volume of qualitative data enabled both depth (narrative insight) and breadth (textual analysis), a rare strength in research on men's mental health. Our use of Braun and Clarke's [72] 6-phase thematic analysis provided methodological rigor, with themes grounded in participants' actual language. Reflexive practices, such as team discussions to check biases, further bolstered trustworthiness.

Another strength lies in our theoretical integration. By aligning findings with frameworks such as hegemonic masculinity, online disinhibition, and Sense of Community Theory, we situate insights within broader scholarship. The study also offers strong ecological validity, as it analyzes real-world online discourse, not lab responses. The inclusion of men from varied regions and age groups enhances the cross-cultural relevance of our results.

However, some limitations must be acknowledged. Our sample, comprising self-selected commenters and volunteers, may not

XSL•FO
RenderX

represent all men, especially those who remain silent or align strongly with traditional norms. The anonymity of online platforms also limits demographic certainty, and while we focused on upvoted posts and comments to reduce outlier bias, quieter dissenting views may be underrepresented. The interpretive nature of thematic analysis, while valuable for depth, introduces potential subjectivity; we addressed this through reflexive dialogue rather than intercoder reliability metrics, per Braun and Clarke's [52] guidelines.

A key limitation of the study lies in the demographic homogeneity of the qualitative sample. The 9 interview participants and 23 survey respondents were drawn mainly from professional and higher-education backgrounds, reflecting the reach of the LinkedIn recruitment channel. This limits the transferability of the findings to broader populations of men with different socioeconomic or cultural contexts. The qualitative insights should therefore be understood as illustrative of discourse patterns within this demographic rather than generalizable to all men. Nevertheless, triangulation with a much larger corpus of Reddit and YouTube data provided valuable contextual grounding that strengthened interpretive credibility.

We also lack data on mental health outcomes, that is, whether participating in these forums leads to tangible help-seeking or symptom relief. The cross-sectional design and focus on English-language platforms further limit generalizability, especially to non-Western or nondigital populations. Finally, while our focus was on supportive dynamics, potential risks such as misinformation or unqualified peer advice in online spaces remain underexplored.

Despite these limitations, this exploratory study provides a nuanced and empirically grounded picture of how men are negotiating emotional stigma and building community in digital spaces. It lays the groundwork for future, outcome-focused research in this domain.

Future Directions

Building on our findings, future research should explore the long-term impact of online mental health communities on men's help-seeking and emotional well-being. Longitudinal studies can assess whether engagement in forums such as Reddit leads to sustained improvements or increased uptake of professional support.

Further work should broaden demographic and cultural scope, especially focusing on underrepresented groups such as older men, adolescents, non-English speakers, and men from LGBTQ+ or racial minority. Understanding how different identities shape digital help-seeking could improve inclusivity in support strategies.

The themes of humor, anonymity, and peer validation offer strong foundations for developing gender-sensitive interventions. Future studies could test digital peer support tools, such as apps or moderated groups, that intentionally integrate these features, measuring their impact on distress, stigma, and connection.

There is also rich potential in exploring symbolic and narrative practices. Studies could examine how acts such as posting anonymously, wearing symbols, or sharing memes promote resilience. Interventions based on storytelling or expressive creativity may offer accessible therapeutic outlets for men.

Finally, attention must be paid to the downsides of online support. Future work should explore moderation strategies, misinformation risks, and how men distinguish helpful from harmful advice in digital spaces. Strengthening these forums without undermining openness is a key challenge.

A multidisciplinary approach, blending psychology, data science, public health, and gender studies, will be crucial to understanding and optimizing how digital communities can support men's mental health at scale.

Conclusions

This study explored how men navigate mental health challenges through anonymous digital platforms and personal narratives. The central finding is that men are increasingly expressing emotional vulnerability in online spaces where they feel safe and heard, challenging persistent stereotypes about masculine emotional suppression. Reddit, YouTube, and similar platforms have become vital outlets for peer support, particularly for men underserved by traditional mental health pathways.

Through humor, shared struggle, and candid dialogue, men are cocreating alternative narratives of masculinity that value empathy over stoicism and connection over silence. These organic support systems, while informal, often provide the understanding and validation men struggle to find in offline contexts. The analysis demonstrates that masculinity is not a fixed barrier but a dynamic set of norms that can shift through community and conversation.

This study makes 4 key contributions to the field. First, methodologically, it combines large-scale digital discourse analysis with qualitative triangulation across multiple platforms and data sources, providing both breadth and depth rarely achieved in men's mental health research. Second, theoretically, it extends inclusive masculinity theory into anonymous online contexts through empirical evidence, demonstrating how digital affordances enable masculine norm negotiation. Third, it introduces the concept of "digitally mediated sanctuaries" to describe online spaces where men practice vulnerability and mutual support with reduced social risk. Fourth, from an infodemiological perspective, it demonstrates how mental health information and peer support narratives circulate and gain legitimacy within male-dominated online communities, revealing patterns of information flow that differ from traditional health communication channels.

The real-world implications are significant for practice and policy. Mental health professionals can design gender-sensitive digital interventions that incorporate features men already use organically, such as humor, anonymity, and peer validation, rather than imposing clinical frameworks that may feel alienating. Public health communicators can develop campaigns that leverage the peer-to-peer dynamics and visual culture already thriving in these spaces. Organizations can recognize and support online peer networks as complementary systems that reduce help-seeking barriers and provide alternative

XSL•FO
RenderX

pathways to emotional support for men hesitant to access formal services.

In essence, this research offers a hopeful message. When given safe, authentic spaces, men not only speak but also listen, validate, and transform each other. The path forward lies in amplifying these grassroots movements and designing mental health strategies that reflect how men already connect, cope, and care. Rather than viewing digital peer support as a substitute for professional care, practitioners and policymakers should understand it as a valuable entry point that meets men where they are. By bridging informal online communities with formal support systems, we can create more inclusive and effective approaches to men's mental health that honor both the challenges men face and the creative solutions they are already building together.

Data Availability

The data that support the findings of this study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author upon reasonable request.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

AS conceptualized the study, conducted data collection from Reddit, YouTube, and qualitative surveys, and led the first round of coding and thematic analysis. MDS independently reviewed and refined the codes and themes, contributing to the analytical rigor and theoretical framing. Both authors contributed to the interpretation of findings and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CHERRIES (Checklist for Reporting Results of Internet E-Surveys) for the men's mental health online study.
[DOCX File , 22 KB - infodemiology_v6i1e81315_app1.docx]

Multimedia Appendix 2
Reddit post illustrating emotional deprivation and validation scarcity among men.
[PNG File , 122 KB - infodemiology_v6i1e81315_app2.png]

Multimedia Appendix 3
Meme from r/MaleMentalHealth depicting the impact of compliment scarcity on men.
[PNG File , 58 KB - infodemiology_v6i1e81315_app3.png]

Multimedia Appendix 4
Meme illustrating societal dismissal of men's mental health.
[PNG File , 519 KB - infodemiology_v6i1e81315_app4.png]

Multimedia Appendix 5
Cartoon depicting masculine norms as self-destructive.
[PNG File , 336 KB - infodemiology_v6i1e81315_app5.png]

Multimedia Appendix 6
Meme critiquing public mockery of male emotional expression.
[PNG File , 478 KB - infodemiology_v6i1e81315_app6.png]

Multimedia Appendix 7
Meme capturing male-specific emotional pain.
[PNG File , 174 KB - infodemiology_v6i1e81315_app7.png]

Multimedia Appendix 8
Meme depicting body shaming directed at men.
[PNG File , 169 KB - infodemiology_v6i1e81315_app8.png]

XSL•FO
RenderX

Multimedia Appendix 9
Post encapsulating contradictory social expectations around male vulnerability.
[PNG File , 50 KB - infodemiology_v6i1e81315_app9.png]

Multimedia Appendix 10
Top 30 most frequent words in Reddit posts and YouTube comments on men's mental health.
[PNG File , 40 KB - infodemiology_v6i1e81315_app10.png]

Multimedia Appendix 11
Emotion distribution in Reddit posts and YouTube comments based on the National Research Council lexicon, showing the frequency of emotion-associated words across 8 categories plus positive and negative valence.
[PNG File , 28 KB - infodemiology_v6i1e81315_app11.png]

References

1. Mental Health, Brain Health and Substance Use (MSD). World Mental Health Report: Transforming Mental Health for All. Geneva: World Health Organization; May 09, 2024.
2. Mental Health Gap Action Programme (mhGAP) guideline for mental, neurological and substance use disorders. World Health Organization. 2023. URL: https://www.who.int/publications/i/item/9789240084278 [accessed 2026-04-18]
3. Mental disorders. World Health Organization. 2025. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders [accessed 2025-02-23]
4. Mental health at work. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work [accessed 2025-02-08]
5. Depression and other common mental disorders. World Health Organization. 2017. URL: https://www.who.int/publications/i/item/depression-global-health-estimates [accessed 2026-04-18]
6. Moitra M, Owens S, Hailemariam M, Wilson KS, Mensa-Kwao A, Gonese G, et al. Global mental health: where we are and where we are going. Curr Psychiatry Rep 2023;25(7):301-311 [FREE Full text] [doi: 10.1007/s11920-023-01426-8] [Medline: 37256471]
7. Chatmon BN. Males and mental health stigma. Am J Mens Health 2020;14(4):1557988320949322 [FREE Full text] [doi: 10.1177/1557988320949322] [Medline: 32812501]
8. Rice S, Oliffe J, Seidler Z, Borschmann R, Pirkis J, Reavley N, et al. Gender norms and the mental health of boys and young men. The Lancet Public Health 2021;6(8):e541-e542. [doi: 10.1016/s2468-2667(21)00138-9]
9. Kim S, Yu S. Men's mental health and interventions tailored to masculinity: a scoping review. J Mens Health 2023;19(11):1-10. [doi: 10.22514/jomh.2023.111]
10. Parent MC, Hammer JH, Bradstreet TC, Schwartz EN, Jobe T. Men's mental health help-seeking behaviors: an intersectional analysis. Am J Mens Health 2018;12(1):64-73 [FREE Full text] [doi: 10.1177/1557988315625776] [Medline: 29226771]
11. Sheikh A, Payne-Cook C, Lisk S, Carter B, Brown JSL. Why do young men not seek help for affective mental health issues? A systematic review of perceived barriers and facilitators among adolescent boys and young men. Eur Child Adolesc Psychiatry 2025;34(2):565-583. [doi: 10.1007/s00787-024-02520-9] [Medline: 39004687]
12. Wyatt Z. The dark side of #positiveVibes: understanding toxic positivity in modern culture. Psychiatry Behav Health 2024;3(1):1-6 [FREE Full text] [doi: 10.33425/2833-5449.0016]
13. Salopek AH, Eastin MS. Toxic positivity intentions: an image management approach to upward social comparisonfalse self-presentation. J comput-mediat commun 2024;29(3):zmae003. [doi: 10.1093/jcmc/zmae003]
14. Feltner ME. Toxic positivity and perceptions of mental health. Senior Theses. 2023. URL: https://scholarcommons.sc.edu/senior_theses/607 [accessed 2026-03-25]
15. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003;129(5):674-697 [FREE Full text] [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
16. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
17. Connell R. Masculinities. New York City: University of California Press; 1995.
18. Herreen D, Rice S, Currier D, Schlichthorst M, Zajac I. Associations between conformity to masculine norms and depression: age effects from a population study of Australian men. BMC Psychol 2021;9(1):32 [FREE Full text] [doi: 10.1186/s40359-021-00533-6] [Medline: 33608063]
19. Wagner AJM, Reifegerste D. Real men don't talk? Relationships among depressiveness, loneliness, conformity to masculine norms, and male non-disclosure of mental distress. SSM - Mental Health 2024;5:100296. [doi: 10.1016/j.ssmmh.2024.100296]

XSL•FO
RenderX

20. Marshall P, Booth M, Coole M, Fothergill L, Glossop Z, Haines J, et al. Understanding the impacts of online mental health peer support forums: realist synthesis. JMIR Ment Health 2024;11:e55750 [FREE Full text] [doi: 10.2196/55750] [Medline: 38722680]
21. Connell RW, Messerschmidt JW. Hegemonic masculinity: rethinking the concept. Gender Soc 2005;19(6):829-859. [doi: 10.1177/0891243205278639]
22. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004;7(3):321-326. [doi: 10.1089/1094931041291295] [Medline: 15257832]
23. Joinson AN. Self‐disclosure in computer‐mediated communication: the role of self‐awareness and visual anonymity. Euro J Social Psych 2001;31(2):177-192. [doi: 10.1002/ejsp.36]
24. Walther JB. Theories of computer-mediated communication and interpersonal relations. In: Mark L K, John A D, editors. The Handbook of Interpersonal Communication. Thousand Oaks, California: The SAGE Handbook of Interpersonal Communication; 2011:443-479.
25. McKenna KYA, Bargh JA. Coming out in the age of the internet: identity "demarginalization" through virtual group participation. J Pers Soc Psychol 1998;75(3):681-694. [doi: 10.1037/0022-3514.75.3.681]
26. De Choudhury M, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. ICWSM 2014;8(1):71-80. [doi: 10.1609/icwsm.v8i1.14526]
27. Crenshaw KW. Race, reform, and retrenchment: transformation and legitimation in antidiscrimination law. Harvard Law Review 1988;101(7):1331. [doi: 10.2307/1341398]
28. Crenshaw KW. Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Droit Société 2021;108(2):465-487. [doi: 10.4324/9780429500480-5]
29. Li H, Liu X, Zheng Q, Zeng S, Luo X. Minority stress, social support and mental health among lesbian, gay, and bisexual college students in China: a moderated mediation analysis. BMC Psychiatry 2023;23(1):746 [FREE Full text] [doi: 10.1186/s12888-023-05202-z] [Medline: 37833656]
30. Yin J. Navigating stress and seeking support on Reddit: a user-centered study of online support-seeking behaviors. Master's thesis. Irvine: University of California; 2024. URL: https://escholarship.org/uc/item/83r1t48j [accessed 2025-07-15]
31. Srivastava A, Gupta T, Cerezo A, Lord SPG, Akhtar MS, Chakraborty T. Critical behavioral traits foster peer engagement in online mental health communities. PLoS One 2025;20(1):e0316906 [FREE Full text] [doi: 10.1371/journal.pone.0316906] [Medline: 39804944]
32. Lundeen LA, McCall JR, Bradshaw AS, LeBlanc EL, Humphrey L. Digital catharsis or harmful exposure? A thematic analysis of self-directed violence Reddit posts. Soc Media Soc 2024;10(3):20563051241263562. [doi: 10.1177/20563051241263562]
33. Grub MF. Reddit as a “Safe Space”: topic modeling of online mental health communities for depression and anxiety. Weizenbaum J Digit Soc 2025;5(3). [doi: 10.34669/WI.WJDS/5.3.3]
34. Globally, on average, 1 man dies by suicide every minute of every day. Movember. 2025. URL: https://ca.movember.com/about/mental-health [accessed 2025-07-14]
35. APA guidelines for psychological practice with boys and men. American Psychological Association (APA). 2018. URL: http://www.apa.org/about/policy/psychological-practice-boys-men-guidelines.pdf [accessed 2025-07-16]
36. Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. Am Psychol 2003;58(1):5-14. [doi: 10.1037/0003-066x.58.1.5] [Medline: 12674814]
37. Seidler ZE, Dawes AJ, Rice SM, Oliffe JL, Dhillon HM. The role of masculinity in men's help-seeking for depression: a systematic review. Clin Psychol Rev 2016;49:106-118. [doi: 10.1016/j.cpr.2016.09.002] [Medline: 27664823]
38. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
39. Duthie G, Reavley N, Wright J, Morgan A. The impact of media-based mental health campaigns on male help-seeking: a systematic review. Health Promot Int 2024;39(4):daae104. [doi: 10.1093/heapro/daae104] [Medline: 39224087]
40. Seidler ZE, Rice SM, Ogrodniczuk JS, Oliffe JL, Dhillon HM. Engaging men in psychological treatment: a scoping review. Am J Mens Health 2018;12(6):1882-1900 [FREE Full text] [doi: 10.1177/1557988318792157] [Medline: 30103643]
41. Fisher K, Seidler ZE, King K, Oliffe JL, Rice SM. Men's anxiety: a systematic review. J Affect Disord 2021;295:688-702. [doi: 10.1016/j.jad.2021.08.136] [Medline: 34517242]
42. Kiselica MS, Englar-Carlson M. Identifying, affirming, and building upon male strengths: the positive psychology/positive masculinity model of psychotherapy with boys and men. Psychotherapy (Chic) 2010;47(3):276-287. [doi: 10.1037/a0021159] [Medline: 22402085]
43. Opozda MJ, Galdas PM, Watkins DC, Smith JA. Intersecting identities, diverse masculinities, and collaborative development: considerations in creating online mental health interventions that work for men. Compr Psychiatry 2024;129:152443 [FREE Full text] [doi: 10.1016/j.comppsych.2023.152443] [Medline: 38113813]
44. Pan X, Hou Y, Wang Q. Are we braver in cyberspace? Social media anonymity enhances moral courage. Comput Hum Behav 2023;148:107880. [doi: 10.1016/j.chb.2023.107880]

XSL•FO
RenderX

45. Schlichthorst M, King K, Reifels L, Phelps A, Pirkis J. Using social media networks to engage men in conversations on masculinity and suicide: content analysis of Facebook campaign data. Soc Media Soc 2019;5(4):2056305119880019. [doi: 10.1177/2056305119880019]
46. Butler J. Gender Trouble: Feminism and the Subversion of Identity. Milton Park: Routledge; 1990.
47. Kozinets RV. Netnography: Redefined. Thousand Oaks, California: Sage; 2015.
48. Anderson E. Inclusive Masculinity. New York, NY: Routledge; 2009.
49. Borinca I, Gkinopoulos T. Challenging traditional masculinity: minority norms and the reimagining of male gender roles. Soc Influ 2025;20(1):2538627. [doi: 10.1080/15534510.2025.2538627]
50. Borinca I, Iacoviello V, Valsecchi G. Men's discomfort and anticipated sexual misclassification due to counter-stereotypical behaviors: the interplay between traditional masculinity norms and perceived men's femininization. Sex Roles 2021;85(3-4):128-141 [FREE Full text] [doi: 10.1007/s11199-020-01210-5] [Medline: 34776595]
51. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research. Thousand Oaks, California: SAGE; 2018.
52. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
53. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks California: SAGE Publications; 2022.
54. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2019;13(2):201-216. [doi: 10.1080/2159676x.2019.1704846]
55. Sim J, Saunders B, Waterfield J, Kingstone T. Can sample size in qualitative research be determined a priori? Int J Soc Res Methodol 2018;21(5):619-634. [doi: 10.1080/13645579.2018.1454643]
56. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
57. Amaya A, Bach R, Keusch F, Kreuter F. New data sources in social science research: things to know before working with Reddit data. Soc Sci Comput Rev 2019;39(5):943-960. [doi: 10.1177/0894439319893305]
58. Fiesler C, Zimmer M, Proferes N, Gilbert S, Jones N. Remember the human: a systematic review of ethical considerations in Reddit research. 2024 Presented at: Proceedings of the ACM on Human-Computer Interaction; january 2024; New York, NY p. 1-33. [doi: 10.1145/3633070]
59. Gliniecka M. The ethics of publicly available data research: a situated ethics framework for Reddit. Soc Media Soc 2023;9(3):20563051231192021. [doi: 10.1177/20563051231192021]
60. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying Reddit: a systematic overview of disciplines, approaches, methods, and ethics. Soc Media Soc 2021;7(2):205630512110190. [doi: 10.1177/20563051211019004]
61. boyd D, Crawford K. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf Commun Soc 2012;15(5):662-679. [doi: 10.1080/1369118X.2012.678878]
62. Markham A, Buchanan E. Recommendations from the AoIR Ethics Working Committee (Version 2.0). 2012. URL: https://pure.au.dk/ws/files/55543125/aoirethics2.pdf [accessed 2023-03-25]
63. Adams NN. 'Scraping' Reddit posts for academic research? Addressing some blurred lines of consent in growing internet-based research trend during the time of Covid-19. Int J Soc Res Methodol 2022;27(1):47-62. [doi: 10.1080/13645579.2022.2111816]
64. Reagle J. Disguising Reddit sources and the efficacy of ethical research. Ethics Inf Technol 2022;24(3):41. [doi: 10.1007/s10676-022-09663-w]
65. McMillan DW, Chavis DM. Sense of community: a definition and theory. J Community Psychol 1986;14(1):6-23. [doi: 10.1002/1520-6629(198601)14:1<6::aid-jcop2290140103>3.0.co;2-i]
66. Anderson E, McCormack M. Inclusive masculinity theory: overview, reflection and refinement. J Gend Stud 2016;27(5):547-561. [doi: 10.1080/09589236.2016.1245605]
67. Tifferet S. Gender differences in social support on social network sites: a meta-analysis. Cyberpsychol Behav Soc Netw 2020;23(4):199-209. [doi: 10.1089/cyber.2019.0516] [Medline: 32096662]
68. Samson AC, Glassco AL, Lee IA, Gross JJ. Humorous coping and serious reappraisal: short-term and longer-term effects. Eur J Psychol 2014;10(3):571-581. [doi: 10.5964/ejop.v10i3.730]
69. Oduwaye O, Kiraz A, Sorakin Y. A review of the literature on the influence of peer support programs on student mental health and well-being. Cyprus Turk J Psychiatry Psychol 2024;6(2):177-182. [doi: 10.35365/ctjpp.24.2.08]
70. Repper J, Carter T. A review of the literature on peer support in mental health services. J Ment Health 2011;20(4):392-411. [doi: 10.3109/09638237.2011.583947] [Medline: 21770786]
71. Geierstanger S, Yu J, Saphir M, Soleimanpour S. Youth mental health first aid training: impact on the ability to recognize and support youth needs. J Behav Health Serv Res 2024;51(4):588-598 [FREE Full text] [doi: 10.1007/s11414-024-09893-4] [Medline: 39090505]
72. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys

XSL•FO
RenderX

LGBTQ+: Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, and more

Edited by W Ahmed; submitted 25.Jul.2025; peer-reviewed by I Borinca, Z Haime; comments to author 11.Oct.2025; revised version received 15.Jan.2026; accepted 29.Jan.2026; published 05.May.2026.

Please cite as:
Shekhar A, Saurombe MD
Online Discussions of Men's Mental Health on Reddit and YouTube: Cross-Sectional Mixed-Methods Infodemiological Study
JMIR Infodemiology 2026;6:e81315
URL: https://infodemiology.jmir.org/2026/1/e81315
doi:10.2196/81315
PMID:

©Anurag Shekhar, Musawenkosi Donia Saurombe. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 05.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Measuring Safety Risks in Online Drug Sales: Empirical Study Based on Complex Network Theory

Rong Jiang[1], Prof Dr; Bingke Xing[2], MS; Lixin Shu[1], Prof Dr

[1]Faculty of Pharmacy, Naval Medical University, Shanghai, China

[2]School of International Pharmaceutical Business, China Pharmaceutical University, Nanjing, China

Corresponding Author:
Lixin Shu, Prof Dr
Faculty of Pharmacy
Naval Medical University
325 Guohe Road, Yangpu District
Shanghai, 200433
China
Phone: 86 021 81871320
Email: luky_shu@sina.com

Abstract

Background: With the rapid development of pharmaceutical e-commerce, online drug sales have led to additional safety risks, while providing convenient services to the public. These risks pose challenges to public health security and regulatory systems.

Objective: This study aimed to quantitatively measure the safety risks of online drug sales.

Methods: We collected data on cases of violation in online drug sales from government websites and established a complex network model consisting of 116 nodes and 152 edges for the safety risks in online drug sales based on complex network theory. We explored network topological properties through complex network characteristic indicators (eg, average path length, diameter, degree, and degree distribution). In addition, we followed the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method to identify key risk factors in online drug sales safety, calculate the relative nearness degree value of each risk node, and rank them accordingly.

Results: Data on 458 cases of violation were collected. The key risk nodes in the complex network model for safety risks in online drug sales were identified as the failure to (1) strictly perform prescription review and dispensing duties, (2) issue prescriptions through internet hospitals, (3) establish drug procurement and acceptance records, and (4) establish drug sales records, in addition to (5) consumers failing to provide prescriptions.

Conclusions: Based on the risk measurement results, regulatory authorities can formulate targeted risk prevention and control measures for key risks, such as failure to strictly perform prescription review and dispensing duties and consumers failing to provide prescriptions. By building a tighter and stronger regulatory network for online drug sales and improving regulatory efficiency, this approach could maintain order in the online drug sales market and enhance the overall level of public health security.

(JMIR Infodemiology 2026;6:e86876) doi:10.2196/86876

KEYWORDS

online drug sales; drug safety risks; risk measurement; complex network; complex network theory

Introduction

In recent years, with the rapid development of internet technology, the medical e-commerce industry has surged to prominence. The online drug sales market has entered a period of rapid development, emerging as a new business model within the pharmaceutical sector. Compared with traditional drug-purchasing methods, online drug sales in China allow consumers to purchase drugs online through various online shopping platforms, such as JD.com, Taobao, and Meituan, breaking the time and space constraints of traditional pharmacies [1]. In the first half of 2025, China's online pharmacy market recorded pharmaceutical sales exceeding 40 billion yuan (US $6 billion), representing a year-on-year increase of 20.36% [2].

However, compared to offline channels, the virtual nature, anonymity, and cross-regional reach of online sales make it a

XSL•FO
RenderX

high-risk channel for circulating counterfeit and substandard drugs and illegal sales worldwide [3,4]. China emphasizes the development of online drug sales, having successively revised and introduced a series of policies to strengthen risk management concepts and enhance systematic governance. In this study, safety risks associated with online drug sales referred to various potential factors or hazards that may lead to issues regarding drug quality and consumer drug safety during pharmaceutical transactions conducted via online platforms. In the literature on safety risks in online drug sales, researchers have used both qualitative and quantitative analytical methods to identify these risks. Ahmed et al [5] reported that online drug sales carry risks, such as drug abuse, inconsistent quality, misleading product claims, and difficulties in protecting consumer rights. Xu and Wang [6], Huang et al [7], and Xie [8] found that the safety risks associated with online drug sales in China include imperfect traceability mechanisms for online drug sales, inadequate logistics and distribution systems, nonstandardized prescription management, insufficient platform oversight, incomplete laws and regulations, and a lack of online pharmaceutical service systems. Some scholars have used relevant theories or models to study the safety risks associated with online drug sales, exploring the causes or effects of these risks. Liu [9] used risk management theory and found that the inadequate regulatory system, weak internal and external constraints within enterprises, and multiple game dynamics in the market environment are the key factors leading to safety risks in online drug sales. Hu et al [10] used failure mode and effects analysis to assess potential safety risks associated with purchasing prescription drugs from online pharmacies, with results indicating that the reuse of paper prescriptions poses the highest risk. Oriakhi et al [11] conducted a cross-sectional study of online pharmacies during the COVID-19 pandemic and found that 47% were “rogue” pharmacies; furthermore, most legitimate pharmacies relied on online consultations instead of prescription reviews, highlighting the serious patient safety risks posed by the lack of safety checks and inadequate regulation.

In recent years, quantitative analysis methods based on case texts have gradually been applied to the field of safety risk research. Zhang et al [12] established a risk evaluation index system based on bus fire accident cases, identifying the primary factors contributing to these incidents. Pathan et al [13] used analytical hierarchy and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) methods to assess flood risk in the city of Navsari, Gujarat, India, identifying vulnerable areas more susceptible to inundation during floods. Shen et al [14] collected investigation reports on construction safety accidents in rural self-built houses from 2014 to 2022 and established a Bayesian network model to elucidate the key causes of these accidents. These findings demonstrate that case-based quantitative analysis methods have matured, but their application in research on the safety risks in online drug sales remains unexplored.

For the past few years, complex network theory, by virtue of its strong advantages in quantitative and visual analysis, has been widely applied in safety analysis across various fields, such as transportation [15], construction [16], coal mining [17], and power systems [18]. Based on this theory, researchers primarily extract risks from case texts, establish correlations between risks, and measure risks using complex network statistical metrics. Wang et al [19] extracted relationships between hazards in Beijing subway accident cases and constructed a network diagram, identifying key hazards using degree, clustering coefficient, and betweenness centrality. However, network indicators, such as degree and betweenness centrality, only reflect risk characteristics from different dimensions and cannot comprehensively measure the importance of risks on their own. Therefore, some scholars use the TOPSIS method to synthesize multiple complex network statistical indicators into a single metric for assessing the extent to which risk affects security. The TOPSIS method is suitable for scenarios involving the ranking and selection of multiple indicators and evaluation objects. By calculating the proximity to the ideal solution, it enables the objective integration of multidimensional indicators, thereby avoiding the one-sidedness associated with a single evaluation dimension. Zhang et al [20] applied the entropy weight–TOPSIS method and evaluation metrics, including degree centrality, betweenness centrality, and PageRank value, to assess the importance of risk nodes in severe traffic accidents in Sichuan Province, China. Online drug sales involve multiple entities and processes, containing numerous safety risk factors that influence each other. Therefore, applying complex network theory to construct the relational structure among safety risk factors of online drug sales, and using the TOPSIS method to integrate multiple network indicators, can scientifically measure risk and has good applicability in research on safety risks in online drug sales.

Currently, China has established a rigorous regulatory framework for online drug sales centered on risk management. Regulatory authorities have invested substantial resources in inspections and enforcement, generating a vast and systematically rich body of publicly available administrative penalty cases. These cases provide a solid data foundation for conducting a reverse analysis of the safety risks in online drug sales models from a risk perspective. However, no existing research has systematically mined and quantitatively analyzed these cases. Therefore, this study aimed to fill this research gap. By collecting cases of violation, we innovatively applied complex network theory to construct a complex network model for safety risks in online drug sales and identified key risks using the TOPSIS method based on network statistical indicators, filling a research gap. The findings will provide evidence-based support for regulators to implement precise governance and optimize resource allocation, while also offering methodological references for risk research and practice in similar markets worldwide.

Methods

Data Collection

We collected cases of violation in online drug sales from January 1, 2014, to June 30, 2024. Data sources included the Peking University Law Database, China Judgments Online, the National Medical Products Administration (NMPA) website, provincial/municipal medical product administration websites, and administrative punishment announcements from market

XSL•FO
RenderX

supervision bureaus. The following search terms were used: “internet drug” (互联网药品) OR “online drug sales” (药品网络销售) OR “online pharmacy” (网上药店).

A total of 772 violation online drug sales cases were retrieved. After removing duplicates, 648 (83.94%) cases remained. Next, we conducted a detailed review of the cases and further excluded some based on the following exclusion criteria: (1) cases violation unrelated to online drug sales, such as online sales of medical devices, cosmetics, health supplements, or offline physical pharmacy operations; (2) cases where the information was too brief to identify a risk transmission chain; and/or (3) cases with extensive content but vague descriptions of illegal acts, using generalizations instead of process details, making it difficult to determine risks and their interrelationships. Finally, 458 (70.68%) cases of violation in online drug sales were obtained.

Two researchers jointly reviewed the facts, processes, and legal grounds for penalties in the 458 cases, identifying specific risk points for each case. Through discussion, similar items were consolidated and duplicates removed to obtain a unique set of risk factors. Subsequently, the two researchers independently grouped the risk factors based on semantic similarity. After resolving any disagreements through discussion, the factors were preliminarily categorized into several risk categories, and three experts were invited to review and confirm the classification results. Adjustments were made based on the experts’ feedback, resulting in a final risk list and risk categories.

Complex Network Theory

Complex network theory commonly uses network diagrams to study system structures and the interrelationships between structure and system function. The key to complex network modeling lies in defining nodes and edges [21]. Based on the fundamental knowledge of complex network theory and the characteristics of online drug sales, we made the following assumptions:

- Assumption 1: The nodes in the model represent risks, and the edges represent the interrelationships between the risks. In the safety risk network model of online drug sales, a node activation indicates that the corresponding risk is active, and this node’s activation may directly trigger violation incidents. However, in most cases, violations result from the intertwined and combined effects of multiple risks. Therefore, the interrelationships between risks can be abstracted as edges within the safety risk network of online drug sales.
- Assumption 2: The model is a directed network, with directions representing the flow of risk transmission. Violations in online drug sales are not caused by a single risk factor but result from the combined effects of multiple interconnected risks. These risks are not isolated but interact with one another. Based on the complete facts of the violations in the penalty document, the sequence of events and the logical causal relationships leading to the risks are identified, and the corresponding links between preceding and subsequent risks are elucidated. Risks with clear triggering or evolutionary relationships are represented as directed edges in a complex network to illustrate the transmission relationships between risks. The risk transmission chain clearly illustrates the progression of cases of violation and precisely reveals the interconnections and chain reactions among risks. Therefore, we simplified violations in online drug sales into a directed network to represent the transmission relationships.

Based on these two assumptions, risk transmission chain data are converted into an adjacency matrix. Using Gephi software, the number of times connections occur between risks is used as weighting to form a complex network composed of multiple risk transmission chains. Analyzing the network’s indicator parameters can reveal its structural characteristics and evolutionary patterns. Common indicators include average path length, diameter, degree, betweenness centrality, eigenvector centrality, and clustering coefficient [22].

Average Path Length and Diameter

The average path length refers to the mean value of all the shortest paths between any two risk nodes, reflecting the strength of mutual influence between factors [23]. It serves as a crucial metric for assessing risk propagation efficiency and connectivity. The network diameter denotes the maximum distance between any two nodes in the network, where the distance between directly connected nodes is 1.

Degree

Degree is one of the key topological properties in complex networks, representing the number of connections a node has with other nodes and reflecting the node’s direct effect on the network [24]. A higher degree indicates that the node has more extensive direct connections with other nodes in the network, potentially making it a critical node. For directed networks, degree centrality includes in-degree, out-degree, and total degree.

Betweenness Centrality

In complex networks, the betweenness centrality measures a node’s connectivity by quantifying the ratio of the number of shortest paths passing through it to the total number of shortest paths in the network [25]. A higher betweenness centrality indicates that a significant number of shortest paths traverse that node. Consequently, the risks represented by the node play a more pivotal role as intermediaries within the risk network, substantially affecting risk propagation across the entire network.

Eigenvector Centrality

Eigenvector centrality is a metric for measuring node importance, considering not only the number of nodes directly connected to it but also the effect of its adjacent nodes [26]. The higher the eigenvector centrality, the more connections a node has with core nodes, indicating a stronger radiating effect and influence within the network.

Clustering Coefficient

The clustering coefficient measures the clustering properties of nodes within a network [27]. It is the ratio of the actual number of edges connecting a node to its neighbors to the maximum possible number of edges, with values ranging from 0 to 1. This

XSL•FO
RenderX

metric intuitively reflects the closeness of connections surrounding a node, indicating whether its neighboring nodes are interconnected.

Risk Measurement

Different metrics in complex networks analyze the effect of nodes from various perspectives. Relying solely on a single metric makes it difficult to fully measure the impact of nodes within a network [28]. To integrate multiple attribute metrics of a single node and thereby assess the effect of individual risks on the final outcome, we used the TOPSIS method to calculate the risk nearness degree C_i. The TOPSIS method is a classical multiattribute decision-making approach that comprehensively considers multiple attribute indicators, avoiding the one-sidedness of single-criterion decision-making. It is widely applied in multiobjective decision-making [29,30] and risk measurement studies [31,32]. The steps in calculation are detailed in Multimedia Appendix 1.

The fundamental approach of the TOPSIS method for risk measurement is as follows: Each risk node within the complex network model for safety risks in online drug sales is treated as a decision scheme in TOPSIS. Four metrics (degree, betweenness centrality, eigenvector centrality, and clustering coefficient) represent the four attributes of each decision scheme. A decision matrix based on the attribute values of each indicator was construct, the distance between each alternative and the ideal solution was measured, and their nearness degree was calculated. The nodes were ranked according to their nearness degree, where a higher degree indicated greater proximity to the positive ideal solution and a stronger influence.

Ethical Considerations

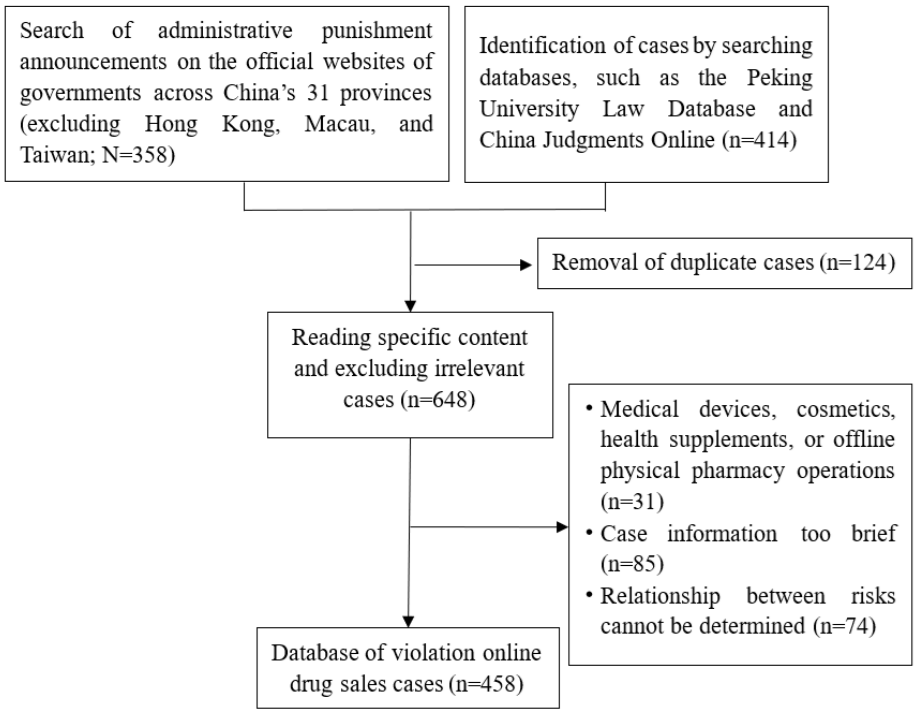
This study did not involve any research on human or animal participants.

Results

Basic Information About Cases of Violation

The case-screening process is shown in Figure 1. We collected a total of 458 cases of violation in online drug sales (Table 1). Over time, the overall number of cases showed an upward trend, accelerating significantly after 2018. This indicates that as internet penetration and e-commerce platforms develop, the scale of online drug sales continues to expand, and illegal incidents gradually increase. Cases of violation primarily occurred in economically developed or densely populated regions, such as Zhejiang Province, Guangdong Province, Fujian Province, and Beijing Municipality. These areas feature high internet penetration rates and substantial pharmaceutical market demand, making them potential hotspots for violations. The entities that commit violations include not only pharmaceutical trading enterprises (eg, retail pharmacies and wholesale drug distributors) and related individuals (eg, licensed pharmacists) but also enterprises that do not originally engage directly in drug trading, including technology companies, trading companies, and food stores. This underscores the need for enhanced oversight, not only of pharmaceutical businesses and individuals, but also of diverse entities operating in the field of online drug sales.

Figure 1. Flowchart for selecting cases of violation related to online drug sales.



XSL•FO
RenderX

Table 1. Basic information about cases of violation in online drug sales (N=458).

Item	Cases, n (%)
Penalty year	
First half of 2024	62 (13.54)
2023	157 (34.28)
2022	76 (16.59)
2021	58 (12.66)
2020	42 (9.17)
2019	26 (5.68)
2018	13 (2.84)
2017	5 (1.09)
2016	1 (0.22)
2015	3 (0.66)
2014	4 (0.87)
Not reported	11 (2.40)
Region	
Zhejiang Province	103 (22.49)
Guangdong Province	63 (13.76)
Fujian Province	38 (8.30)
Beijing Municipality	27 (5.90)
Tianjin Municipality	26 (5.68)
Sichuan Province	21 (4.59)
Other	180 (39.30)
Illegal subject	
Pharmaceutical trading company	267 (58.30)
Individual	67 (14.63)
Technology company	27 (5.90)
Trading company	25 (5.46)
Food stores	14 (3.06)
Other	58 (12.66)

Network Construction Process

Identify Risks

The 458 cases of violation represented actual security incidents that occurred or potential security risks that existed, ultimately leading to the imposition of penalties. Drawing upon these cases of violation as the source material, we identified potential risk points within the cases, extracted 116 risk factors, and categorized them into 8 aspects: business qualifications, 16 (13.79%) risks; drug procurement, 14 (12.07%) risks; transportation and delivery, 9 (7.76%) risks; information display, 11 (9.48%) risks; advertising and promotion, 17 (14.65%) risks; prescription management, 19 (16.38%) risks; pharmaceutical services, 3 (2.59%) risks; and drug sales, 27 (23.28%) risks. Further details are provided in Multimedia Appendix 2.

Identify Risk Transmission Chains

Compared to the traditional drug sales model, online sales involve multiple entities, including e-commerce platforms, online pharmacies, internet hospitals, consumers, and regulatory authorities, as well as multiple operational stages, such as business licensing, information display, prescription verification, drug delivery, and pharmaceutical services. These factors result in complex coupled transmission relationships among risks. Furthermore, when collating cases of violation in online drug sales, violations sometimes result not from a single risk but also from the coupling or derivation of multiple risks under certain conditions. Safety risks online drug sales possess a certain dynamic propagation characteristic.

We extracted and analyzed information from cases of violation in online drug sales, using risks as nodes and arrows to represent the transformation and evolution relationships between risks, thereby forming a chain of safety risk transmission in online

XSL•FO
RenderX

drug sales. A total of 509 risk transmission chains were constructed. Taking the case of a tobacco and liquor retailer in Ningbo City as an example, Textbox 1 illustrates the process of extracting the risk transmission chain from facts about the violation.

Textbox 1. Example of extracting the safety risk transmission chain in online drug sales.

Case source:

Administrative Penalty Information Public Disclosure, Zhejiang Province

Facts about the violation:

Investigation revealed that in July 2021, the involved party, Ningbo *** Tobacco & Alcohol Store, purchased one box (100 pieces) of Yunnan Baiyao Plasters (drug approval number: GYZZ Z20073016) at a price of 17 yuan/box (US $2.51/box), or 0.17 yuan/piece (US $0.02/piece) from the Ningbo Jiangdong Modern Department Store Market but did not retain the purchase invoice. The involved party could not recall the specific stall number. The involved party commenced business on August 5, 2021, primarily engaged in the sales of prepackaged food, daily necessities, and tobacco products, with an operating area of ~80 m^2. Until the case was identified, they had sold a total of 70 pieces of Yunnan Baiyao Plasters at a price of 0.5 yuan/piece (US $0.07/piece). Furthermore, in September 2021, they established a Tmall Shop online store on the Ele.me platform, primarily for retail general merchandise and the sale of Yunnan Baiyao Plasters. In October 2021, they began listing and selling Yunnan Baiyao Plasters on Tmall Shop. The product page contained the text “Yunnan Baiyao Plaster 1 piece ¥0.5”. When the text was clicked, the page displayed content and images including “Yunnan Baiyao Plaster OTC [Hemostasis/ Analgesia/ Anti-inflammatory/ Wound Healing/ Lightweight & Breathable],” “Add to Cart,” “¥20 minimum for delivery,” “7-day no-reason returns supported”, etc. Until the case was identified, the online store had sold a total of 15 pieces of Yunnan Baiyao Plasters at a price of 0.5 yuan/piece (US $0.07/piece).

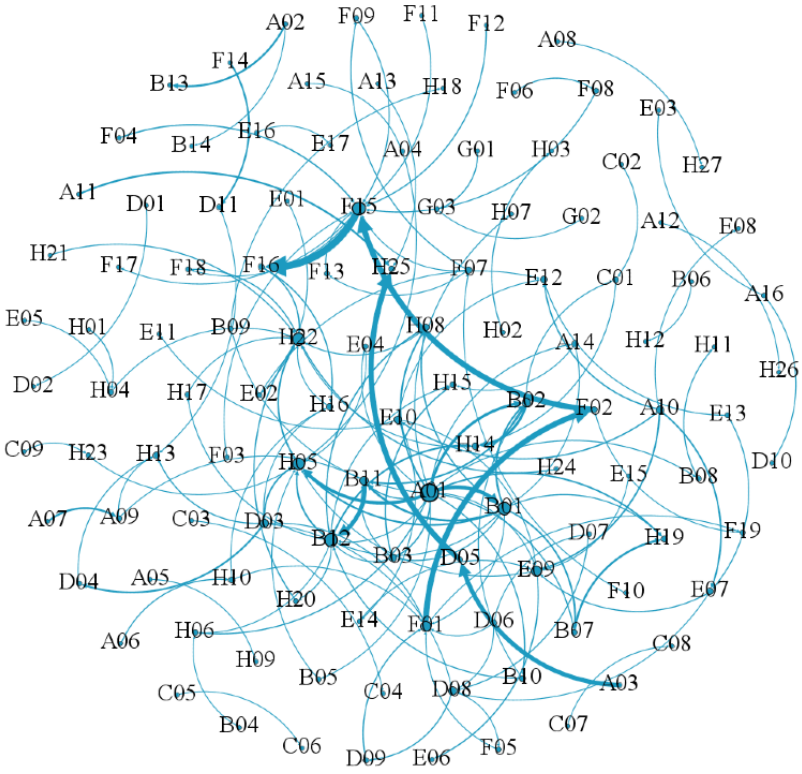
Risk transmission chain:

Failure to obtain a drug business license (A01) → procuring drugs from unqualified individuals or enterprises (B01) → inability to provide drug procurement vouchers (B11) → failure to establish drug procurement and acceptance records (B12) → illegal sale of over-the-counter (OTC) drugs (H08)

Construct Complex Network Models

Based on the aforementioned assumptions and risk transmission chains, we used the social network analysis software Gephi 0.10.1 to construct a complex network model for safety risks in online drug sales. The network consisted of 116 nodes and 152 directed weighted edges. Figure 2 shows the visualization result.

Figure 2. Complex network model for safety risks in online drug sales.



XSL•FO
RenderX

Analysis of Network Topology Characteristics

Average Path Length and Diameter

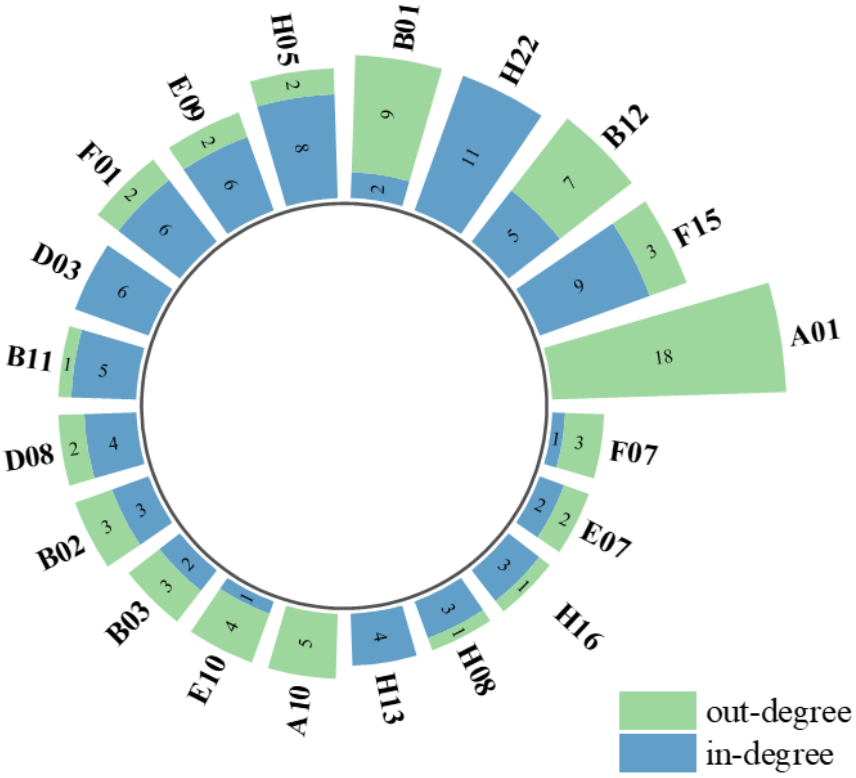
The average path length of the complex network model for safety risks in online drug sales was 2.453, indicating that it takes 2-3 steps to establish a connection between two risk points (relatively small average separation between risk nodes). This implies that in the network model, if a risk occurs at one node, it can be transmitted to other nodes over a short distance, quickly forming a chain reaction that impacts the safety of the entire online drug sales system.

The network diameter of the network model was 8, indicating that it takes a maximum of 8 steps for a risk node to affect another risk node within the network. Therefore, a risk transmission chain can be established within 0-8 steps.

Degree and Degree Distribution

Figure 3 lists the top 20 risk nodes, arranging all nodes in descending order of degree value. The average degree of the complex network model was 2.621, indicating that each node in this network model is connected to an average of 2.621 other nodes. This means that if one risk changes, it can, on average, affect 2-3 other related risks.

Figure 3. Top 20 risk nodes by degree in the complex network model of safety risks in online drug sales.

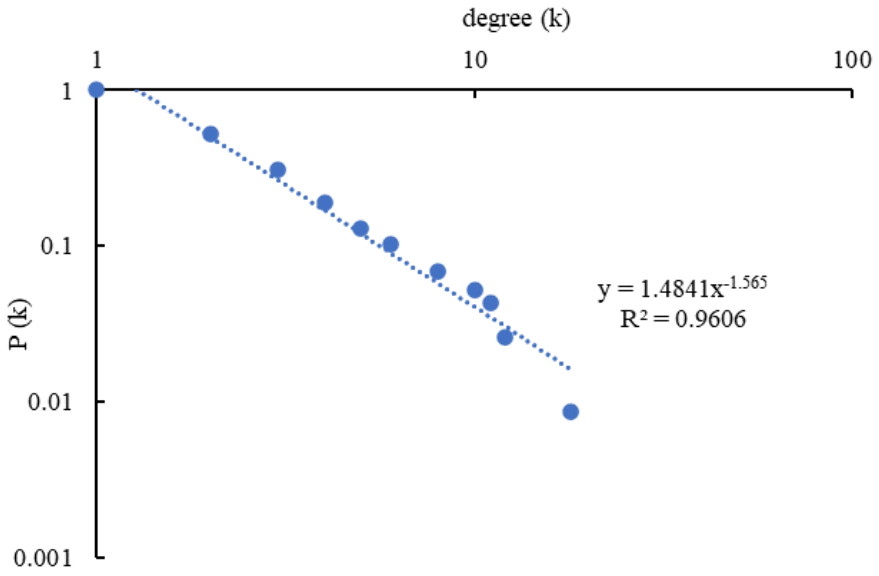


Regarding out-degrees, A01 (failure to obtain a drug business license) had the highest out-degree value of 18. This indicates the risk of failing to obtain a drug business license can propagate to 18 risk nodes within the network, with multiple risk evolution pathways. Regarding in-degrees, H22 (failure to establish drug sales records) had the highest value of 11. This indicates that failure to establish drug sales records can evolve from 11 other risks, making it relatively susceptible to influence from other risks.

The cumulative degree distribution of nodes in the complex network model is shown in Figure 4. The cumulative degree followed a power-law distribution, with the fitting result being $P(k)= \sim 1.4841k^{-1.565}$ and excellent fitting quality ($R^2=0.9606$). This indicates that the complex network model established in this paper exhibits scale-free properties, with most nodes having low degrees and only a small number possessing high degrees. Risks in networks tend to concentrate around core nodes, and the scale-free characteristic endows the network with robustness against random failures and vulnerability to targeted attacks [33]. Therefore, targeted attacks should be launched against high-risk nodes with large degrees within the network to disrupt their risk transmission pathways. This will increase the diameter and average path length of the complex network model, thereby slowing the evolution and spread of risks.

XSL•FO
RenderX

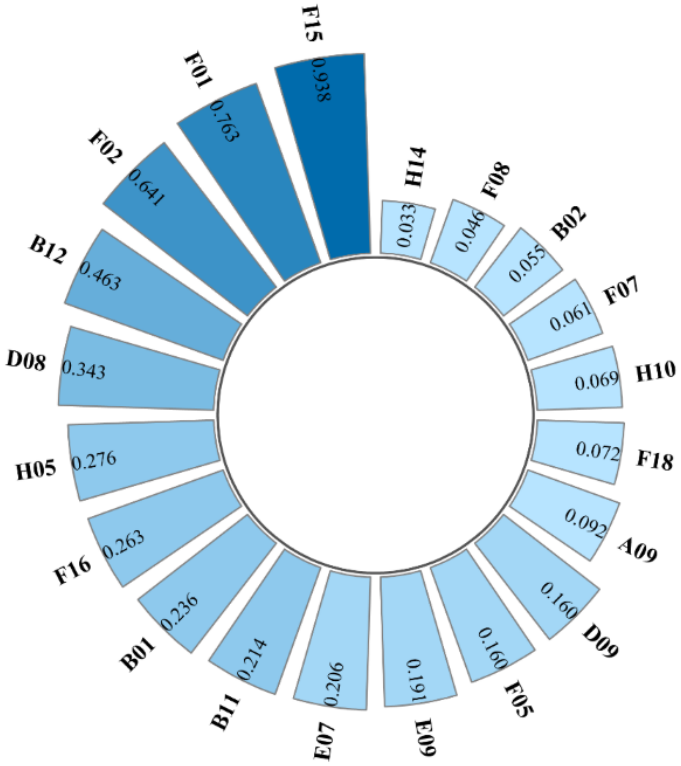
Figure 4. Cumulative degree distribution of nodes in the complex network for safety risks in online drug sales.



Betweenness Centrality

Figure 5 lists the top 20 risk nodes, ranking all risk nodes in descending order of betweenness centrality. F15 (failure to strictly perform prescription review and dispensing duties) exhibited the highest betweenness centrality at 0.938, followed by F01 (consumers failing to provide prescriptions) and F02 (failure to issue prescriptions through internet hospitals), with betweenness centralities of 0.763 and 0.641 respectively. This indicates that within the complex network model, these three nodes lie along the shortest paths connecting numerous other node pairs, serving as critical bridges in risk transmission. For instance, F15 (failure to strictly perform prescription review and dispensing duties) acts as a pivotal node linking risks such as a consumer uploading a blank prescription form and the sale of prescription drugs without prescriptions. A risk represented by a node with high betweenness centrality often triggers further risks along transmission pathways. Therefore, establishing specialized risk management plans for these critical nodes can effectively slow the rate of risk evolution and propagation, mitigating and preventing the chain reaction of multiple risk events.

Figure 5. Top 20 risk nodes by betweenness centrality in the complex network model for safety risks in online drug sales.



XSL•FO
RenderX

Figure 6 lists the top 20 risk nodes, ranking all nodes in descending order of eigenvector centrality. H22 (failure to establish drug sales records) exhibited the highest eigenvector centrality at 1.00, indicating that failure to establish drug sales records is linked to multiple critical risk nodes. Among the risk nodes with elevated eigenvector centrality, various drug sales–related risk nodes were encompassed, such as H05 (illegal drug sales), H14 (selling imported drugs without drug approval documents), and H23 (drug sales records are not authentic or complete). This implies that the definition of eigenvector centrality aligns with our practical understanding of safety in online drug sales. Specifically, most risks in the sales segments are typically linked to key risk nodes in the complex network model. The combined influence of these risk nodes increases the probability of sales risks occurring.

Figure 6. Top 20 risk nodes by eigenvector centrality in the complex network model for safety risks in online drug sales.



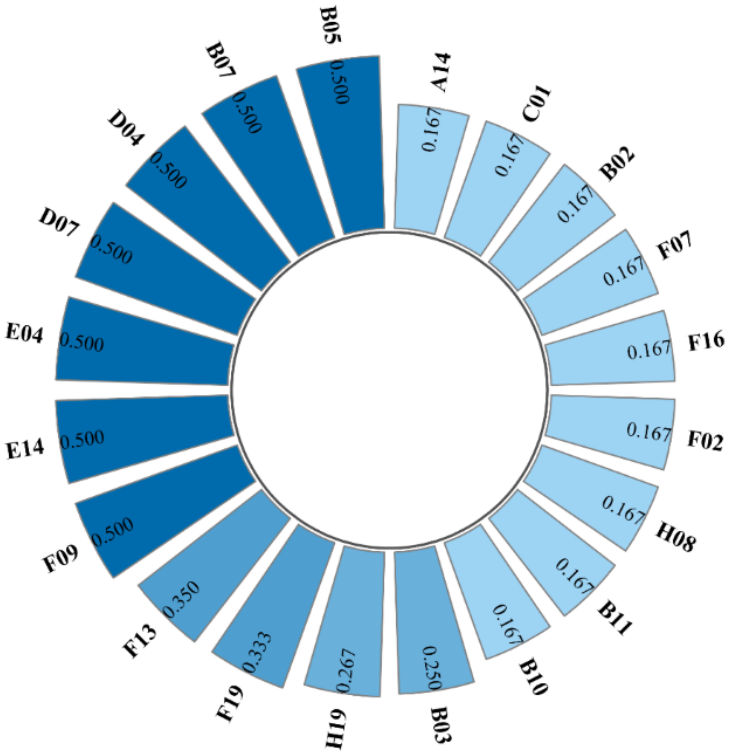
Clustering Coefficient

Figure 7 lists the top 20 risk nodes, ranking all nodes in descending order of clustering coefficient. Ten risks nodes, including B05 (using others' information to buy drugs from various hospitals), B07 (procuring counterfeit drugs), and D04 (failing to display qualification information of pharmacists or other pharmaceutical technical personnel) among others, exhibited the highest clustering coefficient of 0.50. This indicates a clustering effect among adjacent risk nodes for these risks. For example, when someone uses another person's information to purchase drugs at hospitals across different regions, the strong correlation between adjacent risks in the network, such as A01 (failure to obtain a drug business license) and H05 (illegal drug sales), may trigger a chain reaction within a localized area.

The average clustering coefficient of the complex network model was calculated to be 0.077. Using Gephi 0.10.1 software to generate a random network with the same number of nodes and edges yielded an average clustering coefficient of 0.008. Therefore, compared to a random network of the same scale, the complex network model exhibits a higher average clustering coefficient. This indicates the presence of strong local clustering structures within the network, where risks tend to concentrate more readily in specific segments.

XSL•FO
RenderX

Figure 7. Top 20 risk nodes by clustering coefficient in the complex network model for safety risks in online drug sales.



Comprehensive Ranking of Network Risks

The importance of risk nodes in the complex network for safety risks in online drug sales is influenced by multiple factors. Different indicators in complex network theory analyze the effect of nodes from different perspectives. Relying solely on a single indicator makes it difficult to fully measure the importance of nodes in the network [28]. Therefore, we adopted the TOPSIS method, effectively integrating four indicators from the complex network model (degree, betweenness centrality, eigenvector centrality, and clustering coefficient) to scientifically and reasonably measure the importance of each risk node in the network. Based on the TOPSIS calculation results, the risk nodes were ranked in descending order of C ; the top 15 nodes are shown in Table 2, while the complete ranking of all nodes is provided in Multimedia Appendix 3.

Table 2. Ranking of the top 15 risk nodes by C_i value for nodes in the complex network model for safety risks in online drug sales.

Node	C_i in descending order
F15	0.5408
F01	0.4638
F02	0.4419
B12	0.3956
H22	0.3872
F19	0.3561
H19	0.3547
B07	0.3518
E14	0.3477
F09	0.3477
F13	0.3477
B05	0.3471
D04	0.3471
D07	0.3471
E04	0.3471

XSL•FO
RenderX

In this paper, a higher relative nearness degree value for a node indicated that the node was more important in the complex network model for safety risks in online drug sales and represented a key node in the risk network. As shown in Table 2, the top-ranked risks in the complex network model included F15 (failure to strictly perform prescription review and dispensing duties), F01 (consumers failing to provide prescriptions), and F02 (failure to issue prescriptions through internet hospitals), with relative nearness degrees of 0.5408, 0.4638, and 0.4419, respectively. These risk nodes occupied critical positions within the network, and once manifested, they rapidly propagated to other risk nodes in the network. Among these, F15 (failure to strictly perform prescription review and dispensing duties) exhibited the highest relative nearness degree at 0.5408, making it the most critical risk factor in the complex network for safety risks in online drug sales that requires prioritized prevention and control.

Discussion

Principal Findings

With the rapid development of internet technology, significant changes have occurred in public drug-purchasing patterns. Online drug sales have gradually become a primary channel for the public to purchase drugs due to their convenience and efficiency [34]. However, this has also led to a series of issues and risks, such as the sale of prescription drugs without prescriptions and discrepancies between drug information displayed online and the actual products. Our study introduced complex network theory into the research on safety risks in online drug sales by constructing a network model to analyze risk propagation patterns and using the TOPSIS method to measure risks. The results showed that the key risk nodes in the model are F15 (failure to strictly perform prescription review and dispensing duties), F01 (consumers failing to provide prescriptions), F02 (failure to issue prescriptions through internet hospitals), B12 (failure to establish drug procurement and acceptance records), and H22 (failure to establish drug sales records).

We collected a total of 458 cases of violation. Based on the facts of violations and combined with legal and regulatory requirements, potential risk points in these cases were identified. Ultimately, 116 risks were extracted from the cases and categorized into 8 aspects: business qualifications, drug procurement, transportation and delivery, information display, advertising and promotion, prescription management, pharmaceutical services, and drug sales. Based on complex network theory, a complex network model for safety risks in online drug sales was established, comprising 116 nodes and 152 directed weighted edges.

Analysis of the topological properties of the complex network model revealed that A01 (failure to obtain a drug business license) has the highest total degree, with the maximum risk transmission or reception relationships. Although purchasing medicines from legal online pharmacies has been accepted in higher-income countries, the market has always been plagued by illegal suppliers, posing a significant risk to patient safety [35]. Burrell [36] analyzed the operational structure of illegal online pharmacies from the perspective of cross-border cybercrime, pointing out that it exhibits an adaptive strategy and high profit and low risk characteristics and is deeply intertwined with public health safety risks. Meanwhile, the network's cumulative degree follows a power-law distribution, exhibiting scale-free characteristics. F15 (failure to strictly perform prescription review and dispensing duties) exhibits the highest betweenness centrality, serving as a crucial bridge in risk transmission. Consistent with our findings, Oriakhi et al [11] investigated the operation of online pharmacies and found that among the 116 online pharmacies evaluated, 47% were illegal, often selling prescription drugs without prescriptions or consultations. H22 (failure to establish drug sales records) has the highest eigenvector centrality, connecting to multiple important risk nodes in the network. Ten risk nodes, including B05 (using others' information to buy drugs from various hospitals), B07 (procuring counterfeit drugs), and D04 (failure to display qualification information of pharmacists or other pharmaceutical technical personnel), have the highest clustering coefficients, indicating a clustering effect among adjacent risk nodes of these risks.

The TOPSIS method results indicate that in the complex network model, the top-ranked risks include F15 (failure to strictly perform prescription review and dispensing duties), F01 (consumers failing to provide prescriptions), F02 (failure to issue prescriptions through internet hospitals), B12 (failure to establish drug procurement and acceptance records), and H22 (failure to establish drug sales records). These risk nodes occupy crucial positions within the network and can rapidly propagate to other risk nodes. A study conducted in Saudi Arabia on the online purchase of antibiotics found that before selling antibiotics, up to 33.3% of online pharmacies do not require customers to provide a prescription [37]. The study further pointed out that 45% of these websites are fake or only sell local antibiotics [37]. Ahmed et al [38] conducted a systematic review on the issue of substandard and counterfeit drugs sold online, identifying challenges, such as the lack of knowledge and awareness among consumers and doctors, as well as the use of internet platforms and social media for direct-to-consumer advertising. Furthermore, Xiao [39] conducted a systematic analysis of the regulatory policies for online pharmacies in China and reported that the revision of the Drug Administration Law in 2019 was the first national-level legal framework for regulating online pharmacies in China. However, as artificial intelligence (AI)–driven online pharmacies are rapidly emerging, issues such as the responsibility boundaries of online platforms and the legal status of AI doctors and pharmacists still face legal and ethical controversies. These new business models will further exacerbate the complexity of safety risks in online drug sales. Digital health interventions, training programs for medical service providers, and public awareness campaigns tailored to specific situations are key approaches to bridging the gap between cognition and practice [40]. To mitigate the aforementioned risks, regulatory authorities should popularize the knowledge of prescription drugs to the public through various channels and media, including definitions, online purchasing procedures, and usage risks, thereby guiding the public to voluntarily purchase prescription drugs with prescriptions. Online pharmacies engaged in prescription drug

XSL•FO
RenderX

sales should regularly organize licensed pharmacists to participate in relevant training or continuing education courses to enhance their sense of responsibility and professional competence. Furthermore, standardized and unified electronic prescription review processes should be established to ensure the professionalism and compliance of online prescription reviews. In addition, regulatory authorities should conduct regular inspections of online drug operating enterprises, focusing on whether they have established complete and accurate drug procurement, acceptance, and sales records, as required, and whether these records enable drug traceability.

Strengths and Limitations

This study quantitatively measured risks in online drug sales and identified key risks, providing scientific guidance for online drug sales regulation and helping enhance the risk management capabilities of regulatory authorities. Despite these contributions, this study has some limitations.

First, because the extraction of risk nodes was performed manually, coding bias may exist in the process of translating the contents of cases of violation into risk nodes due to subjective judgment. To mitigate this potential bias, the following measures were taken: two researchers independently coded the cases, followed by cross-checking and discussion; three experts were invited to review and validate the risk classification results; and final adjustments were made based on expert feedback. Nevertheless, completely eliminating coding bias remains difficult. Future research could use automated text analysis methods for cross-validation to further reduce coding bias.

In addition, the online drug sales process involves numerous uncertain risks, and risk evolution patterns are relatively complex. Future research should continuously supplement violation case data and use multiple methods to deeply explore the patterns of risk propagation and key risk nodes in online drug sales, such as infectious disease models and node simulation attacks.

Conclusion

As an emerging drug procurement method, online drug sales provide significant convenience to consumers, while also involving numerous risks. Based on the collected cases of violation, this study used complex network theory to construct a complex network model for safety risks in online drug sales. Using network indicators and the TOPSIS method, we measured the safety risks in online drug sales and identified key risks. The results showed that the key risk nodes in the complex network model include F15 (failure to strictly perform prescription review and dispensing duties), F01 (consumers failing to provide prescriptions), F02 (failure to issue prescriptions through internet hospitals), B12 (failure to establish drug procurement and acceptance records), and H22 (failure to establish drug sales records). Therefore, targeted risk prevention and control measures should be developed for these key risks to promote the healthy and orderly development of the online drug sales industry.

Funding

This study was supported with grants from the National Natural Science Foundation of China (grant number 72304280) and the Key Support Project of the Jiangsu Provincial Social Science Applied Research Excellence Program (grant number 24SYA-031).

Data Availability

The datasets generated and analyzed during the study are available from the corresponding author upon reasonable request.

Authors' Contributions

RJ and LXS determined the research topic and design. RJ and BKX performed the database search, case selection, and model construction. RJ, BKX, and LXS participated in data analysis and interpretation. RJ was responsible for drafting and revising the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Steps in calculation used in the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method.
[DOCX File , 19 KB - infodemiology_v6i1e86876_app1.docx]

Multimedia Appendix 2
Safety risks in online drug sales.
[DOCX File , 33 KB - infodemiology_v6i1e86876_app2.docx]

Multimedia Appendix 3
Ranking of C_i values for all nodes in the complex network model for safety risks in online drug sales.
[DOCX File , 18 KB - infodemiology_v6i1e86876_app3.docx]

XSL•FO
RenderX

References

1. Liu L. Research on Regulatory Issues in Online Pharmaceutical Sales. Hainan University 2020:1-48. [doi: 10.27073/d.cnki.ghadu.2020.000728]
2. Online pharmacy drug sales exceeded 40 billion yuan in the first half of 2025, marking a 20.36% year-on-year increase. Commentary on the New Healthcare Reform. Xin Yi Gai Ping. 2025 Oct. URL: http://www.xygpl.com/contents/3/19507.html [accessed 2026-01-17]
3. Sng YJ, Kwok D, Goh E, Tan A, Teo J, Chan CL. The problems with online health product sales: how can regulations be improved? Drug Saf 2024 Jun 14;47(6):529-533 [FREE Full text] [doi: 10.1007/s40264-024-01414-8] [Medline: 38483769]
4. Miller R, Wafula F, Onoka CA, Saligram P, Musiega A, Ogira D, et al. When technology precedes regulation: the challenges and opportunities of e-pharmacy in low-income and middle-income countries. BMJ Glob Health 2021 May;6(5):e005405 [FREE Full text] [doi: 10.1136/bmjgh-2021-005405] [Medline: 34016578]
5. Ahmed Shaikh Z, Buzdar MH, Pahore MR, Hussain Mirani I, Noor M. The emergent business of internet pharmacies: convenience, risks, regulatory policies and future. J Pharm Res Int 2019 Nov 15;31(5):1-6 [FREE Full text] [doi: 10.9734/jpri/2019/v31i530311]
6. Xu Y, Wang W. Tips for mitigating risks in online drug sales. Pharm Econ J 2020 Dec 24;2:1-2 [FREE Full text] [doi: 10.38275/n.cnki.nyyjj.2020.002187]
7. Huang B, Wu S, Zhang Z. Risks and supervision countermeasures of online drug sales in China. China Pharm 2022;31(10):18-21. [doi: 10.3969/j.issn.1006 - 4931.2022.10.004]
8. Xie C. Existing risks of online drug sales in China and regulatory countermeasures. China Health Stand Manage 2025;16(05):44-47. [doi: 10.3969/j.issn.1674-9316.2025.05.011]
9. Liu L. Risk identification, assessment, and response for prescription drugs online sales. Health Econ Res 2022;39(04):19-23. [doi: 10.14055/j.cnki.33-1056/f.2022.04.005]
10. Hu Q, Hu H, Hu M, Zhang J, Gou L, Shi S, et al. Use of failure mode and effect analysis to reduce patient safety risks in purchasing prescription drugs from online pharmacies in China. Front Med (Lausanne) 2022 Jul 19;9:913214 [FREE Full text] [doi: 10.3389/fmed.2022.913214] [Medline: 35928298]
11. Oriakhi OJ, Almomani H, Patel N, Donyai P. The characteristics and operations of "online pharmacies" investigated in relation to medicines popularised during the coronavirus pandemic: a cross-sectional study. Front Pharmacol 2024 Feb 20;15:1346604 [FREE Full text] [doi: 10.3389/fphar.2024.1346604] [Medline: 38444937]
12. Zhang D, Yang H, Lv W. Research on risk evaluation indicator system for bus fire accidents. Saf Environ Eng 2019;26(06):137-142. [doi: 10.13578/j.cnki.issn.1671-1556.2019.06.020]
13. Pathan AI, Girish Agnihotri P, Said S, Patel D. AHP and TOPSIS based flood risk assessment- a case study of the Navsari City, Gujarat, India. Environ Monit Assess 2022 Jun 17;194(7):509 [FREE Full text] [doi: 10.1007/s10661-022-10111-x] [Medline: 35713716]
14. Shen J, Wang S, Zhang X. Analysis of causes and risk factors for safety hazards in rural self-built houses based on case mining. J Eng Manage 2024;38(03):71-75. [doi: 10.13991/j.cnki.jem.2024.03.013]
15. Deng J, Liu S, Shu Y, Hu Y, Xie C, Zeng X. Risk evolution and prevention and control strategies of maritime accidents in China's coastal areas based on complex network models. Ocean Coastal Manage 2023 Apr;237:106527. [doi: 10.1016/j.ocecoaman.2023.106527]
16. Tang B, Guo S, Li J, Lu W. Exploring the risk transmission characteristics among unsafe behaviors within urban railway construction accidents. J Civ Eng Manage 2022;28(6):443-456 [FREE Full text] [doi: 10.3846/jcem.2022.16924]
17. Cheng L, Guo A, Liu L. Risk measurement of coal mine gas explosion based on topological network algorithm. J Xi'an Univ Sci Technol 2022;42(02):268-275. [doi: 10.13800/j.cnki.xakjdxxb.2022.0210]
18. Amani AM, Jalili M. Power grids as complex networks: resilience and reliability analysis. IEEE Access 2021;9:119010-119031 [FREE Full text] [doi: 10.1109/access.2021.3107492]
19. Wang Z, Bhamra RS, Wang M, Xie H, Yang L. Critical hazards identification and prevention of cascading escalator accidents at metro rail transit stations. Int J Environ Res Public Health 2020 May 13;17(10):3400 [FREE Full text] [doi: 10.3390/ijerph17103400] [Medline: 32414127]
20. Zhang S, Xiong H, Li X, Yuan H, Zhang D, Liu Q. Causation analysis of serious road traffic accidents in Sichuan Province of China based on a hybrid HFACS-CN model. Traffic Inj Prev 2025 Nov 12:1-10 [FREE Full text] [doi: 10.1080/15389588.2025.2575049] [Medline: 41223434]
21. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proc Natl Acad Sci U S A 2004 Mar 16;101(11):3747-3752 [FREE Full text] [doi: 10.1073/pnas.0400087101] [Medline: 15007165]
22. Guo S, Zhou X, Tang B, Gong P. Exploring the behavioral risk chains of accidents using complex network theory in the construction industry. Physica A 2020 Dec;560:125012 [FREE Full text] [doi: 10.1016/j.physa.2020.125012]
23. Latora V, Marchiori M. Efficient behavior of small-world networks. Phys Rev Lett 2001 Oct 17;87(19):198701 [FREE Full text] [doi: 10.1103/physrevlett.87.198701]
24. Feng C, Gou H, Yang X. A vulnerability analysis of express network based on complex network theory. J Transp Eng Inf 2020;18(1):9-15. [doi: 10.3969/j.issn.1672-4747.2020.01.002]

XSL•FO
RenderX

25. Freeman LC. A set of measures of centrality based on betweenness. Sociometry 1977 Mar;40(1):35-41 [FREE Full text] [doi: 10.2307/3033543]
26. Wang Y. Research on the identification of safety risks and its evolution law of metro shield construction. Lanzhou Univ Technol 2023:1-89 [FREE Full text] [doi: 10.27206/d.cnki.ggsgu.2023.000982]
27. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998 Jun 04;393(6684):440-442 [FREE Full text] [doi: 10.1038/30918] [Medline: 9623998]
28. Sun J, Li M, Xiu X. Causal analysis of water traffic accidents based on complex network theory. J Dalian Maritime Univ 2023;49(02):80-90. [doi: 10.16411/j.cnki.issn1006-7736.2023.02.009]
29. Qu G, Zhang Z, Qu W, Xu Z. Green supplier selection based on green practices evaluated using fuzzy approaches of TOPSIS and ELECTRE with a case study in a Chinese internet company. Int J Environ Res Public Health 2020 May 07;17(9):3268 [FREE Full text] [doi: 10.3390/ijerph17093268] [Medline: 32392876]
30. Cao J, Xu F. Entropy-based fuzzy TOPSIS method for investment decision optimization of large-scale projects. Comput Intell Neurosci 2022 Jul 18;2022:4381293 [FREE Full text] [doi: 10.1155/2022/4381293] [Medline: 35898784]
31. Deng F, Li Y, Lin H, Miao J, Liang X. A BWM-TOPSIS hazardous waste inventory safety risk evaluation. Int J Environ Res Public Health 2020 Aug 10;17(16):5765 [FREE Full text] [doi: 10.3390/ijerph17165765] [Medline: 32784982]
32. Dogan B, Oturakci M, Dagsuyu C. Action selection in risk assessment with fuzzy Fine-Kinney-based AHP-TOPSIS approach: a case study in gas plant. Environ Sci Pollut Res Int 2022 Sep 02;29(44):66222-66234 [FREE Full text] [doi: 10.1007/s11356-022-20498-2] [Medline: 35499730]
33. Lynn C, Holmes C, Palmer S. Emergent scale-free networks. PNAS Nexus 2024 Jul;3(7):236 [FREE Full text] [doi: 10.1093/pnasnexus/pgae236] [Medline: 38966012]
34. Fu L. How to establish a regulatory mechanism for online pharmaceutical sales. China Qual Supervision 2023;11:60-62.
35. Fittler A, Abanmy NO, Serefko A, Rehman IU, Vida RG. Editorial: Internet pharmacies and the online pharmacy market: trends, perspectives and challenges. Front Pharmacol 2024 Sep 25;15:1489396 [FREE Full text] [doi: 10.3389/fphar.2024.1489396] [Medline: 39386023]
36. Burrell DN. Counterfeit pharmaceuticals and illegal online pharmacies as cyber-enabled organized crime and a dark public health risk. RAIS J Social Sci 2026;10(1):240-260. [doi: 10.5281/zenodo.19661131]
37. Abanmy NO, Alshehri SA, Alsudais RA. Antibiotic purchasing through online pharmacies: a mystery shopper study. J Taibah Univ Med Sci 2025 Oct;20(5):687-695 [FREE Full text] [doi: 10.1016/j.jtumed.2025.09.002] [Medline: 41140707]
38. Ahmed J, Modica de Mohac L, Mackey TK, Raimi-Abraham BT. A critical review on the availability of substandard and falsified medicines online: incidence, challenges and perspectives. J Med Access 2022 Feb 15;6:23992026221074548 [FREE Full text] [doi: 10.1177/23992026221074548] [Medline: 36204527]
39. Xiao P. China's online pharmacy: its regulatory policy and a future research agenda for the emerging AI age. Med Law Rev 2025 Jul 14;33(3):fwaf024. [doi: 10.1093/medlaw/fwaf024] [Medline: 40674723]
40. Pal E, Merchant L, Koffi AK, Mehta R, Rusatira JC, Kubiak L, et al. Assessing knowledge, attitudes, and practices and demand-side interventions for combating substandard and falsified medicines: a scoping review. J Pharm Policy Pract 2025 Sep 12;18(1):2550369 [FREE Full text] [doi: 10.1080/20523211.2025.2550369] [Medline: 40958951]

Abbreviations

TOPSIS: Technique for Order Preference by Similarity to Ideal Solution

Edited by T Mackey; submitted 31.Oct.2025; peer-reviewed by Z Wan; comments to author 02.Jan.2026; revised version received 18.May.2026; accepted 22.Jun.2026; published 07.Jul.2026.

Please cite as:
Jiang R, Xing B, Shu L
Measuring Safety Risks in Online Drug Sales: Empirical Study Based on Complex Network Theory
JMIR Infodemiology 2026;6:e86876
URL: https://infodemiology.jmir.org/2026/1/e86876
doi:10.2196/86876
PMID:42412541

©Rong Jiang, Bingke Xing, Lixin Shu. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 07.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a

XSL•FO
RenderX

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

XSL•FO
RenderX

Original Paper

Comparative Content Analysis and Expert Physician Risk Assessment of Aesthetic Procedures Promoted on TikTok in Taiwan: Quantitative Study

Cheng-Jung Wu[1,2,3,4], MD, PhD; Sheng-Yu Wu[5], MS; Cheng-Yu Tsai[6,7,8], PhD; Arnab Majumdar[6], PhD; Jeffrey Yang[9], MS; Jinn-Moon Yang[10,11], PhD; Lok-Yee Joyce Li[12,13], MD

[1]Department of Otolaryngology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[2]Department of Otolaryngology, Shuang Ho Hospital, Taipei Medical University, New Taipei, Taiwan
[3]College of Biomedical Science and Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
[4]Allergy and Rhinology, Quality HealthCare Center, Taoyuan, Taiwan
[5]Department of Industrial Engineering and Industrial Management, National Tsing Hua University, HsinChu, Taiwan
[6]Department of Civil and Environmental Engineering, London, United Kingdom
[7]School of Respiratory Therapy, College of Medicine, Taipei Medical University, Taipei, Taiwan
[8]TMU Research Center of Artificial Intelligence in Medicine and Health, Taipei, Taiwan
[9]Department of Biology, University of California, Riverside, CA, United States
[10]Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, HisnChu, Taiwan
[11]Department of Biological Science and Technology, National Chiao Tung University, HsinChu, Taiwan
[12]Department of Environmental and Occupational Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan
[13]School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Lok-Yee Joyce Li, MD
Department of Environmental and Occupational Medicine
National Taiwan University Hospital Hsin-Chu Branch
No 25, Lane 442, Section 1, Jingguo Road
Hsinchu, 300
Taiwan
Phone: 1 886 3 532 6151
Email: b101102173@tmu.edu.tw

Abstract

Background: Algorithm-driven social media platforms such as TikTok are increasingly creating stratified layers in the cosmetic medical market, thereby influencing patient decisions and safety. In Taiwan, TikTok has 2 parallel markets: one is a formal tier promoting “cosmetic surgery tourism” to the public and the other is an underground tier targeting Southeast Asian migrant workers, providing informal, high-risk services.

Objective: In this study, we quantified the clinical risks embedded within these hierarchical markets and demonstrated how digital platforms exacerbate health inequalities through algorithm-driven social media content delivery.

Methods: We conducted a dual-track content analysis of 60 TikTok videos (n=30 per group). A panel of 6 specialist physicians independently evaluated the videos using the newly developed Clinical Legitimacy and Risk Scoring (CLRS) scale. This scale assesses videos across 4 dimensions: depicted environment, operator identity, risk communication, and communication channels. Interrater reliability was evaluated using Fleiss κ.

Results: The CLRS was used by 6 specialist physicians and showed a significant safety difference between the two groups. The medical tourism group (model A) had an average video score of 3.2 (SD 0.7), with the primary content type of short-form videos being “surgery experience vlogs” (9/30, 30%). Concomitantly, although professionalism was apparent in these videos, underlying risks were regularly obscured. In contrast, the migrant worker group (model B) had an extremely low average score of 1.1 (SD 0.3), indicating a complete deviation from medical standards (P<.001) and thus posing a higher risk to patient safety. The videos in model B were mainly “home surgery demonstrations” (15/30, 50%). The interrater reliability among physicians

XSL•FO
RenderX

was high (Fleiss κ=0.85; P<.001). The primary coding team compiled the baseline descriptive data, whereas the board-certified physician panel independently evaluated the specific CLRS metrics.

Conclusions: The algorithmic architecture of TikTok creates a stratified marketplace, reinforcing socioeconomic stratification and health inequities. Specifically, regarding the informal aesthetic medical tier for migrant workers, the risk of infection has become a serious and urgent public health threat. Results of this study indicate that responsible public health interventions are urgently needed for the informal aesthetic medical tier targeting migrant workers and that there is a need to reassess TikTok's social responsibility in reviewing high-risk medical content in short-form videos.

(JMIR Infodemiology 2026;6:e88227) doi:10.2196/88227

KEYWORDS

social media; TikTok; medical tourism; health disparities; public health; esthetic surgery; migrant health; misinformation

Introduction

The combination of social media and medical aesthetics has created a new sales model for the advertising, awareness, and consumption patterns of medical aesthetic, surgical and nonsurgical treatments. Among social media platforms, TikTok, with its large number of users and powerful algorithm-driven short-form video distribution, has shown significant influence not only in reshaping aesthetic standards but also in driving the consumption behaviors of impressionable audiences, especially young people [1,2]. Although some studies have begun to explore the dermatologic, physiological, and psychosocial risks brought about by promoting medical cosmetic skin care trends to teenagers on these platforms, relatively little attention has been paid to how these platforms, within specific national and social contexts, facilitate the dissemination of information about and access channels for invasive medical cosmetic procedures [3,4].

This study focused on the TikTok platform in Taiwan, a society with a well-established domestic medical aesthetics industry. Although local medical aesthetic resources are widely available, 2 significant trends have emerged under the promotion of short-form videos by TikTok. The first is that medical aesthetic tourism continues to be popular among young people in Taiwan, who travel to South Korea or China for major medical aesthetic surgeries. This formal or semiformal market is characterized by groups consisting of online agents, influencers, and clinics that use TikTok to promote all-inclusive surgical packages. The second trend is more covert, existing within Taiwan's community of more than 830,000 Southeast Asian migrant workers, which also includes an estimated 90,000 undocumented migrant workers [5,6]. Because of multiple obstacles such as language barriers, economic limitations, and limited access to formal medical aesthetic channels in Taiwan, an underground medical aesthetic market has flourished. There is some promotion of this informal market on TikTok, but promotion mainly occurs through closed networks with specific language use, providing medical aesthetic services such as injectable fillers or thread lifting performed by unlicensed practitioners (usually fellow citizens in the beauty industry) in nonclinical settings.

Traditionally, tools such as DISCERN have long been used as the gold standard for assessing the quality of consumer health information. The DISCERN instrument is a widely used health information evaluation tool, jointly developed by the University of Oxford in the United Kingdom and patient organizations. It is designed to help patients, caregivers, and the public evaluate the quality and reliability of publications or website content related to health care and treatment options. However, DISCERN was fundamentally designed for long-form texts and websites. In modern times, applying DISCERN to fast-paced, algorithm-driven short-form videos creates significant structural gaps, and it is not suitable for evaluating such media. DISCERN often fails to capture critical multimedia dynamics, such as visual environments, unverified operator identities, and real-time interactive communication channels, which are characteristic of modern social media promotional methods. Therefore, as a new, specialized tool is needed to adapt to the changing times, we developed the new Clinical Legitimacy and Risk Scoring (CLRS) scale to assess the safety of these health care–related short-form video platforms.

The TikTok platform has developed algorithms aimed at maximizing user engagement. These algorithms inevitably isolate the two distinct demographic groups mentioned above, resulting in two parallel and unique medical aesthetics ecosystems [3]. From a physician's perspective, we assumed that both ecosystems harbor inherent risks. However, the medical aesthetics risks and health hazards associated with these two ecosystems differ in nature and severity. This outcome exacerbates social stratification and health inequalities. This is the first study to provide a scoring method based on the CLRS for these two demographic groups and conduct a comparative content analysis and expert risk assessments of videos aimed at these two groups on Taiwan's TikTok. In this research, video types were systematically classified and risk scores were quantified and verified by a panel of 6 specialist physicians. This study aimed to analyze the degree of threat to public health posed by these two parallel cosmetic medicine markets, elucidate the mechanisms by which social media algorithms operationalize health disparities among marginalized groups, and test the hypothesis that algorithmic personalization actively directs disadvantaged groups toward dangerous, unregulated medical practices.

Methods

Overview

For this study, we established two completely new and entirely different TikTok accounts to simulate real user settings. We

XSL•FO
RenderX

applied a systematic search strategy using different high-traffic hashtags for these 2 new accounts [3].

Account A (medical tourism track; model A) was set up to represent a young Taiwanese woman interested in plastic surgery and cosmetic medical treatments. The user interface was set to traditional Chinese. This account was used to actively search for and interact with relevant content using hashtags such as "#KoreanPlasticSurgery" and "#MainlandCosmetic."

Account B (informal migrant worker track; model B) was set up to represent a Southeast Asian female migrant worker working in Taiwan. The user interface was set to Vietnamese. The hashtags and keywords used for searching included "#fillerTaiwan," "#CosmeticTaiwan," and "#suntikputihTaiwan."

For both accounts, researchers repeatedly liked and saved relevant medical aesthetics videos to train TikTok's "For You" page algorithm. From May 2025 to July 2025, 60 independent videos with the highest view counts were collected, including 30 (50%) videos from model A and 30 (50%) videos from model B. To ensure content diversity, only 1 video per creator was included. The inclusion and exclusion criteria were strictly enforced. Videos were included if they were 15 to 600 seconds in length, contained specific medical advice or explicit price promotions, and had been posted within the past 12 months. Videos were excluded if they were duplicates or lacked real human elements (eg, static slideshow presentations without integrated dynamic audiovisual content).

Ethical Considerations

The study protocol was reviewed and approved by the Taipei Medical University Joint Institutional Review Board (TMU-JIRB; N202511073).

Development and Validation of the CLRS

The development of the CLRS tool used a deductive-inductive hybrid approach. In terms of deduction, the tool drew on core bioethical principles, such as the principle of nonmaleficence, as well as local medical advertising regulations (eg, Taiwan's Medical Care Act) to establish benchmark standards for medical legality.

In terms of induction, through a preliminary analysis of 60 nonsample medical aesthetics videos, we found that traditional tools such as DISCERN were not suitable for evaluating short-form medical videos. DISCERN's evaluation methods overlooked certain phenomena in new social media, such as cross-border medical intermediaries, ultra–low-cost package deals, and invasive procedures conducted in home environments, which cannot be assessed using traditional DISCERN. These insights were transformed into specific indicators across 4 key dimensions to quantify platform-specific safety signals. The CLRS was specifically designed for this study to address the lack of visually oriented tools for assessing the quality of short-form videos. However, the research team anticipates that this scale will have strong practical utility and will be widely applied to various medical topics and short-form video platforms in the future.

Data Coding and Analysis

Video Genre Classification

Drawing on methodologies used to analyze social media content, we classified videos into different "genres" based on common narrative formats. The analysis phase was also a multilevel evaluation. Initially, the primary coding team, composed of 2 independent researchers, watched the videos and outlined the baseline characteristics of each short-form video, while classifying each video as model A or model B. The researchers conducted a comprehensive assessment of the short-form videos, including audio and visual elements, and on-screen subtitles, rather than relying solely on the textual transcripts. Although the content targeted 2 different language groups (traditional Chinese and Vietnamese), both coders were core members of the research team. They used a single, unified coding book and received synchronized consensus training to ensure consistency. Any disagreements in classification were discussed and resolved through a consensus meeting led by a senior consulting physician. In this study, the primary coded variables included promotion model (A vs B), video type (eg, surgical procedure vlog and home treatment demonstration), and baseline indicators. These variables were mapped before conducting clinical evaluations to establish a descriptive framework for the platform.

Expert Physician Review and Risk Scoring

To critically quantify the clinical risks brought about by TikTok videos, a group of specialist physicians was recruited as the expert panel. The panel consisted of 2 dermatologists, 2 facial plastic surgeons, and 2 plastic surgeons, each with more than 10 years of clinical experience in aesthetic medicine. In this single-blind study, panel members were unaware of the study hypotheses and the dual-track study design. Each video was viewed at least twice, after which the panel members rated it using the innovative standardized CLRS tool.

The 6-member expert panel rated videos using digital CLRS rating cards built into a securely encrypted Google Forms interface. The CLRS assessment covers 4 domains, namely environment, operator identity, risk communication, and communication channels. Scores for each domain ranged from 1 point (highest risk and lowest legitimacy) to 5 points (lowest risk and highest legitimacy). The final rating was calculated as the average of these 4 scores (Table 1). This digital rating system eliminated transcription errors during the data conversion process, thereby assuring data accuracy.

XSL•FO
RenderX

Table 1. Clinical Legitimacy and Risk Scoring system.

Scores	Domain 1: depicted environment	Domain 2: operator identity	Domain 3: risk communication	Domain 4: communication channel
1 (highest risk)	Private residence, hotel room, or back room of a nonmedical business	Unspecified peer (eg, "sister" or "friend") or self-proclaimed beautician with no credentials shown	Active misinformation (eg, "100% safe" or "no side effects")	Encrypted, anonymous private chat (eg, Telegram or WhatsApp)
2	Unspecified or ambiguous location	Vaguely described as "experienced" or "professional," without evidence	Complete absence of any mention of risk, pain, or recovery	Private social media direct messages (eg, TikTok or Instagram)
3	Appears to be a private beauty salon or spa rather than a medical clinic	Presented as a nurse or technician, but credentials are not verifiable	Mentions minor, transient side effects only (eg, "a little swelling")	Public-facing but personal social media account
4	Appears to be a formal medical clinic, but accreditation is not specified	Presented as a physician, but the name or credentials are difficult to verify	Acknowledges a recovery period and common side effects	Official business page on a social media platform (eg, Facebook Business Page)
5 (lowest risk)	Appears to be an accredited hospital or major surgical center	Presented as a named, board-certified surgeon with verifiable credentials	Mentions specific serious potential complications and contraindications	Official clinic or hospital website with publicly available contact information

Statistical Analysis

Descriptive statistics were used to summarize the video characteristics. Mean scores of the CLRS for each model were calculated based on the average ratings from the 6 board-certified physicians. Interrater reliability for the CLRS was assessed using Fleiss κ. Group differences in mean CLRS total scores and domain scores between models A and B were evaluated using the Mann-Whitney *U* test, given the ordinal nature of the scoring scale and the nonnormal distribution of scores. A 2-tailed *P* value of <.05 was considered statistically significant. All statistical analyses were performed using SPSS (version 26.0; IBM Corp).

Results

General Characteristics and Promotional Strategies

An analysis of videos on TikTok showed that the promotional market for medical aesthetic videos clearly has distinct tiers. Promotion system model A is built on narratives of the "aspiration to become beautiful," using testimonials from influencers or celebrities and the legitimacy of clinical medical environments. Model B, on the other hand, is a completely different setting. The strategy of model B is rooted in leveraging trust within peer communities while avoiding legal medical environments and placing more emphasis on convenience, extreme cost-effectiveness, and privacy. Key differences are summarized in Table 2.

Table 2. Comparative analysis of 2 tiers of aesthetic procedure promotion on TikTok.

Comparison metrics	Model A: medical tourism (to China or South Korea)	Model B: informal procedures (for migrant workers)
Promotional strategy	Aspirational results, influencer testimonials, luxurious clinic environments, and package deals	In-group trust ("by us, for us"), extreme affordability, privacy, and the "at-home" service convenience
Common procedures	Invasive surgeries, including rhinoplasty, blepharoplasty, orthognathic (jaw) surgery, and breast augmentation	Minimally invasive procedures, including dermal filler injections (nose, lips, and chin), thread lifts, and "whitening" injections
Cost and payment	High cost, often presented as "value" packages (US $5000-$15,000); payment through formal channels	Extremely low cost, often priced per unit (eg, US $100-$300 per syringe); payment typically made in cash
Risk communication	Minimal; focuses on the recovery period, while severe complications are rarely mentioned	None; often includes explicit, false claims such as "100% safe" and "no side effects"
Depicted environment	Modern, brightly lit clinics and hospital operating rooms	Private residences, hotel rooms, or back rooms of other businesses
Operator identity	Presented as board-certified, experienced surgeons and medical teams	Presented as a fellow migrant worker, a "beautician," or an experienced peer
Communication channel	Public-facing business accounts (eg, official LINE, WeChat, or Instagram accounts)	Encrypted, private messaging apps (eg, WhatsApp or Telegram) or private Facebook groups

Video Genre Distribution

After categorizing the videos by type, 2 different narrative styles were detected (Table 3). Model A mainly used a "surgical journey vlog" format (9/30, 30%), which promotes a parasocial connection between influencers and viewers, making the surgical experience more personal [7,8]. In contrast, model B videos often feature "surgical demonstrations in personal studios"

XSL•FO
RenderX

(15/30, 50%), thereby normalizing medical aesthetic procedures performed in unsafe environments.

Table 3. Distribution of video genres by promotion model (N=60).

Genres	Model A: medical tourism (n=30), n (%)	Model B: informal procedures (n=30), n (%)
Surgery journey vlog	9 (30)	0 (0)
Clinic promotion or tour	7 (23.3)	0 (0)
Agent question and answer	5 (16.7)	0 (0)
Dramatic transformation	6 (20)	4 (13.3)
At-home procedure demonstration	0 (0)	15 (50)
Client testimonial	1 (3.3)	8 (26.7)
Product or service advertisement	2 (6.7)	3 (10)

Physician-Rated CLRS

Six board-certified physicians used the CLRS system to evaluate the 60 videos. Scores were assigned to promote the evaluation of these two different types of videos. The CLRS provides a quantitative measure of the safety differences between the videos. The average total score for model A videos was 3.2 (SD 0.7), while the average total score for model B videos was extremely low at 1.1 (SD 0.3). The interrater reliability among the 6 expert physicians was significant (Fleiss κ=0.85; P<.001), indicating a high level of consistency in risk assessment.

The CLRS analysis revealed the specific risk areas for each model (Table 4). Model A short-form videos scored highest in the "depicted environment" domain (mean 4.1, SD 0.6), successfully shaping the image of clinical medical professionals. However, this group scored lowest in the "risk communication" domain (mean 2.2, SD 0.7), confirming that the video failed to adequately inform patients about serious complications that may arise from surgery or treatment. Model B videos scored the highest risk level across all 4 domains, with the lowest scores in the "operator identity" and "risk communication" domains (mean 1.0, SD 0.2 and mean 1.0, SD 0.1, respectively). This finding indicates that these videos completely and systematically deviated from any standards of medical quality and safety.

Table 4. Mean physician-rated Clinical Legitimacy and Risk Scoring (CLRS) by domain and promotion model.

CLRS domains	Domain 1: depicted environment	Domain 2: operator identity	Domain 3: risk communication	Domain 4: communication channel	Overall CLRS score
Model A: medical tourism, mean (SD)	4.1 (0.6)	3.8 (0.9)	2.2 (0.7)	2.8 (0.7)	3.2 (0.7)
Model B: informal procedures, mean (SD)	1.1 (0.3)	1.0 (0.2)	1.0 (0.1)	1.3 (0.5)	1.1 (0.3)

Discussion

Principal Findings

As confirmed by this study, the innovative CLRS system can integrate platform-specific features (eg, communication encryption and environmental tracking) into quantifiable metrics, successfully filling the gap in DISCERN's analysis of short-form videos. In the context of modern social media, the CLRS, as a fast and visual filtering mechanism, can effectively identify potentially dangerous peer-to-peer medical advertising, thereby complementing DISCERN. DISCERN is the gold standard for assessing the quality of health information; however, because TikTok is a visually oriented platform and its videos are short form, the traditional long-form DISCERN assessment tool is not entirely applicable. Therefore, we referred to the core principles of DISCERN and developed the CLRS by synthesizing established health communication standards, such as the DISCERN instrument and the American Medical Association (AMA) benchmarks (X-9), while adapting them to the unique characteristics of short-form video platforms. The CLRS was developed using a deductive-inductive hybrid approach. The deductive method refers to the principle of nonmaleficence in medical ethics and clinical guidelines (eg, Taiwan's Ministry of Health and Welfare regulations on medical aesthetics advertising), which define what constitutes high-risk behaviors. The other method was inductive, which was used during a pilot study of 60 videos, in which phenomena not captured by existing scales were discovered, such as inducing cross-border medical treatment, presenting low-cost promotional packages, and emphasizing surgeries in nonmedical settings.

From these observations, specific subitems of the CLRS were inductively derived. The "clinical legitimacy" domain draws on regulatory standards for medical advertising (eg, local regulatory reference), while the "risk" domain was inductively derived from thematic analyses of prevalent aesthetic procedure promotion on TikTok, focusing on safety signals that DISCERN may overlook, such as the glamorization of invasive procedures in nonclinical settings.

The main findings of this study indicate that TikTok in Taiwan is not a single source of medical information but rather a stratified platform in which access to cosmetic medical procedures and related risks is heavily influenced by socioeconomic status, ethnicity, and legal identity.

XSL•FO
RenderX

Paradoxically, although such platforms are expected to enhance information transparency and equalize the cost of cosmetic medical procedures, our analysis showed the opposite. Instead, the platform fosters a more polarized and opaque market, intensifying medical risks caused by information asymmetry. These 2 tiers present fundamentally different types of risks, a finding supported by empirical evidence from the quantitative risk assessment and highly validated by a panel of expert physicians.

In our study, these 2 quantified risks, information asymmetry and physical threats, represent fundamentally different categories of risk across 2 tiers. The CLRS quantitative risk assessment table was validated by a strong consensus from a panel of medical experts, providing an empirical medical basis for this conclusion. The risk associated with model A mainly lies in the quality of care and information asymmetry. Therefore, patients participating in medical aesthetic tourism face significant risks in verifying physicians' qualifications and managing postoperative care [9]. The risk associated with model B pertains to immediate patient safety. Its CLRS score was as low as 1.1, confirming that it represents a severe public health crisis and risk of infection. The risks are urgent and multiple: practitioners without a medical background could cause catastrophic complications such as vascular embolism and severe infections [10]; there is a high chance of using illegal and nonsterile counterfeit products, leading to serious local infections and inflammation [10]; procedures performed in unhygienic environments pose a very high risk of infection and transmission of bloodborne pathogens; and the staff in this unit completely lack the ability to handle emergency complications.

The divide between these two tiers on TikTok is a direct manifestation of the broader social class disparities and health inequalities within the health care ecosystem. TikTok's digital stratification has already shifted from a virtual interface to a tangible expression of social class differences, health inequities, and socioeconomic gaps. In this study, the divide between these two tiers on TikTok in cosmetic medicine is a direct reflection of structural inequalities. The trend toward medical aesthetic tourism (model A) is jointly driven by medical intermediary agencies and globally minded consumers influenced by transnational media flows [1]. The emergence of the model B market is a direct expression of these structural inequalities. It is a form of "health care" that arises not because it is explicitly prohibited by the formal system but because structural barriers (eg, high costs, language barriers, and the well-founded fear of deportation among undocumented migrant workers when interacting with any formal institution) make straightforward medical aesthetics unattainable. Due to high costs, language barriers, or fear of immigration authorities, migrant workers may be unable to access aesthetic treatments through Taiwan's formal medical system [6,11]. The rise of this underground market indicates how marginalized groups, when structurally excluded from the formal medical aesthetic system, use peer networks to meet these aesthetic needs, even if it comes at extremely dangerous health risks. The emergence of this underground economy fills a gap and demonstrates how marginalized groups use flexible methods to fulfill these aesthetic needs when the formal system fails to reach them, despite exceedingly dangerous consequences. Short-form videos on TikTok have become a normalized channel for disseminating these risks.

Local health authorities in Taiwan face a dual challenge and contradictions of punitive regulatory enforcement, in terms of the government facing challenges in monitoring advertisements for cross-border medical services (model A) and domestic unlicensed medical practices (model B). These challenges highlight the regulatory failure of health authorities. Although authorities have established support systems for workers, such as the 1955 hotline, the main enforcement mechanism against undocumented migrant workers remains largely punitive. However, from a public health perspective, these punitive measures may have counterproductive effects. By increasing the risk of punishment when discovered, these measures may exacerbate undocumented workers' fear of any interaction with official health care units in Taiwan. This fear, combined with targeted short-form video content on TikTok, may push them further toward underground unlicensed medical aesthetic providers. Immigration enforcement may unintentionally worsen this public health crisis. This also raises key questions about platform responsibility. The way algorithms are designed may create short-form video spaces that normalize high-risk behaviors. Platforms are not neutral providers of medical aesthetic content but active participants in creating and sustaining these high-risk medical aesthetic markets.

This study has several limitations. First, a limited number of videos were selected through TikTok's algorithm, which may not represent the full picture of content on the TikTok platform. In addition, because models A and B represent simulations using only a single account, the results may reflect their individual algorithmic paths, rather than patterns that can be consistently reproduced across different users. Given that TikTok's recommendation engine is highly personalized and opaque, relying on a single simulated account to test models A and B produces localized trajectories. This lack of sample diversity makes it impossible to generalize the findings to a broader user base or reliably replicate the dataset in future studies. Moreover, the specific content pushed by the social media platform will change over time and cannot be directly reproduced.

Future research should analyze multiple accounts representing different demographic groups to better account for algorithmic variability. Second, this analysis was based solely on video content, and it could not measure the actual incidence of real-world complications. Third, this was a cross-sectional study, which could not evaluate the longitudinal results produced by the continuous contact algorithm. To address this limitation, future studies should adopt a hybrid research design incorporating longitudinal follow-up. Furthermore, in-depth qualitative interviews with users represented by model A and model B may provide deeper insight into their motivations and subsequent clinical complications.

Conclusions

Social media platforms such as TikTok have gone beyond mere information dissemination; their algorithms enhance individual engagement with the platform through carefully curated personalization. These algorithms may, in turn, strengthen and

XSL•FO
RenderX

exacerbate existing socioeconomic inequalities. In this study, the CLRS analysis verified by specialist physicians indicated that this digital stratification actively amplifies risks associated with medical aesthetics. For the public, the platform serves as a gateway to international medical tourism while concealing critical postoperative care and infection risks. For marginalized migrant worker communities, it acts as an efficient channel, guiding them toward unregulated, potentially life-threatening medical aesthetic procedures.

Ultimately, this study issues an urgent, data-driven call to action for Taiwan's public health authorities. Rather than relying solely on traditional, punitive prevention, a more proactive harm reduction strategy should be implemented. To be specific, effective policies must combine culturally tailored, multilingual health outreach with targeted administrative reforms. These approaches may contribute to dismantling the systemic barriers that prevent migrant workers from accessing professional medical aesthetic services in Taiwan. Furthermore, a paradigm shift regarding "platform accountability" is imperative. When algorithms explicitly craft a lucrative market for fatal medical negligence targeting vulnerable groups, the platform ceases to be a neutral medium. Future regulatory frameworks must compel social media corporations to take responsibility for the foreseeable public health harms propagated by their algorithmic architectures.

Acknowledgments

The funding source had no role in the design, conduct, analysis, or reporting of the study. During the preparation of this manuscript, the authors used a generative AI tool (Gemini) to assist with English grammar corrections in certain paragraphs of the Discussion and Conclusion sections. Following this, the manuscript was professionally reviewed and refined by a human English editing service in Taiwan. The authors take full responsibility for the content, accuracy, and integrity of the final publication.

Data Availability

The data presented in this study are available from the corresponding author upon reasonable request.

Funding

This study was supported by the Taipei Medical University New Faculty Research Grant (grant TMU113-AE1-B04) and the National Science and Technology Council of Taiwan (grant 114-2314-B-038-117).

Authors' Contributions

Conceptualization: CJW, LYJL, CYT
Data curation: SYW, CYT, AM, JY
Formal analysis: CJW, JMY, AM, JY
Investigation: CJW, AM, JY, JMY
Methodology: JMY, LYJL, JY
Writing—original draft: CJW, SYW, CYT, JMY
Writing—review and editing: CYT, JMY, LYJL
Supervision: LYJL, CJW

Conflicts of Interest

None declared.

References

1. Wilcox BL, Kunkel D, Cantor J, Dowrick P, Linn S, Palmer E. Report of the APA Task Force on Advertising and Children. American Psychological Association. 2004. URL: https://www.apa.org/pi/families/resources/advertising-children.pdf [accessed 2026-07-08]
2. Ali R, Komarova V, Aslam T, Peleckis K. The impact of social media marketing on youth buying behaviour in an emerging country. Entrep Sustain Issues 2022 Jun 30;9(4):125-138. [doi: 10.9770/jesi.2022.9.4(6)]
3. Hales M, Rigali S, Paller A, Liszewski W, Lagu T. Pediatric skin care regimens on TikTok. Pediatrics 2025 Jul 01;156(1):e2024070309. [doi: 10.1542/peds.2024-070309] [Medline: 40484399]
4. Rieder EA, Andriessen A, Cutler V, Gonzalez ME, Greenberg JL, Lio P, et al. Dermatology in contemporary times: building awareness of social media's association with adolescent skin disease and mental health. J Drugs Dermatol 2023 Aug 01;22(8):817-825. [doi: 10.36849/jdd.7596] [Medline: 37556525]
5. Liang LF. Everyday vulnerability: work and health experiences of live-in migrant care workers in Taiwan. In: Horn V, Schweppe C, Böcker A, Bruquetas-Callejo M, editors. The Global Old Age Care Industry. Singapore: Palgrave Macmillan; 2021.

XSL•FO
RenderX

6. Tsai TI, Lee SY. Health literacy as the missing link in the provision of immigrant health care: a qualitative study of Southeast Asian immigrant women in Taiwan. Int J Nurs Stud 2016 Feb;54:65-74. [doi: 10.1016/j.ijnurstu.2015.03.021] [Medline: 25865515]
7. Brestovanský M, Sekerešová D. Parasocial interaction does not function as a relationship mediator between the influencer's online content and the adolescent's prosocial moral reasoning and behaviour. Commun Today 2022;13(1):130-144 [FREE Full text]
8. Horton D, Wohl RR. Mass communication and para-social interaction: observations on intimacy at a distance. Psychiatry 1956 Aug;19(3):215-229. [doi: 10.1080/00332747.1956.11023049]
9. Hery D, Schwarte B, Patel K, Elliott JO, Vasko S. Plastic surgery tourism: complications, costs, and unnecessary spending? Aesthet Surg J Open Forum 2024;6(45):ojad113 [FREE Full text] [doi: 10.1093/asjof/ojad113] [Medline: 38213470]
10. Hong GW, Hu H, Chang K, Park Y, Lee KWA, Chan LKW, et al. Adverse effects associated with dermal filler treatments: part II vascular complication. Diagnostics (Basel) 2024 Jul 18;14(14):1111 [FREE Full text] [doi: 10.3390/diagnostics14141555] [Medline: 39061692]
11. Lee TH, Jarupreechachan W, Tseng YC. Exploring low-income migrant workers' health information-seeking behavior during COVID-19 in Taiwan: a qualitative study. Digit Health 2022;8(5):20552076221133764-20552076221133764 [FREE Full text] [doi: 10.1177/20552076221133764] [Medline: 36312850]

Abbreviations

AMA: American Medical Association
CLRS: Clinical Legitimacy and Risk Scoring
TMU-JIRB: Taipei Medical University Joint Institutional Review Board

Edited by C Baur; submitted 21.Nov.2025; peer-reviewed by A Shafer, N Hassan; comments to author 23.Mar.2026; accepted 19.Jun.2026; published 19.Aug.2026.

Please cite as:
Wu CJ, Wu SY, Tsai CY, Majumdar A, Yang J, Yang JM, Li LYJ
Comparative Content Analysis and Expert Physician Risk Assessment of Aesthetic Procedures Promoted on TikTok in Taiwan: Quantitative Study
JMIR Infodemiology 2026;6:e88227
URL: https://infodemiology.jmir.org/2026/1/e88227
doi:10.2196/88227
PMID:

©Cheng-Jung Wu, Sheng-Yu Wu, Cheng-Yu Tsai, Arnab Majumdar, Jeffrey Yang, Jinn-Moon Yang, Lok-Yee Joyce Li. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 19.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Emotional Expression and Mental Health Support in BTS Fandom Communities Using Natural Language Processing on YouTube Comments: Cross-Sectional Study

Nari Yoo[1], MA, PhD; Aaron H Rodwin[2], MSW, PhD; Michael Park[3], MSW, PhD; Sangpil Youm[4], MA, PhD; Sou Hyun Jang[5], MA, PhD

[1]School of Social Work, University of Michigan, Ann Arbor, MI, United States

[2]Silver School of Social Work, New York University, New York, NY, United States

[3]School of Social Work, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

[4]Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL, United States

[5]Department of Sociology, Korea University, Seoul, Republic of Korea

Corresponding Author:
Sou Hyun Jang, MA, PhD
Department of Sociology
Korea University
145 Anam-ro, Anam-dong, Seongbuk-gu
Seoul, 02841
Republic of Korea
Phone: 82 (2) 3290 2073
Email: soujang@korea.ac.kr

Abstract

Background: The global rise of K-pop has shaped youth culture and online communities, particularly through BTS, a South Korean boy band with an international fanbase known as ARMY (Adorable Representative MC for Youth). Music fandoms are increasingly engaging with digital platforms such as YouTube not only for entertainment but also as spaces for emotional expression and mutual support. Despite growing interest in the mental health potential of music-based coping strategies, limited research has examined how fandom cultures differentially express emotional needs and supportive interactions online.

Objective: This study investigates specific mental health language patterns and coping mechanisms expressed by BTS fans in online spaces, examining how different linguistic features (including self-referential language and emotional expression patterns) may reflect psychological states and mental health needs. We used YouTube comments from fan-curated “sad” or “depression” playlists of BTS. We further included YouTube comments from equivalent Taylor Swift playlists as a reference group.

Methods: Using natural language processing and Linguistic Inquiry and Word Count 2022 software, we analyzed a total of 13,224 YouTube comments: 11,772 comments on BTS "sad playlist" videos and 1452 comments on Taylor Swift equivalents. Statistical comparisons were conducted to evaluate differences in comment length, word count, pronoun use, and emotional valence. Representative comments were examined to contextualize the emotion classification results.

Results: BTS original comments were significantly longer (mean 253.38, SD 703.65 characters) and had higher word counts (mean 38.93, SD 88.54 words) than Taylor Swift original comments (length: mean 89.84, SD 330.96 characters; word count: mean 16.08, SD 64.78 words; P<.001). BTS fans used more first-person singular pronouns (mean 10.24%, SD 9.57% vs mean 7.43%, SD 9.41%) and expressed greater sadness (1691/5341, 31.7% vs 75/452, 16.6%). In contrast, Taylor Swift fans exhibited higher admiration (86/452, 19% vs 429/5341, 8%). Among reply comments, BTS fans demonstrated more caring (242/1729, 14% vs 7/128, 5.5%), gratitude (294/1729, 17% vs 15/128, 11.7%), and optimism (162/1729, 9.4% vs 6/128, 4.7%). Linguistic analysis also revealed a broader international user base for BTS, including higher proportions of Spanish (719/11,772, 6.11%) and Portuguese (222/11,772, 1.89%) comments. Examination of comment content showed that fans used these spaces to disclose personal struggles, express gratitude for the community, and offer peer support, with many describing the fandom as a safe space for emotional expression they could not access elsewhere.

Conclusions: The findings show that comments on BTS fan playlists included more emotional disclosure and more supportive replies than those on the Taylor Swift comparison, consistent with fan communities operating as informal sites of peer support

XSL•FO
RenderX

online. These results suggest implications for culturally responsive, community-based, and digitally mediated mental health interventions among youth and global populations.

(JMIR Infodemiology 2026;6:e74397) doi:10.2196/74397

KEYWORDS

BTS; emotional expression; K-pop fandom; LIWC; mental health; music-based coping; natural language processing; online peer support; Taylor Swift; YouTube comments

Introduction

Overview

Mental health conditions are a growing public health concern, with rates of depression, anxiety, and suicidality increasing across populations [1,2]. Suicide rates in the United States have risen by 57% among individuals aged 10-24 years between 2007 and 2018 [3,4], with similar trends observed globally [5]. Despite the availability of treatments, engagement with mental health services remains low, with many individuals experiencing structural, social, and psychological barriers to care [6,7]. This treatment gap has led researchers to examine alternative sources of mental health expression, support, and engagement, particularly in digital spaces.

Social media platforms contain large-scale, user-generated text that can provide insight into how individuals discuss mental health [8], seek support [9], and describe their psychological states [10]. YouTube, as one of the largest video-sharing platforms, hosts discussions around mental health through comments, vlogs, and music engagement [11,12]. Prior research has demonstrated that individuals often disclose psychological distress in these settings, creating opportunities to study mental health–related language and emotional expression at scale [13].

To specifically investigate coping mechanisms related to depressive mood, this study focuses on comments from YouTube playlists explicitly curated by fans and labeled as “sad” or “depression” playlists. This targeted approach allows for an in-depth examination of emotional expression and support-seeking in a context where users likely engage with music for mood regulation [14-16]. As interactions on online platforms, such as YouTube, have been described as a “clinical whitespace,” they provide new avenues for early detection and monitoring, which holds importance for clinical practice [17]. Further, music use has long been associated with emotional regulation, coping, and psychological well-being [18,19]. It can support emotional processing, evoke memories, and provide an outlet for self-expression [20,21].

Music, Emotions, and Mental Health

Music is widely used as a tool for emotional processing and psychological well-being [18]. It can influence mood regulation, social connection, and personal identity [20]. Studies have found that music-based interventions can improve mental health outcomes, particularly in therapeutic settings [19,22,23]. However, music engagement can also serve maladaptive functions with mixed effects, as some individuals may use music to cope in ways that reinforce distress and intensify rumination [24,25]. A growing body of research, including reports from the World Health Organization [19] and the National Institutes of Health [22], suggests that psychosocial interventions that integrate music are associated with improved engagement in treatment and mental health outcomes [23,26]. This research highlights the multisensory nature of music and its capacity to evoke emotions, imagination, and activate the senses [19,22].

Music Engagement and Mental Health Expression

Music listening is often an emotionally charged experience, and online platforms provide opportunities for individuals to share their interpretations, reactions, and personal connections to music. Prior studies have examined how song lyrics reference mental health themes and how music can influence help-seeking behaviors [27,28]. The impact of music on mental health extends beyond lyrical content to the ways in which listeners engage with and discuss music online [29,30].

Studies on music-based online communities suggest that fandoms may operate as spaces for social support, shared identity, and mental health discourse [31-33]. While music engagement can provide comfort and validation, it can also reinforce distress through maladaptive listening patterns, including rumination and avoidance [25,34,35]. An increasing body of work has begun to examine the connection between music, fandom, and well-being. For example, Laffan and colleagues [36] revealed that BTS fans’ psychological sense of community was a protective factor against the impacts of cyberbullying. Recent work has also addressed the psychological dimensions of music fandom: Pownall [37] offered reflections on how Taylor Swift’s lyrics relate to fans’ developmental experiences, while Baudinette and Scholes [38] found that K-pop fandom provides affective spaces of security and support for identity exploration. However, less research has examined how fan communities articulate emotional needs related to depressive mood in online spaces such as YouTube comments, and how these expressions manifest linguistically.

While numerous K-pop groups have gained international popularity, BTS (a K-pop boy band) is distinctive for the consistent themes of self-love, personal growth, and mental health awareness [39]. For example, a recent Billboard survey [40] reported that 62% of respondents said K-pop helps relieve stress and enhance a sense of community belonging. This focus aligns closely with the study’s objectives, making BTS’s music and fan interactions especially relevant. BTS’s massive and highly engaged global fanbase, known as ARMY (Adorable Representative MC for Youth), provides a novel data source for examining the intersection of music engagement and mental health across different cultures and demographics. For these reasons, BTS was selected as the focus for this study to gain perspectives on the intersection of music engagement and mental health [41,42].

XSL•FO
RenderX

While the focus of this study is BTS for the reasons discussed above, we also include Taylor Swift as a reference group, as she is the world's most streamed artist [43], and her music also contains mental health–related content [29,44]. Both fandoms illustrate the reach of parasocial relationships in the digital age [44-46]. However, the BTS ARMY stands out in its global reach and cultural diversity. While Swift's fandom is predominantly American, presented as Miss Americana [47], BTS has cultivated a truly international following, bridging language and culture [39].

Mental Health Discourse on YouTube Comments

Social media platforms provide large-scale data sources for analyzing mental health–related language behavior [48] and help-seeking behavior [9]. While prior research has primarily used data from platforms such as Reddit [49] and Twitter [50], YouTube has become a relevant platform for studying mental health discourse, as users engage with content that reflects their emotional states, including music playlists labeled with mental health–related themes [12,51]. Prior studies have documented how YouTube comments can provide a form of peer support, self-expression, or stigma reduction [13].

In contrast to platforms such as Reddit, where discussions are more text-based, YouTube's comment sections are tied to audiovisual content, influencing the way users articulate emotional experiences and allowing for more contextually grounded emotional expression [11]. Music-based playlists labeled as "sad" or "depression" provide a space where users engage in discussions about their mental health experiences, often in response to the emotional tone of the music. Prior research has suggested that such spaces can operate as informal support systems while also carrying risks of reinforcing distress [51].

Objective

This study examines how BTS fans express emotional needs and engage in peer support within YouTube comment sections on "sad" or "depression" playlists. Using natural language processing (NLP) techniques, we analyze linguistic patterns and emotional expressions in fan comments collected between 2018 and 2024. We address the following two research questions: (1) What linguistic characteristics and emotional tones characterize BTS fan comments across this period? (2) How do linguistic features relate to different types of emotional expression in these online communities? Taylor Swift fan comments were included as a reference group to contextualize findings.

Methods

Dataset

This study uses established NLP techniques to analyze the linguistic and emotional characteristics of YouTube comments within mental health–themed music engagement. YouTube was selected as the data source for this study, given its position as one of the largest platforms for music-related fan engagement, with comment sections that allow extended textual responses suitable for linguistic analysis of emotional expression and peer support [11]. The data consists of YouTube comments collected from the BTS "sad" or "depression" playlists. While the primary conceptual focus of this study is on BTS fandom, we also included Taylor Swift's "sad" or "depression" playlists as a reference group to enhance interpretability and context for results. The YouTube comments were collected using the YouTube application programming interface via Google. By using the search terms "sad playlist" and "depression playlist" with "BTS" or "Taylor Swift," we collected the video IDs for 16 videos for the BTS playlist and 14 videos for the Taylor Swift playlist.

The collected comments were categorized into 2 types: original comments and replies. Original comments refer to stand-alone comments made directly on the video. Reply comments, on the other hand, are responses to original comments, representing fan engagement with other fans and their opinions. In the analysis, we included 11,772 comments from 9062 users for BTS and 1452 comments from 1067 users for Taylor Swift. A total of 46 comments (23 BTS and 23 Taylor Swift) consisted of a single character and returned no analyzable text in Linguistic Inquiry and Word Count 2022 software (LIWC), so the linguistic analyses in the Results section are based on 13,178 comments (BTS: n=11,749; Taylor Swift: n=1429), and the models are based on 12,786 comments after listwise deletion of comments with no LIWC dictionary match. The data were collected in May 2024, covering comments published from October 21, 2018, to May 23, 2024, representing the entire range available on the data collection date. This study uses a cross-sectional design, analyzing aggregated patterns across this period rather than examining temporal trends. This approach enables characterization of overall linguistic and emotional patterns within fan communities while maximizing sample size for robust statistical analysis.

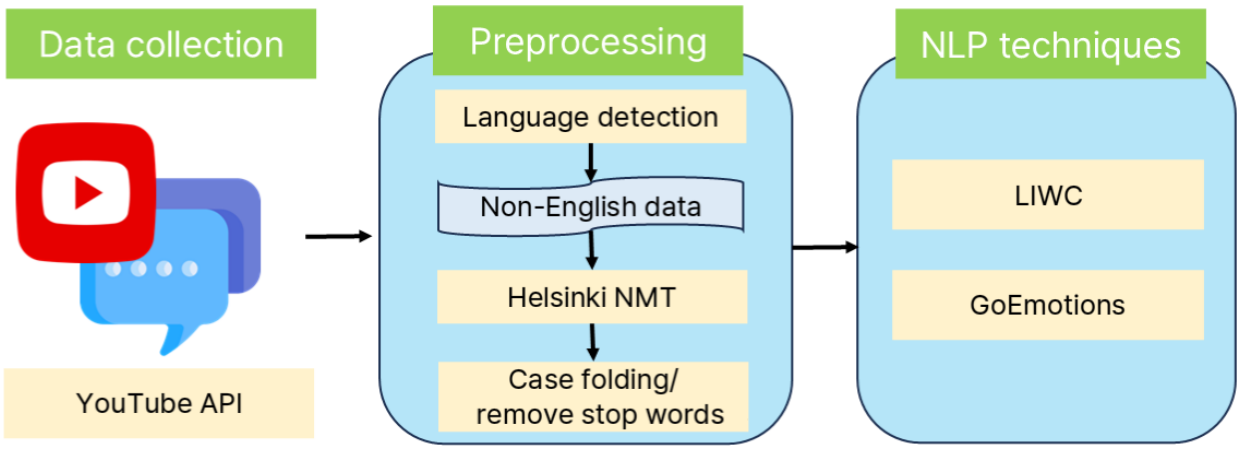
We included the comments written in non-English languages using the Helsinki model [52]. The Helsinki-NLP models were selected for their broad language coverage across over 1000 language pairs and demonstrated strong performance in open-domain translation benchmarks [53]. Among publicly available neural machine translation systems, OPUS-MT (Open Parallel Corpus machine translation) models have been widely adopted in multilingual NLP research due to their accessibility and consistent performance across low- and high-resource languages [53]. BTS fans used a diverse range of languages in their comments, with English (n=8963, 76.14%) being the most common, followed by Spanish (n=719, 6.11%), Portuguese (n=222, 1.89%), Somali (n=140, 1.19%), German (n=98, 0.83%), Indonesian (n=94, 0.8%), Korean (n=76, 0.65%), and Russian (n=69, 0.59%). In contrast, Taylor Swift fans are significantly more likely to use English (n=1166, 80.3%), with Spanish (n=42, 2.89%) and German (n=35, 2.41%) being the only other languages representing more than 2% of the comments. Other languages included Dutch (n=15, 1.03%), Norwegian (n=13, 0.9%), Indonesian (n=11, 0.76%), Somali (n=11, 0.76%), Welsh (n=10, 0.69%), Estonian (n=9, 0.62%), Afrikaans (n=7, 0.48%), Danish (n=7, 0.48%), Romanian (n=7, 0.48%), and Portuguese (n=6, 0.41%; $\chi^2_{45}=151.52$; $P<.001$).

XSL•FO
RenderX

Data Analysis

A study design is depicted in Figure 1. The research began with applying NLP techniques to the entire dataset of YouTube comments. The data preprocessing stage involved standardizing text by converting to lowercase, removing URLs, special characters, and excessive whitespace. Stop words were removed using the standard English stop-word list from the Natural Language Toolkit library. Tokenization was performed using the *spaCy* library's default English tokenizer, which segments text based on punctuation and whitespace. For comments translated into English, these preprocessing steps were applied post translation.

Figure 1. Flowchart of study design. API: application programming interface; LIWC: Linguistic Inquiry and Word Count 2022 software; NLP: natural language processing; NMT: neural machine translation.



Linguistic analysis was conducted using the LIWC with its default English dictionary [54]. We focused on first-person singular pronouns ("I") as a key linguistic marker because previous research has established strong associations between higher "I" usage and depression, psychological distress, and self-focus during emotional processing [55,56]. Similarly, first-person plural pronouns ("we") were analyzed to examine collective identity and social connection [57]. Word count and comment length were treated as indicators of emotional expressiveness and disclosure depth, which typically increase when processing difficult emotions or seeking support. Independent sample *t* tests were conducted to compare the differences between BTS's and Taylor Swift's comments.

The Bidirectional Encoder Representations from Transformers (BERT)–based GoEmotions model was used to predict 27 types of emotions [58] to analyze the range of emotions experienced by fans as they engaged with the music. The GoEmotions model was selected as one of the most fine-grained emotion classification models available, trained on 58,000 Reddit comments [58]. The model has been widely used in social media emotion analysis due to its granular emotion taxonomy and strong benchmark performance [59], and it was applied to comments that were either originally in English or translated into English [60,61]. We used the pretrained Google BERT model fine-tuned on the GoEmotions Reddit dataset without further fine-tuning on our specific dataset, using default inference parameters.

After inference, chi-square tests were conducted to compare emotions across BTS and Taylor Swift. To contextualize the quantitative emotion classification results, we systematically examined the content of comments within each emotion category. This approach allowed for an in-depth exploration of how fans articulated mental health–related experiences and engaged in peer support. Representative comments from each emotion category were selected to illustrate the range of emotional expressions and support dynamics present in the data.

To analyze how fandom affiliation and linguistic features are associated with emotional expression in comments, we constructed multilevel logistic regression models for the 5 most frequent emotions identified in our initial analysis. We used multilevel modeling with random intercepts for video ID to account for potential clustering effects, as comments within the same video might share similar emotional characteristics. The dependent variable was the binary presence (1) or absence (0) of each emotion as classified by the BERT-based GoEmotions model. Independent variables included fandom (BTS vs Taylor Swift), comment type (reply vs original), number of replies received, comment length, word count, first-person singular pronoun usage ("I"), first-person plural pronoun usage ("we"), and affiliation score from LIWC analysis.

Ethical Considerations

This study analyzed only publicly available YouTube comments and involved no interaction or intervention with individuals and no access to private identifiable information. It therefore did not meet the regulatory definition of human subjects research under 45 CFR 46.102, and institutional review board review was not sought. Ethical considerations nonetheless remain regarding user consent and privacy. While public data may not legally require informed consent, ethical guidelines emphasize respecting user expectations and minimizing harm. Users may not anticipate their comments being analyzed for research, raising concerns about privacy and autonomy. Following best practices in social media research [62], we did not analyze user IDs, usernames, or other identifiable information, ensuring that individual commenters could not be traced. Further, we still follow common ethical standards, such as removing all

XSL•FO
RenderX

personally identifiable information from the comments that we report. No informed consent was obtained and no compensation was provided because the study involved no recruitment of, or interaction with, individuals.

Results

Linguistic Analysis

Word frequency analysis showed that the term “ARMY” was mentioned in 1253 (10.64%) BTS-related comments, whereas “Swiftie” was mentioned in 10 (0.69%) Taylor Swift–related comments.

The LIWC analysis revealed significant differences in several linguistic features between comments related to BTS and Taylor Swift (Table 1). For original comments, the average length of comments was significantly longer for BTS (mean 253.38, SD 703.65 characters) compared to Taylor Swift (mean 89.84, SD 330.96 characters), with a *P* value of <.001. The word counts also showed a significant difference, with BTS-related comments averaging 38.93 (SD 88.54) words and Taylor Swift–related comments averaging 16.08 (SD 64.78) words (*P*<.001).

Table 1. Comparison of linguistic characteristics across BTS and Taylor Swift comments. Length is in characters; word count is in words; I, we, and affiliation are Linguistic Inquiry and Word Count 2022 software scores expressed as a percentage of total words.

Characteristic	Total	BTS	Taylor Swift	*P* value[a]
Original comments				
Count, n	9597	8522	1075	N/A[b]
Number of replies, mean (SD)	0.9 (8.8)	0.9 (9.3)	0.5 (2.3)	.10
Length, mean (SD)	235.1 (674.2)	253.4 (703.6)	89.8 (331.0)	<.001
Word count, mean (SD)	36.4 (86.5)	38.9 (88.5)	16.1 (64.8)	<.001
I, mean (SD)	9.9 (9.6)	10.24 (9.57)	7.4 (9.4)	<.001
We, mean (SD)	0.5 (2.2)	0.5 (2.2)	0.4 (2.1)	.08
Affiliation, mean (SD)	1.4 (3.9)	1.4 (3.7)	1.2 (5.2)	.08
Replies				
Count, n	3581	3227	354	N/A
Length, mean (SD)	99.1 (179.6)	104.0 (185.8)	54.8 (98.2)	<.001
Word count, mean (SD)	18.1 (35.2)	19.2 (36.7)	8.7 (11.4)	<.001
I, mean (SD)	7.9 (11.9)	8.0 (11.9)	6.4 (11.2)	.02
We, mean (SD)	0.9 (4.3)	1.0 (4.4)	0.4 (2.3)	.02
Affiliation, mean (SD)	2.3 (6.6)	2.4 (6.8)	1.0 (3.7)	<.001

[a]*P* values are calculated based on the *t* tests for the continuous variables (number of replies, length, word count, I, we, and affiliation).

[b]N/A: not applicable.

The usage of the pronoun “I” was higher in BTS-related comments (mean 10.24%, SD 9.57%) compared to Taylor Swift (mean 7.43, SD 9.41), with a *P* value of <.001. The usage of the pronoun “we” did not show a significant difference (*P*=.08). For reply, as depicted in Table 1, the average length of comments was significantly longer for BTS (mean 103.98, SD 185.75 characters) compared to Taylor Swift (mean 54.77, SD 98.23 characters), with a *P* value of <.001.

The word counts also showed a significant difference, with BTS-related comments averaging 19.17 (SD 36.74) words and Taylor Swift–related comments averaging 8.75 (SD 11.40) words (*P*<.001). The usage of the pronoun “I” was higher in BTS-related comments (mean 8.04, SD 11.92) compared to Taylor Swift (mean 6.43, SD 11.23), with a *P* value of .02. The usage of the pronoun “we” was significantly higher in BTS-related comments (mean 1.00, SD 4.44) compared to Taylor Swift (mean 0.43, SD 2.29), with a *P* value of .02. Affiliation scores were significantly higher for BTS (mean 2.43, SD 6.85) compared to Taylor Swift (mean 0.97, SD 3.69), with a *P* value of <.001.

Emotion Classification

As depicted in Table 2, the distribution of emotion based on the emotion classification of BTS- and Taylor Swift–related comments was significantly different across both the comments and replies (comments: χ^2_{23}=189.0; replies: χ^2_{24}=66.3; both *P*<.001).

XSL•FO
RenderX

Table 2. Emotion classification across BTS and Taylor Swift (TS) comments and replies. Percentages are calculated among comments (or replies) that the GoEmotions classifier assigned to a nonneutral emotion category: comments: n=5793 (BTS: n=5341; Taylor Swift: n=452); replies: n=1857 (BTS: n=1729; Taylor Swift: n=128).

Emotion	Original comments				Replies			
	Total, n (%)	BTS, n (%)	TS, n (%)	BTS-TS difference (%)	Total, n (%)	BTS, n (%)	TS, n (%)	BTS-TS difference (%)
Admiration	515 (8.9)	429 (8)	86 (19)	–11	170 (9.2)	150 (8.7)	20 (15.6)	–6.9
Amusement	106 (1.8)	90 (1.7)	16 (3.5)	–1.8	54 (2.9)	41 (2.4)	13 (10.2)	–7.8
Anger	55 (0.9)	49 (0.9)	6 (1.3)	–0.4	10 (0.5)	9 (0.5)	1 (0.8)	–0.3
Annoyance	41 (0.7)	39 (0.7)	2 (0.4)	0.3	15 (0.8)	13 (0.8)	2 (1.6)	–0.8
Approval	131 (2.3)	107 (2)	24 (5.3)	–3.3	84 (4.5)	76 (4.4)	8 (6.3)	–1.9
Caring	152 (2.6)	143 (2.7)	9 (2)	0.7	249 (13.4)	242 (14)	7 (5.5)	8.5
Confusion	47 (0.8)	42 (0.8)	5 (1.1)	–0.3	14 (0.8)	12 (0.7)	2 (1.6)	–0.9
Curiosity	31 (0.5)	23 (0.4)	8 (1.8)	–1.4	11 (0.6)	11 (0.6)	0 (0)	0.6
Desire	212 (3.7)	207 (3.9)	5 (1.1)	2.8	45 (2.4)	43 (2.5)	2 (1.6)	0.9
Disappointment	171 (3)	167 (3.1)	4 (0.9)	2.2	21 (1.1)	17 (1)	4 (3.1)	–2.1
Disapproval	33 (0.6)	28 (0.5)	5 (1.1)	–0.6	12 (0.6)	9 (0.5)	3 (2.3)	–1.8
Disgust	10 (0.2)	10 (0.2)	0 (0)	0.2	4 (0.2)	4 (0.2)	0 (0)	0.2
Embarrassment	6 (0.1)	4 (0.1)	2 (0.4)	–0.3	2 (0.1)	2 (0.1)	0 (0)	0.1
Excitement	19 (0.3)	16 (0.3)	3 (0.7)	–0.4	10 (0.5)	8 (0.5)	2 (1.6)	–1.1
Fear	73 (1.3)	73 (1.4)	0 (0)	1.4	12 (0.6)	12 (0.7)	0 (0)	0.7
Gratitude	886 (15.3)	825 (15.4)	61 (13.5)	1.9	309 (16.6)	294 (17)	15 (11.7)	5.3
Joy	257 (4.4)	248 (4.6)	9 (2)	2.6	73 (3.9)	66 (3.8)	7 (5.5)	–1.7
Love	1029 (17.8)	917 (17.2)	112 (24.8)	–7.6	276 (14.9)	256 (14.8)	20 (15.6)	–0.8
Nervousness	18 (0.3)	17 (0.3)	1 (0.2)	0.1	2 (0.1)	2 (0.1)	0 (0)	0.1
Optimism	103 (1.8)	94 (1.8)	9 (2)	–0.2	168 (9)	162 (9.4)	6 (4.7)	4.7
Pride	0 (0)	0 (0)	0 (0)	0	1 (0.1)	1 (0.1)	0 (0)	0.1
Realization	64 (1.1)	61 (1.1)	3 (0.7)	0.4	17 (0.9)	16 (0.9)	1 (0.8)	0.1
Remorse	42 (0.7)	39 (0.7)	3 (0.7)	0	43 (2.3)	39 (2.3)	4 (3.1)	–0.8
Sadness	1766 (30.5)	1691 (31.7)	75 (16.6)	15.1	248 (13.4)	237 (13.7)	11 (8.6)	5.1
Surprise	26 (0.4)	22 (0.4)	4 (0.9)	–0.5	7 (0.4)	7 (0.4)	0 (0)	0.4

In the original comments category, sadness showed the largest difference between the 2 fan communities, with a 15.1% higher prevalence in BTS comments (1691/5341, 31.7%) compared to Taylor Swift comments (75/452, 16.6%), suggesting that BTS fans more frequently express their personal struggles and negative feelings when commenting on BTS content. Love and admiration showed notable differences in the opposite direction. Love was more prevalent in Taylor Swift comments with a –7.6% difference (BTS: 917/5341, 17.2%; Taylor Swift: 112/452, 24.8%), and admiration showed an –11% difference (BTS: 429/5341, 8%; Taylor Swift: 86/452, 19%).

In the "reply" category, the emotional differences were equally distinct but with different patterns. Caring showed the largest difference with an 8.5% higher prevalence in BTS replies (BTS: 242/1729, 14%; Taylor Swift: 7/128, 5.5%), followed by gratitude with a 5.3% difference (BTS: 294/1729, 17%; Taylor Swift: 15/128, 11.7%), and optimism with a 4.7% difference (BTS: 162/1729, 9.4%; Taylor Swift: 6/128, 4.7%). These differences suggest a stronger emphasis on community support and positive encouragement within BTS fan interactions. Conversely, Taylor Swift–related replies showed a higher prevalence in emotions related to fan engagement, with amusement showing a –7.8% difference (BTS: 41/1729, 2.4%; Taylor Swift: 13/128, 10.2%) and admiration showing a –6.9% difference (BTS: 150/1729, 8.7%; Taylor Swift: 20/128, 15.6%). This suggests that Taylor Swift fans engage in more lighthearted exchanges and explicit artist appreciation in their interactions.

Table 3 presents the results of multilevel logistic regression models predicting 5 key emotions in YouTube comments. After controlling for linguistic features and comment characteristics, BTS fans had significantly higher odds of expressing sadness (odds ratio [OR] 2.15, 95% CI 1.24-3.76; P=.007), gratitude (OR 2.18, 95% CI 1.40-3.39; P<.001), and caring (OR 3.10, 95% CI 1.77-5.45; P<.001) compared to Taylor Swift fans.

XSL•FO
RenderX

These findings support our descriptive analysis and suggest that BTS fans more frequently use YouTube comment sections as spaces for emotional vulnerability and mutual support. In particular, the substantially higher odds of expressing caring emotions among BTS fans align with the strong sense of community observed in comment content. Reply comments showed distinct emotional patterns compared to original comments, with significantly lower odds of expressing sadness (OR 0.32, 95% CI 0.28-0.37; *P*<.001), love (OR 0.70, 95% CI 0.61-0.81; *P*<.001), and admiration (OR 0.74, 95% CI 0.61-0.89; *P*=.001), but dramatically higher odds of expressing caring (OR 4.38, 95% CI 3.53-5.44; *P*<.001). This pattern suggests a functional differentiation within these online communities: original comments tend to serve as venues for emotional disclosure and artist appreciation, while replies function primarily as supportive responses to these disclosures, creating a reciprocal support dynamic. First-person singular pronoun ("I") usage showed significant associations with all 5 emotions, increasing the odds of expressing sadness (OR 1.05, 95% CI 1.04-1.05; *P*<.001) and love (OR 1.04, 95% CI 1.04-1.05; *P*<.001), while decreasing the odds of expressing gratitude (OR 0.99, 95% CI 0.98-1.00; *P*=.001), admiration (OR 0.93, 95% CI 0.92-0.94; *P*<.001), and caring (OR 0.97, 95% CI 0.96-0.98; *P*<.001). This pattern suggests that self-referential language is more commonly used when disclosing personal emotional experiences and struggles, while other, more focused emotions such as gratitude, admiration, and caring are expressed with less self-reference. First-person plural pronoun ("we") usage had a positive association with expressing love (OR 1.06, 95% CI 1.04-1.09; *P*<.001) but a negative association with gratitude (OR 0.92, 95% CI 0.88-0.95; *P*<.001), potentially reflecting how collective identity reinforces positive emotional attachment to the artist while individual expressions dominate in gratitude.

Table 3. Multilevel logistic regression models predicting emotions in YouTube comments (n=30 groups for each emotion). Emotions were classified using the Bidirectional Encoder Representations from Transformers–based GoEmotions model. Random effects for video ID (n=30) were included to account for clustering within videos. Word count, length, and pronoun usage (I and we) were derived from Linguistic Inquiry and Word Count 2022 software analysis.

Variables	Sadness		Love		Gratitude		Admiration		Caring	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
BTS fan (reference: Taylor Swift fan)	2.15 (1.24-3.76)	.007	1.02 (0.81-1.28)	.87	2.18 (1.40-3.39)	<.001	0.98 (0.71-1.36)	.93	3.10 (1.77-5.45)	<.001
Reply comment (reference: nonreply)	0.32 (0.28-0.37)	<.001	0.70 (0.61-0.81)	<.001	0.94 (0.81-1.09)	.41	0.74 (0.61-0.89)	.001	4.38 (3.53-5.44)	<.001
Number of replies	1.00 (1.00-1.01)	.26	0.99 (0.98-1.01)	.32	1.00 (0.99-1.01)	.96	1.00 (1.00-1.01)	.25	1.00 (0.97-1.02)	.74
Length	1.00 (1.00-1.00)	<.001	1.00 (1.00-1.00)	.01	1.00 (1.00-1.00)	.006	1.00 (1.00-1.00)	.06	1.00 (1.00-1.00)	.03
Word count	1.01 (1.01-1.02)	<.001	1.00 (1.00-1.01)	.002	1.01 (1.00-1.01)	.002	1.00 (1.00-1.01)	.67	1.01 (1.00-1.01)	.003
I	1.05 (1.04-1.05)	<.001	1.04 (1.04-1.05)	<.001	0.99 (0.98-1.00)	.001	0.93 (0.92-0.94)	<.001	0.97 (0.96-0.98)	<.001
We	0.98 (0.95-1.00)	.09	1.06 (1.04-1.09)	<.001	0.92 (0.88-0.95)	<.001	0.99 (0.95-1.03)	.48	1.00 (0.98-1.03)	.77
Affiliation	1.02 (1.01-1.04)	<.001	0.99 (0.97-1.01)	.17	1.02 (1.01-1.03)	.006	0.99 (0.97-1.01)	.40	1.01 (0.99-1.03)	.37
Random effect variance	0.39 (0.18-0.82)	<.001	0.01 (0.00-0.15)	.08	0.19 (0.10-0.39)	<.001	0.05 (0.02-0.18)	<.001	0.07 (0.01-0.30)	.001

[a]OR: odds ratio.

Emotional Expression in BTS Fan Comments

To further contextualize the quantitative emotion classification results, we examined the content of comments within each emotion category. Table 4 presents representative examples that illustrate how fans expressed mental health–related experiences and engaged in peer support within the comment sections.

Commenters across multiple emotion categories described BTS-related YouTube comment sections as spaces where they could share experiences they felt unable to disclose elsewhere. One user wrote, "Let's be honest, the only place we can really express our feelings is in the YouTube comments, because no one knows us and won't judge us because of that." Another stated, "I really need to thank YouTube comment sections, because it's the only place I can vent," before describing experiences of childhood sexual abuse, a friend's suicide, a parent's cancer diagnosis, and a recent suicide attempt. Such comments suggest that the perceived anonymity and shared fan identity created conditions for disclosure that users did not find in their offline environments.

XSL•FO
RenderX

Table 4. Examples of emotional expression in BTS fan comments. Percentages are of BTS comments (n=5341) or BTS replies (n=1729) that the GoEmotions classifier assigned to a nonneutral emotion category. Quotations are reproduced verbatim; ellipses mark omitted text.

Emotion and type	n (%)	Examples
Sadness		
Comment	1691 (31.7)	"My brother died of suicide last two years ago... I started having depression after when my big brother died. Two years later, I started thinking to cut myself and suicide too. But when I watch BTS 'No More Dream'... I continue watching BTS"
Reply	237 (13.7)	"I feel you cause I have the same problems, and I also have no one to talk to... my heart is weaker than all the people who thinks I'm strong"
Love		
Comment	917 (17.2)	"I'm one of the 2 billion ARMYs that love and cherish you. Are you okay? No really are you?... as the old me was bullied for 4 years, had problems a 16-year-old should never go through"
Reply	256 (14.8)	"Army is a place to fall down on when you can't stand. U r not alone. U r loved and u r part of this family"
Gratitude		
Comment	825 (15.4)	"I can't go to a therapist since I'm a minor and they will tell my parents that they will be extremely angry... thank you for reading about my problems"
Reply	294 (17)	"Home is not the best place for me right now... People I thought cared about me, they just abandoned me... but I know I can trust ARMY"
Admiration		
Comment	429 (8)	"HEY YOU! Yes, you, sitting behind your screen reading this. I don't know you and you certainly don't know me... you're a beautiful, wonderful, talented person"
Reply	150 (8.7)	"I'm proud of you for surviving 100% of your bad days. It takes a lot of courage, and that itself makes you such an amazing existence"
Joy		
Comment	248 (4.6)	"I had never been into a fandom that treats me like their family! I am so glad that I've come into this fandom, I've never been so attached into any bands, this is the one and only. They saved many people from depression and other mental health problems."
Reply	66 (3.8)	"Ever since I became an ARMY I've been happier. And trying more. And not as suicidal as I was before"
Desire		
Comment	207 (3.9)	"I always bottled up my emotions that sometimes it blows up on the wrong time... I wanted to cry so bad, but I just can't. It's hurting me more. Make me start to think I was not worthy"
Reply	43 (2.5)	"I miss the days I used to be loved by my parents but now they hate me because I listen to BTS... I even tried committing suicide, but I was safe"
Disappointment		
Comment	167 (3.1)	"I have been struggling with anxiety for almost all of my life but now it's gotten out of control... I couldn't get into med school, keep on gaining more weight and lost all of my friends"
Reply	17 (1)	"I want to study hard and get good marks, but I always find myself procrastinating... I feel like I'm wasting my time unintentionally"
Caring		
Comment	143 (2.7)	"It's okay to cry, you're not weak nor sensitive. It's alright to make mistakes, you're not perfect. Forgive yourself of past mistakes, you weren't as wise as today"
Reply	242 (14)	"I've seen so many comments in this comment section of ARMYs who don't feel well and want someone to talk to... I'm not a therapist or something but I care and love every single ARMY member so if you want someone to talk to,, I'm always ready to listen."
Approval		
Comment	107 (2)	"ARMY had comforted me during my panic attacks even before I liked BTS. Their music stopped me from suiciding and it is the only thing that helps me through my depression"
Reply	76 (4.4)	"It's no use of going hard on yourself for that... It's long gone into past which you can't turn back. Pressuring and hurting yourself won't help"
Optimism		

XSL•FO
RenderX

Emotion and type	n (%)	Examples
Comment	94 (1.8)	"I hope you know that the emptiness and sorrow you feel are valid... I promise you it's only temporary and what you're feeling is what makes you human"
Reply	162 (9.4)	"Dear army, I hope you will find inner peace someday. Your chest might feel heavy as if the world is against you. But you will find yourself through this"
Fear		
Comment	73 (1.4)	"I am here because I suffer from depression and I am afraid to tell my parents and friends... I locked myself in the room doing self-harm"
Reply	12 (0.7)	"I understand you. My parents also fight all the time for my brother and I, the truth at this point I'm afraid to lose him since he suffers from anxiety and even though my mom said that I wanted to take him to a psychologist my dad doesn't understand that this is serious."
Remorse		
Comment	39 (0.7)	"I have been a BTS fan since 2015, they got me through a lot over the years, but I lost touch in 2022, for a number of reasons, but now I am back at my safe place that protected me over the years from depression. I am sorry for the past year me who allowed others to treat her the way they did."
Reply	39 (2.3)	"Everyone loses friends, but that's how you know who the real ones are. Armies don't hate you, we love you"

Fans frequently described the BTS fan community as a surrogate family or safe space. As shown in Table 4, a joy-coded comment (248/5341, 4.6%) stated, "I had never been into a fandom that treats me like their family! I am so glad that I've come into this fandom... They saved many people from depression and other mental health problems." A remorse-coded comment (39/5341, 0.7%) similarly described returning to "my safe place that protected me over the years from depression." The framing of these spaces as nonjudgmental environments where fans could share without fear appeared across emotion categories.

Sadness was the most prevalent emotion in original comments (1691/5341, 31.7%). These comments included disclosures of depression, grief, self-harm, and suicidal ideation. As 1 commenter shared (Table 4), "My brother died of suicide last two years ago... I started having depression after when my big brother died. Two years later, I started thinking to cut myself and suicide too. But when I watch BTS 'No More Dream'... I continue watching BTS." Sadness-coded replies (237/1729, 13.7%) often expressed solidarity: "I feel you cause I have the same problems, and I also have no one to talk to... my heart is weaker than all the people who thinks I'm strong."

Gratitude-coded comments (825/5341, 15.4%) and replies (294/1729, 17%) frequently linked expressions of thanks to descriptions of mental health struggles and barriers to formal care. One comment stated, "I can't go to a therapist since I'm a minor and they will tell my parents that they will be extremely angry... thank you for reading about my problems." A gratitude-coded reply described, "Home is not the best place for me right now... People I thought cared about me, they just abandoned me... but I know I can trust ARMY." These examples illustrate how structural barriers (age restrictions, concerns about parental reactions, and lack of trusted relationships) shaped fans' reliance on the fandom community.

Desire-coded comments (207/5341, 3.9%) and replies (43/1729, 2.5%) captured expressions of emotional suppression and unmet needs. One comment described, "I always bottled up my emotions that sometimes it blows up on the wrong time... I wanted to cry so bad, but I just can't. It's hurting me more. Make me start to think I was not worthy." A desire-coded reply revealed, "I miss the days I used to be loved by my parents but now they hate me because I listen to BTS... I even tried committing suicide, but I was safe."

Fear-coded comments (73/5341, 1.4%) and replies (12/1729, 0.7%) captured anxiety about disclosure and concerns about family members' mental health. One comment stated, "I am here because I suffer from depression and I am afraid to tell my parents and friends... I locked myself in the room doing self-harm." A fear-coded reply shared, "My parents also fight all the time for my brother and I, the truth at this point I'm afraid to lose him since he suffers from anxiety and even though my mom said that I wanted to take him to a psychologist my dad doesn't understand that this is serious."

Reply comments exhibited distinct emotional patterns reflecting peer support. While caring represented only 2.7% (143/5341) of original comments, it constituted 14% (242/1729) of replies. A caring-coded reply stated, "I've seen so many comments in this comment section of ARMYs who don't feel well and want someone to talk to... I'm not a therapist or something but I care and love every single ARMY member so if you want someone to talk to,, I'm always ready to listen." This explicit acknowledgment of the comment section as a space for mental health disclosure, combined with an offer of informal support, illustrates the peer support dynamics operating within these spaces.

Optimism similarly increased from 1.8% (94/5341) in original comments to 9.4% (162/1729) in replies. A comment stated, "I hope you know that the emptiness and sorrow you feel are valid... I promise you it's only temporary and what you're feeling is what makes you human," while a reply offered, "Dear army, I hope you will find inner peace someday. Your chest might feel heavy as if the world is against you. But you will find yourself through this."

Admiration appeared in both comments (429/5341, 8%) and replies (150/1729, 8.7%), often taking the form of direct

XSL•FO
RenderX

encouragement. A comment addressed fellow fans: "HEY YOU! Yes, you, sitting behind your screen reading this. I don't know you and you certainly don't know me... you're a beautiful, wonderful, talented person." A reply affirmed, "I'm proud of you for surviving 100% of your bad days. It takes a lot of courage, and that itself makes you such an amazing existence."

Love-coded comments (917/5341, 17.2%) and replies (256/1729, 14.8%) emphasized connection and belonging. One comment shared, "I'm one of the 2 billion ARMYs that love and cherish you. Are you okay? No really are you?... as the old me was bullied for 4 years, had problems a 16-year-old should never go through." A reply emphasized community belonging: "Army is a place to fall down on when you can't stand. U r not alone. U r loved and u r part of this family."

Approval-coded comments (107/5341, 2%) and replies (76/1729, 4.4%) validated experiences and discouraged self-criticism. One comment described, "ARMY had comforted me during my panic attacks even before I liked BTS. Their music stopped me from suiciding and it is the only thing that helps me through my depression." A reply counseled, "It's no use of going hard on yourself for that... It's long gone into past which you can't turn back. Pressuring and hurting yourself won't help."

Joy-coded replies (66/1729, 3.8%) contained references to mental health improvement alongside ongoing challenges: "Ever since I became an ARMY I've been happier. And trying more. And not as suicidal as I was before." Remorse-coded replies (39/1729, 2.3%) offered reassurance: "Everyone loses friends, but that's how you know who the real ones are. Armies don't hate you, we love you."

Discussion

Principal Results

This study investigated the linguistic characteristics and emotional expressions in YouTube comments on BTS's "sad" or "depression" playlists. Our findings reveal the diverse ways in which fans engage with the music along with other fans. For example, our findings capture how aspects of BTS music (eg, self-affirming lyrics and narratives, and melodies that carry emotion beyond language), along with the nature of fan engagement (eg, sense of community and fandom), are reflected in how fans express emotion and support one another in these spaces. Overall, our study adds to existing literature that has documented the positive effects of music engagement on mental health [19,63].

Our findings extend this work to a digital setting, YouTube comment sections. It also adds to the literature that has captured the mental health content in popular song lyrics [27,32], the impact of mental health content of popular song lyrics on important outcomes (eg, help-seeking [28]), and how fan engagement with BTS music may impact listeners' mental health [33].

The LIWC analysis showed that BTS-related original comments were significantly longer, had a higher word count, and contained more first-person singular pronouns compared to Taylor Swift original comments. Higher use of first-person singular pronouns ("I") among BTS fans may suggest stronger personal reflection or emotional investment; it may also reflect demographic differences, such as younger age [55] or greater psychological vulnerability. Further, the significantly higher prevalence of sadness expressed by BTS fans suggests that they are using the YouTube platform as a space to process and share their emotional struggles. This deeper emotional engagement is consistent with the music and possibly a sense of relatability to BTS's lyrics' message. In contrast, while expressing admiration for the artist, Taylor Swift fans appear to use the platform more for the simple enjoyment of the music.

Our attention to the individual and collective nature of music engagement fits within the Dynamic Music Engagement model [33], which sets out several routes through which music engagement can affect mental health. Specifically, the model understands music engagement in the context of the listeners, artists, fan community, and their interactions [33]. For example, we found a higher prevalence of caring, gratitude, and optimism in BTS comments compared to Taylor Swift's comments in the "reply" category. These findings suggest the strength of the BTS ARMY community in building emotional connection as an online bonding social capital [64] and support among fans worldwide, transcending language and cultural barriers [31,65]. This finding is aligned with the fact that ARMY fandom is known as the most dedicated community among K-pop fans, providing a communal experience [63]. This also builds on prior research that found that a sense of community, reflection, connection, and comfort were key experiences shared in response to listening to BTS music [33].

Building on our NLP findings, the examination of comment content revealed that BTS music and fandom may work as a coping mechanism for fans dealing with difficult emotions, trauma, and life experiences. These findings, highlighting music's use in managing difficult emotions and life experiences, align with prior research [21,24,66-68]. For example, fans may find comfort, emotional release, and a sense of understanding through the music and lyrics due to its capacity to address underlying emotions and experiences. This relates to seminal research [18], which delineates principles underlying music-evoked emotions, including resonance (eg, emotional contagion), understanding (eg, meaning-making), and social functions (eg, relational experiences).

We also found some evidence that music engagement served "unhealthy" functions at times. Comments classified under emotions such as sadness and desire included some evidence related to how some fans and listeners engaged with music in ways that may have intensified negative emotions or avoidant coping (eg, social isolation, blocking others out, and rumination). These findings add to some prior literature that has documented potential "unhealthy" and maladaptive functions of music, even when music is used with the intention of feeling better [24,35]. While more research is needed, our findings suggest a potential need for developing digital music interventions that maximize music's positive impact on mental health [69] while mitigating potential risks.

Our findings also show the barriers to mental health support that BTS fans described, with many reporting a lack of access

XSL•FO
RenderX

to or understanding of mental health services [70]. In the absence of formal mental health support, BTS and the ARMY community may fill the gap by providing a virtual space for understanding, validation, and encouragement [71]. It is aligned with the previous finding that ARMY forms a transcultural fandom with a strong sense of in-group identity, which is also linked to a broader identification with humanity [72]. This digital community provides a sense of belonging and mutual social support among its members and is perceived as a virtual home [73]. These findings emphasize the need for mental health resources, particularly for populations relying on online platforms for connection and communication.

Limitations

First, the study relied on YouTube comments as the primary data source, which may not fully represent the wider BTS or Taylor Swift fandom. Fans who actively engage in YouTube comments may have distinct characteristics or experiences compared to those who do not comment. Future research could involve in-depth interviews with BTS fans from diverse backgrounds to gain a deeper understanding of their lived experiences and how BTS has impacted their mental health and well-being. This study highlights the potential of BTS music and fandom to provide mental health support and build a sense of belonging, particularly among those from Asian backgrounds, as reported in Zhang and Ong [74].

Second, the study included Taylor Swift comments as a reference group; however, the sample size for Taylor Swift comments (n=1452) was substantially smaller than that for BTS comments (n=11,772). While a comparative analysis was not the focus of this paper, future research should aim for more balanced sample sizes to ensure robust comparative analyses.

Third, the NLP tools used in this study, while powerful, have limitations in interpreting figurative and community-specific language. Fanspeak, or specialized terminology and inside jokes unique to fandoms (eg, “Easter Eggs” for Taylor Swift fans or “borahae” for BTS fans), may not be fully captured or understood by lexicon-based tools such as LIWC, which relies on fixed dictionaries that may not cover evolving slang and community-specific expressions [75], or even context-aware models such as GoEmotions, potentially leading to misinterpretation of some community-specific expressions. Further, detecting sarcasm or distinguishing between literal and figurative expressions of emotion remains a significant challenge for current NLP techniques [76].

Fourth, while the Helsinki-NLP models are well-regarded for multilingual translation [52], a formal validation of translation accuracy for every language pair, especially for subtle emotional expressions, was beyond this study’s scope. Automated translation of emotion-laden text presents inherent challenges, as affective expressions are often culturally embedded and may lack direct equivalents across languages [77,78]. This limitation extends to subsequent emotion classification: the GoEmotions model was trained exclusively on English Reddit comments [58] and culturally specific emotional expressions, such as the Korean concepts jeong (affectionate attachment) or han (deep sorrow), may be rendered into English approximations that do not fully capture the original affective meaning. Consequently, some semantic or cultural nuances in non-English comments might not have been perfectly captured, potentially affecting emotion classification accuracy for this subset of data. Our primary analyses, however, remain robust due to the large volume of English-native comments (approximately 10,129/13,224, 76.6% of the dataset) and the consistent patterns observed across languages.

Fifth, the study did not examine differences in emotional expression patterns and linguistic characteristics across demographic groups within the BTS fandom, such as age, gender, or geographic location. Public YouTube comment data does not include user demographic information, and the YouTube application programming interface does not provide access to commenter demographics [11,79]. This limitation is common across social media research, where demographic variables must often be inferred rather than directly observed. Future research could investigate how these factors shape fans’ engagement with music and its impact on their mental health.

Sixth, our cross-sectional design does not capture temporal trends in emotional expression across the 2018-2024 study period. Future research could build on these findings by using longitudinal designs to examine how emotional expression patterns within fan communities evolve over time. The study period (2018-2024) covers events that may shape fan engagement in ways that warrant temporal analysis, including the COVID-19 pandemic, which affected both online mental health discourse [80] and music listening behaviors [81], and artist-specific events such as album releases and mental health advocacy campaigns.

Finally, while NLP provides insights into large-scale linguistic patterns, it is not designed to measure the depth of human experience or to work as a diagnostic instrument for mental health conditions [82,83]. Our findings reflect patterns of emotional expression and supportive interaction as articulated in text, rather than directly assessing internal psychological states or predicting clinical outcomes. The study identifies expressions of peer support and self-expression, but should not be equated with clinical assessment. Future research, particularly with qualitative methods, can provide a deeper context.

Overgeneralizing from commenters to the fandom as a whole also risks stereotyping fan communities [84].

Despite these limitations, this study provides a foundation for future research on the role of music and fandom in mental health, particularly across diverse cultural contexts. Addressing these limitations and ethical considerations can enhance the validity and impact of future studies, contributing to a deeper understanding of how music and artist fandoms can support mental health. Methodologically, the NLP methods demonstrated in this study offer accessible tools for mental health researchers and practitioners interested in understanding emotional expression in online communities. LIWC and pretrained emotion classification models such as GoEmotions provide validated approaches that do not require model development, making large-scale analysis of fan discourse feasible for interdisciplinary teams working at the intersection of mental health and digital media.

XSL•FO
RenderX

Conclusions

This study has several implications related to mental health. Our findings provide a resource to clinicians who wish to integrate BTS music and its themes into their work, particularly with clients from Asian and Hispanic/Latino backgrounds. Music that matches fans' emotions and experiences can be a useful component of psychosocial interventions to enhance engagement and therapeutic benefits, especially for young people [23]. The language distribution in the methods section, which revealed a significant proportion of Spanish (719/11,772, 6.11%) and Portuguese (222/11,772, 1.89%) comments among BTS fans, highlights the relevance of BTS music to Hispanic/Latino populations.

The supportive nature of the BTS ARMY community emphasizes the importance of promoting peer support and community building for individuals struggling with mental health issues. Mental health organizations can build similar online communities and safe spaces for people from diverse cultural backgrounds to connect and share their experiences, such as the example that BTS ARMY Help Center partnered with Crisis Text Line for mental health support [85].

The challenges fans face in accessing mental health support point to the need for culturally sensitive and accessible services, with mental health providers striving to create inclusive environments and tailor their approaches to the unique needs and preferences of people from diverse cultural backgrounds. The role of BTS music on fans' mental health and well-being demonstrates the potential for music and celebrity influence to promote positive mental health messages and reduce stigma [86,87]. Collaborations between mental health organizations and influential artists such as BTS can help raise awareness and encourage help-seeking behaviors in ways that draw on digital technology [88,89]. This does not imply using platforms such as YouTube directly as therapeutic tools, as they are not designed for such purposes and lack the necessary safety and privacy infrastructures for clinical intervention. Instead, collaborations could involve (1) artists partnering with established mental health advocacy groups (eg, National Alliance on Mental Illness and Active Minds) or global health organizations (eg, World Health Organization and United Nations Children's Fund [UNICEF]) to cocreate awareness campaigns, as exemplified by BTS's LOVE MYSELF campaign with UNICEF [90]; (2) insights from online fan interactions informing the design of bespoke, ethically sound digital mental health tools; and (3) artists using their platforms to direct fans toward verified mental health resources and evidence-based interventions [86,87].

Our findings suggest several potential applications for mental health practitioners. Clinicians working with those from diverse cultural backgrounds might ask about clients' engagement with music fandoms as an entry point for discussing emotional experiences and coping strategies. The prevalence of comments expressing music as a coping mechanism and community belonging suggests that fandom participation may act as an informal supplement to formal mental health care for some individuals. However, clinicians should also remain alert to potentially maladaptive patterns [25,34,35], such as excessive reliance on parasocial relationships or using music to avoid difficult emotions. Digital mental health interventions might consider drawing on the positive aspects of music engagement and community support while providing guidance toward additional evidence-based coping strategies. Such interventions could also benefit from understanding how fan communities operate across multiple platforms. Fan interactions on YouTube, with its extended comment format, may differ from those on Twitter, Reddit, Weverse, or TikTok, each with distinct affordances and community norms [91,92]. Mental health practitioners and digital intervention designers seeking to engage with fandom communities should consider these platform-specific dynamics when developing outreach strategies.

Authors' Contributions

NY, AHR, MP, SY, and SHJ contributed to the conceptualization. NY conducted data collection, data analysis, and manuscript drafting. SY provided support with data analysis. AHR contributed to writing the literature review and discussion sections. All authors reviewed and approved the final version of the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Close J, Arshad SH, Soffer SL, Lewis J, Benton TD. Adolescent health in the post-pandemic era: evolving stressors, interventions, and prevention strategies amid rising depression and suicidality. Pediatr Clin North Am 2024;71(4):583-600. [doi: 10.1016/j.pcl.2024.04.002] [Medline: 39003003]
2. Hoffmann JA, Alegría M, Alvarez K, Anosike A, Shah PP, Simon KM, et al. Disparities in pediatric mental and behavioral health conditions. Pediatrics 2022;150(4):e2022058227 [FREE Full text] [doi: 10.1542/peds.2022-058227] [Medline: 36106466]
3. Czeisler M, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
4. Protecting Youth Mental Health: The U.S. Surgeon General's Advisory. Washington (DC): US Department of Health and Human Services; 2021.

XSL•FO
RenderX

5. Rice T, Livshin A, Rihmer Z, Walther A, Bhuiyan M, Boude AB, et al. International trends in male youth suicide and suicidal behaviour. Acta Neuropsychiatr 2024;36(6):402-422. [doi: 10.1017/neu.2024.37] [Medline: 39397775]
6. Munson MR, Jaccard J, Smalling SE, Kim H, Werner JJ, Scott LD J. Static, dynamic, integrated, and contextualized: a framework for understanding mental health service utilization among young adults. Soc Sci Med 2012;75(8):1441-1449. [doi: 10.1016/j.socscimed.2012.05.039] [Medline: 22800921]
7. Key substance use and mental health indicators in the United States: results from the 2024 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. 2025. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt56287/2024-nsduh-annual-national-report.pdf [accessed 2025-12-31]
8. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2021 Presented at: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media; 2013 July 8–11; Cambridge, MA, USA p. 128-137. [doi: 10.1609/icwsm.v7i1.14432]
9. Sharma E, De Choudhury M. Mental health support and its relationship to linguistic accommodation in online communities. USA: ACM; 2018 Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; 2018 April 21–26; Montreal, QC, Canada p. 1-13. [doi: 10.1145/3173574.3174215]
10. Boettcher N. Studies of depression and anxiety using Reddit as a data source: scoping review. JMIR Ment Health 2021;8(11):e29487 [FREE Full text] [doi: 10.2196/29487] [Medline: 34842560]
11. Thelwall M. Social media analytics for YouTube comments: potential and limitations. Int J Soc Res Methodol 2018;21(3):303-316. [doi: 10.1080/13645579.2017.1381821]
12. McLellan A, Schmidt-Waselenchuk K, Duerksen K, Woodin E. Talking back to mental health stigma: an exploration of YouTube comments on anti-stigma videos. Comput Hum Behav 2022;131:107214. [doi: 10.1016/j.chb.2022.107214]
13. King CM, McCashin D. Commenting and connecting: a thematic analysis of responses to YouTube vlogs about borderline personality disorder. Internet Interv 2022;28:100540 [FREE Full text] [doi: 10.1016/j.invent.2022.100540] [Medline: 35493438]
14. Garrido S, Schubert E. Moody melodies: do they cheer us up? A study of the effect of sad music on mood. Psychol Music 2013;43(2):244-261. [doi: 10.1177/0305735613501938]
15. Van den Tol AJM, Edwards J. Exploring a rationale for choosing to listen to sad music when feeling sad. Psychol Music 2011;41(4):440-465. [doi: 10.1177/0305735611430433]
16. Gelfand A. Sad music listening: what is it good for? An exploratory study of motivations for listening to sad music [dissertation]. Indiana State University. 2025. URL: https://tinyurl.com/mr48bk5r [accessed 2026-02-02]
17. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 2018;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]
18. Koelsch S. Music-evoked emotions: principles, brain correlates, and implications for therapy. Ann N Y Acad Sci 2015;1337:193-201. [doi: 10.1111/nyas.12684] [Medline: 25773635]
19. Fancourt D, Finn S. What is the evidence on the role of the arts in improving health and well-being? WHO Regional Office for Europe. 2019. URL: https://www.ncbi.nlm.nih.gov/books/NBK553773/ [accessed 2026-08-14]
20. McFerran KS, Garrido S, Saarikallio S. A critical interpretive synthesis of the literature linking music and adolescent mental health. Youth Soc 2013;48(4):521-538. [doi: 10.1177/0044118X13501343]
21. Travis R. Rap music and the empowerment of today's youth: evidence in everyday music listening, music therapy, and commercial rap music. Child Adolesc Soc Work J 2012;30(2):139-167. [doi: 10.1007/s10560-012-0285-x]
22. Edwards E, St Hillaire-Clarke C, Frankowski DW, Finkelstein R, Cheever T, Chen WG, et al. NIH music-based intervention toolkit: music-based interventions for brain disorders of aging. Neurology 2023;100(18):868-878 [FREE Full text] [doi: 10.1212/WNL.0000000000206797] [Medline: 36639235]
23. Rodwin AH, Shimizu R, Travis R, James KJ, Banya M, Munson MR. A systematic review of music-based interventions to improve treatment engagement and mental health outcomes for adolescents and young adults. Child Adolesc Social Work J Springer Science and Business Media LLC 2023;40(4):537-566 [FREE Full text] [doi: 10.1007/s10560-022-00893-x] [Medline: 36407676]
24. Saarikallio S, Erkkilä J. The role of music in adolescents' mood regulation. Psychol Music 2007;35(1):88-109. [doi: 10.1177/0305735607068889]
25. Silverman MJ. Music-based affect regulation and unhealthy music use explain coping strategies in adults with mental health conditions. Community Ment Health J 2020;56(5):939-946. [doi: 10.1007/s10597-020-00560-4] [Medline: 31997124]
26. Geipel J, Koenig J, Hillecke TK, Resch F, Kaess M. Music-based interventions to reduce internalizing symptoms in children and adolescents: a meta-analysis. J Affect Disord 2018;225:647-656. [doi: 10.1016/j.jad.2017.08.035] [Medline: 28889050]
27. Kresovich A, Reffner Collins MK, Riffe D, Carpentier FRD. A content analysis of mental health discourse in popular rap music. JAMA Pediatr 2021;175(3):286-292 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.5155] [Medline: 33284350]
28. Niederkrotenthaler T, Tran US, Gould M, Sinyor M, Sumner S, Strauss MJ, et al. Association of Logic's hip hop song "1-800-273-8255" with lifeline calls and suicides in the United States: interrupted time series analysis. BMJ 2021;375:e067726. [doi: 10.1136/bmj-2021-067726] [Medline: 34903528]

XSL•FO
RenderX

29. Garfinkle-Crowell S. Taylor Swift has rocked my psychiatric practice. The New York Times. 2023. URL: https://www.nytimes.com/2023/06/17/opinion/taylor-swift-mental-health.html [accessed 2026-02-02]
30. Lazzer ASD. 7 reasons Taylor Swift is good for her fans' mental health. Psychology Today. 2023. URL: https://www.psychologytoday.com/us/blog/meaningfull/202310/taylor-swift-may-be-good-for-mental-health-per-new-research [accessed 2026-02-02]
31. Hieftje KD. How a Korean boy band saved my daughter. J Adolesc Health 2022;71(3):376-377. [doi: 10.1016/j.jadohealth.2022.02.022] [Medline: 35422362]
32. Jo W, Kim MJ. Tracking emotions from song lyrics: analyzing 30 years of K-pop hits. Emotion 2023;23(6):1658-1669. [doi: 10.1037/emo0001185] [Medline: 36355673]
33. Bhattacharya A, Backonja U, Le A, Antony R, Si Y, Lee J. Understanding the influence of music on people's mental health through dynamic music engagement model. In: Information for a Better World: Normality, Virtuality, Physicality, Inclusivity. Lecture Notes in Computer Science. Cham: Springer Nature; 2023 Presented at: iConference; 2023 Mar 13; Switzerland p. 91-108. [doi: 10.1007/978-3-031-28035-1_8]
34. Miranda D, Claes M. Music listening, coping, peer affiliation and depression in adolescence. Psychol Music 2009;37(2):215-233. [doi: 10.1177/0305735608097245]
35. Saarikallio S, Gold C, McFerran K. Development and validation of the Healthy-Unhealthy Music Scale. Child Adolesc Ment Health 2015;20(4):210-217 [FREE Full text] [doi: 10.1111/camh.12109] [Medline: 26726295]
36. Laffan DA, Stenson A, Flood C. The role of cyberbullying victimization in the relationship between adult BTS fans' psychological sense of community and wellbeing. J Community Psychol 2023;51(4):1479-1494. [doi: 10.1002/jcop.22924] [Medline: 35988136]
37. Pownall M. A (psychology) love letter to Taylor Swift. The British Psychological Society. 2023. URL: https://www.bps.org.uk/psychologist/psychology-love-letter-taylor-swift [accessed 2026-02-02]
38. Baudinette T, Scholes KE. K-pop fandom's affective role in shaping knowledge of gender and sexuality among LGBTQ+ fans in Australia and the Philippines. Sexualities 2024;28(4):1566-1585. [doi: 10.1177/13634607241275855]
39. Chung DJ, Koo KR. The role of cultural diversity in creating value: a case study of South Korea's pop band BTS. Asia Pac Bus Rev 2023;29(5):1275-1299. [doi: 10.1080/13602381.2023.2243242]
40. Billboard. Billboard K-pop fandom in the U.S.: a billboard reader survey. 2025. URL: https://edition.pagesuite.com/html5/reader/production/default.aspx?pubname=&edid=617f82c1-03ff-45cb-8da4-2589808f8816 [accessed 2026-02-02]
41. Ahn P, Cho M, Gonzalez VV, Neutill R, Nguyen MT, Wong Y. For youth. Bangtan remixed. Duke University Press 2024:379-383. [doi: 10.1215/9781478059615-035]
42. Tizzard DA. Who likes BTS? A survey of 400,000 fans around the world. The Korea Times. 2021. URL: https://www.koreatimes.co.kr/www/opinion/2024/07/782_306147.html [accessed 2026-02-02]
43. Stacy G. The top songs, artists, podcasts, and listening trends of 2023 revealed. Spotify. 2023. URL: https://newsroom.spotify.com/2023-11-29/top-songs-artists-podcasts-albums-trends-2023/ [accessed 2026-02-02]
44. Zafina N, Sinha A. Celebrity-fan relationship: studying Taylor Swift and Indonesian swifties' parasocial relationships on social media. Media Asia 2024;51(4):533-547. [doi: 10.1080/01296612.2024.2304422]
45. Kanozia R, Ganghariya G. More than K-pop fans: BTS fandom and activism amid COVID-19 outbreak. Media Asia 2021;48(4):338-345. [doi: 10.1080/01296612.2021.1944542]
46. Ryan Bengtsson L, Edlom J. Commodifying participation through choreographed engagement: the Taylor Swift case. AAM 2023;13(2):65-79. [doi: 10.1108/aam-07-2022-0034]
47. Prins A. On good girls and woke white women: Miss Americana and the performance of popular white womanhood. Celebr Stud 2022;13(1):102-107. [doi: 10.1080/19392397.2021.2023852]
48. Garg M. Mental health analysis in social media posts: a survey. Arch Comput Methods Eng 2023;30(3):1819-1842 [FREE Full text] [doi: 10.1007/s11831-022-09863-z] [Medline: 36619138]
49. Kamarudin NS, Beigi G, Liu H. A study on mental health discussion through Reddit. 2021 Presented at: 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM) IEEE; 2021 Sep 24; IEEE p. 637-643. [doi: 10.1109/icsecs52883.2021.00122]
50. Di Cara NH, Maggio V, Davis OSP, Haworth CMA. Methodologies for monitoring mental health on Twitter: systematic review. J Med Internet Res 2023;25:e42734 [FREE Full text] [doi: 10.2196/42734] [Medline: 37155236]
51. Chae M, Lee SS. A study on the comfort brought by music on YouTube: focusing on Generation Z. Nanotechnol Percept 2024;20(S4). [doi: 10.62441/nano-ntp.v20is4.26]
52. Tiedemann J, Thottingal S. OPUS-MT? Building open translation services for the world. 2020 Presented at: Proceedings of the 22nd Annual Conference of the European Association for Machine Translation; 2020 Nov; Lisboa, Portugal p. 479-480 URL: https://researchportal.helsinki.fi/en/publications/opus-mt-building-open-translation-services-for-the-world [accessed 2026-03-17]
53. Tiedemann J, Aulamo M, Bakshandaeva D, Boggia M, Grönroos SA, Nieminen T, et al. Democratizing neural machine translation with OPUS-MT. Lang Resour Eval 2023;58(2):713-755. [doi: 10.1007/s10579-023-09704-w]

XSL•FO
RenderX

54. Boyd RL, Ashokkumar A, Seraj S, Pennebaker JW. The Development and Psychometric Properties of LIWC-22. Austin, TX: University of Texas at Austin; 2022:1-47.
55. Pennebaker JW, Stone LD. Words of wisdom: language use over the life span. J Pers Soc Psychol 2003;85(2):291-301. [doi: 10.1037/0022-3514.85.2.291] [Medline: 12916571]
56. Tackman AM, Sbarra DA, Carey AL, Donnellan MB, Horn AB, Holtzman NS, et al. Depression, negative emotionality, and self-referential language: a multi-lab, multi-measure, and multi-language-task research synthesis. J Pers Soc Psychol 2019;116(5):817-834. [doi: 10.1037/pspp0000187] [Medline: 29504797]
57. Lee SH, Tak JY, Kwak EJ, Lim TY. Fandom, social media, and identity work: the emergence of virtual community through the pronoun “we”. Psychol Pop Media 2020;9(4):436-446. [doi: 10.1037/ppm0000259]
58. Demszky D, Movshovitz-Attias D, Ko J, Cowen A, Nemade G, Ravi S. GoEmotions: a dataset of fine-grained emotions. 2020 Presented at: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020 July; Online p. 4040-4054. [doi: 10.18653/v1/2020.acl-main.372]
59. Koufakou A, Nieves E, Peller J. Towards a new benchmark for emotion detection in NLP: a unifying framework of recent corpora. 2024 Presented at: Proceedings of the 2nd GenBench Workshop on Generalisation (Benchmarking) in NLP; 2024 Nov; Miami, FL, USA p. 196-206. [doi: 10.18653/v1/2024.genbench-1.13]
60. Ring O, Szabó MK, Guba C, Váradi B, Üveges I. "Approaches to sentiment analysis of Hungarian political news at the sentence level". Lang Resour Eval 2024;58(4):1233-1261. [doi: 10.1007/s10579-023-09717-5]
61. Salmerón-Ríos A, García-Díaz JA, Pan R, Valencia-García R. Fine grain emotion analysis in Spanish using linguistic features and transformers. PeerJ Comput Sci 2024;10:e1992 [FREE Full text] [doi: 10.7717/peerj-cs.1992] [Medline: 38855234]
62. Tanner JP, Takats C, Lathan HS, Kwan A, Wormer R, Romero D, et al. Approaches to research ethics in health research on YouTube: systematic review. J Med Internet Res 2023;25:e43060 [FREE Full text] [doi: 10.2196/43060] [Medline: 37792443]
63. Singh B, Vaswani K, Paruchuri S, Saarikallio S, Kumaraguru P, Alluri V. "Help! I need some music!": analysing music discourse and depression on Reddit. PLoS One 2023;18(7):e0287975 [FREE Full text] [doi: 10.1371/journal.pone.0287975] [Medline: 37471415]
64. Fenton A, Keegan BJ, Parry KD. Understanding sporting social media brand communities, place and social capital: a netnography of football fans. Commun Sport 2021;11(2):313-333. [doi: 10.1177/2167479520986149]
65. Lee W, Kao G. “Make it right”: why #BlackLivesMatter(s) to K-pop, BTS, and BTS ARMYs. IASPM J 2021;11(1):70-87. [doi: 10.5429/2079-3871(2021)v11i1.7en]
66. Rooks M. Novel applications of music and digital media in global health intervention and education initiatives during the COVID-19 pandemic: a case study of BTS and ARMY [master’s thesis]. New Haven, CT: Yale University; 2022. URL: https://elischolar.library.yale.edu/ysphtdl/2196/ [accessed 2026-02-02]
67. Carlson E, Wilson J, Baltazar M, Duman D, Peltola HR, Toiviainen P, et al. The role of music in everyday life during the first wave of the coronavirus pandemic: a mixed-methods exploratory study. Front Psychol 2021;12:647756 [FREE Full text] [doi: 10.3389/fpsyg.2021.647756] [Medline: 34017286]
68. Nicholas J, Boydell K, Christensen H. Self-management in young adults with bipolar disorder: strategies and challenges. J Affect Disord 2017;209:201-208. [doi: 10.1016/j.jad.2016.11.040] [Medline: 27930913]
69. Devlin K. Shaping the digital space: exploring relationships in online music therapy session delivery. Nord J Music Ther 2022;31(3):203-213. [doi: 10.1080/08098131.2022.2026453]
70. Talebi M, Matheson K, Anisman H. The stigma of seeking help for mental health issues: mediating roles of support and coping and the moderating role of symptom profile. J Appl Soc Psychol 2016;46(8):470-482. [doi: 10.1111/jasp.12376]
71. Haulenbeek SK. Mental health, community, commitment, and the growth of individual identity in the participatory fandom of the BTS ARMY. ODU Undergraduate Res J 2022;9(1). [doi: 10.25778/rd1r-ep02]
72. Newton D, Ferenczi N. Into the purple ocean: the formation and dynamics of a transcultural fandom as a result of cultural diffusion through K-pop. Psychol Pop Media 2024;13(4):760-766. [doi: 10.1037/ppm0000519]
73. Yulianto JE, Rembulan CL. Weaving the real and the virtual: young urban Indonesians’ lived experiences of making psychological homes. Vulnerable Child Youth Stud Informa UK Limited 2024;19(4):676-694. [doi: 10.1080/17450128.2024.2402396]
74. Zhang A, Ong C. "We are bulletproof": the transcultural power of fandom in #StopAsianHate. Sociol Inq 2024;94(2):391-424. [doi: 10.1111/soin.12590]
75. Bahgat M, Wilson S, Magdy W. LIWC-UD: classifying online slang terms into LIWC categories. 2022 Presented at: 14th ACM Web Science Conference 2022; 2022 June 26-29; Barcelona, Spain p. 422-432. [doi: 10.1145/3501247.3531572]
76. Băroiu AC, Trăuşan-Matu Ş. Automatic sarcasm detection: systematic literature review. Information 2022;13(8):399. [doi: 10.3390/info13080399]
77. Hershcovich D, Frank S, Lent H, de LM, Abdou M, Brandl S, et al. Challenges and strategies in cross-cultural NLP. 2022 Presented at: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); 2022 May 22–27; Dublin, Ireland p. 6997-7013. [doi: 10.18653/v1/2022.acl-long.482]

XSL•FO
RenderX

78. Troiano E, Klinger R, Padó S. Lost in back-translation: emotion preservation in neural machine translation. 2020 Presented at: Proceedings of the 28th International Conference on Computational Linguistics; 2020 Dec 8–13; Barcelona, Spain (Online) p. 4340-4354. [doi: 10.18653/v1/2020.coling-main.384]
79. Cesare N, Grant C, Nguyen Q, Lee H, Nsoesie EO. How well can machine learning predict demographics of social media users. arXiv Preprint posted online on February 6, 2017 [FREE Full text] [doi: 10.48550/arXiv.1702.01807]
80. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: observational Study. J Med Internet Res 2020;22(10):e22635 [FREE Full text] [doi: 10.2196/22635] [Medline: 32936777]
81. Fink LK, Warrenburg LA, Howlin C, Randall WM, Hansen NC, Wald-Fuhrmann M. Viral tunes: changes in musical behaviours and interest in coronamusic predict socio-emotional coping during COVID-19 lockdown. Humanit Soc Sci Commun 2021;8(1):180. [doi: 10.1057/s41599-021-00858-y]
82. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020;3:43 [FREE Full text] [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
83. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
84. Stanfill M. “They’re losers, but I know better”: intra-fandom stereotyping and the normalization of the fan subject. Crit Stud Media Commun 2013;30(2):117-134. [doi: 10.1080/15295036.2012.755053]
85. McNamara B. BTS ARMY help center partners with crisis text line for mental health support. Teen Vogue. 2020. URL: https://www.teenvogue.com/story/bts-army-help-center-crisis-text-line [accessed 2026-02-02]
86. Gronholm PC, Thornicroft G. Impact of celebrity disclosure on mental health-related stigma. Epidemiol Psychiatr Sci 2022;31:e62 [FREE Full text] [doi: 10.1017/S2045796022000488] [Medline: 36039976]
87. Calhoun AJ, Gold JA. "I feel like I know them": the positive effect of celebrity self-disclosure of mental illness. Acad Psychiatry 2020;44(2):237-241. [doi: 10.1007/s40596-020-01200-5] [Medline: 32100256]
88. 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]
89. Draganidis A, Fernando AN, West ML, Sharp G. Social media delivered mental health campaigns and public service announcements: a systematic literature review of public engagement and help-seeking behaviours. Soc Sci Med 2024;359:117231 [FREE Full text] [doi: 10.1016/j.socscimed.2024.117231] [Medline: 39278158]
90. UNICEF. UNICEF and BTS celebrate success of "groundbreaking" LOVE MYSELF campaign. UNICEF. 2021. URL: https://www.unicef.org/press-releases/unicef-and-bts-celebrate-success-groundbreaking-love-myself-campaign [accessed 2026-02-02]
91. Lee JM, Lau G. Uniting global fans and revolutionizing K-pop marketing: Weverse, a global fan community platform. Asia Pac Bus Rev 2025:23. [doi: 10.1080/13602381.2025.2534346]
92. Kim MS, Wang S, Kim S. Effects of online fan community interactions on well-being and sense of virtual community. Behav Sci (Basel) 2023;13(11):897 [FREE Full text] [doi: 10.3390/bs13110897] [Medline: 37998644]

Abbreviations

ARMY: Adorable Representative MC for Youth
BERT: Bidirectional Encoder Representations from Transformers
LIWC: Linguistic Inquiry and Word Count 2022 software
NLP: natural language processing
OR: odds ratio
UNICEF: United Nations Children’s Fund

Edited by T Nguyen; submitted 24.Mar.2025; peer-reviewed by DA Laffan, Y Zhang; comments to author 21.May.2025; revised version received 02.Feb.2026; accepted 12.Mar.2026; published 31.Aug.2026.

Please cite as:
Yoo N, Rodwin AH, Park M, Youm S, Jang SH
Emotional Expression and Mental Health Support in BTS Fandom Communities Using Natural Language Processing on YouTube Comments: Cross-Sectional Study
JMIR Infodemiology 2026;6:e74397
URL: https://infodemiology.jmir.org/2026/1/e74397
doi:10.2196/74397
PMID:41823482

XSL•FO
RenderX

©Nari Yoo, Aaron H Rodwin, Michael Park, Sangpil Youm, Sou Hyun Jang. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 31.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Social Media Constructions of Alcohol as Emotional Coping and Social Connection in Ghana: Qualitative Study

Rachel GA Thompson[1,2,3*]; Evershed Kwasi Amuzu[4*]; Tricia Thompson[3,5,6*]; Mark Kwaku Gyasi[3*]; Orlando Harris[7*]; Glenn-Milo Santos[7*]; Jerry John Ouner[6*]

[1]Language Centre, College of Humanities, University of Ghana, Accra, Greater Accra, Ghana

[2]Institute for Global Health Sciences, University of California, San Francisco, San Francisco, CA, United States

[3]Africa Interdisciplinary Research Institute, Accra, Greater Accra, Ghana

[4]Department of Linguistics, College of Humanities, University of Ghana, Accra, Greater Accra, Ghana

[5]Push Aid Africa, Accra, Greater Accra, Ghana

[6]Department of Family Health Care Nursing, School of Nursing, University of California, San Francisco, San Francisco, CA, United States

[7]Department of Community Health Systems, School of Nursing, University of California, San Francisco, San Francisco, CA, United States

[*]all authors contributed equally

Corresponding Author:
Jerry John Ouner
Department of Family Health Care Nursing
School of Nursing
University of California, San Francisco
490 Illinois St
Suite 92H
San Francisco, CA, 94158
United States
Phone: 1 4154764442
Email: jerry.nutor@ucsf.edu

Abstract

Background: Alcohol is a leading contributor to disease and injury worldwide; yet, it remains embedded in cultural practices associated with pleasure, social connection, and emotional relief. While global research has documented both the epidemiological burden and the social meanings of drinking, little is known about how alcohol-related meanings are circulated and normalized within African social media environments.

Objective: This study explores how alcohol is represented in Ghanaian social media discourse, with particular attention to emotional coping, social interaction, and everyday well-being.

Methods: The study draws on a qualitative dataset of publicly available user-generated content (N=189) from widely used social media platforms, Facebook and TikTok. It uses discourse analysis within a social constructionist framework, thereby treating language as a site where alcohol-related meanings are produced, negotiated, and circulated on online platforms.

Results: Alcohol was routinely constructed as a resource for coping with stress, heartbreak, and economic strain, while also functioning as a facilitator of friendship, leisure, and communal belonging. Alcohol was further normalized as part of everyday mood regulation, with users presenting drinking as a means of enhancing relaxation, pleasure, and bodily comfort. A broader commercial-lay hybrid discourse was identified, in which promotional messaging, therapeutic claims, and peer communication became increasingly intertwined. Through these discourses, alcohol was associated with vitality, emotional restoration, and well-being.

Conclusions: By situating these constructions within Ghana's broader cultural and digital context, the study demonstrates how global narratives of drinking for self-care and leisure are adapted and recirculated through local social media practices. The study contributes one of the first qualitative analyses of alcohol-related social media discourse in West Africa and highlights the importance of public health communication strategies that engage with culturally meaningful and digitally circulating narratives rather than relying solely on risk-based messaging.

(JMIR Infodemiology 2026;6:e102546) doi:10.2196/102546

XSL•FO
RenderX

KEYWORDS

alcohol; social media discourse; emotional coping; social drinking; digital health communication; Ghana

Introduction

Background

Alcohol maintains a paradoxical role within both public health and daily social life. On the one hand, it is a leading contributor to disease, injury, and mortality worldwide [1]. On the other hand, it remains deeply embedded in cultural practices associated with pleasure, social connection, leisure, and emotional relief [2,3]. In many societies, drinking is represented as a normal way to manage stress, strengthen relationships, and transition into relaxation or celebration. These meanings are not fixed but are continually produced, negotiated, and reinforced through everyday communication [4]. The public health burden associated with alcohol use is substantial. The World Health Organization (WHO) [3] estimates that alcohol contributes to 2.6 million deaths annually worldwide. Alcohol consumption is associated with more than 200 diseases and injuries, including liver cirrhosis, cardiovascular diseases, cancers, road traffic injuries, interpersonal violence, and mental health disorders [5-7]. The burden is especially pronounced in low- and middle-income countries, with the WHO African Region recording some of the highest alcohol-attributable mortality and disability-adjusted life year rates globally [8]. Research further shows strong bidirectional relationships between alcohol use and mental health conditions such as depression and anxiety, where psychological distress may contribute to alcohol use while heavy drinking simultaneously worsens emotional well-being [9].

In Ghana, these global patterns are reflected in national trends. WHO [3] estimates total alcohol per capita consumption at 4.5 liters of pure alcohol annually. Alcohol use remains a significant public health concern, particularly among adults aged 15 years and older, with consumption patterns generally higher among men than women. According to a GhanaWeb [10] report citing the Mental Health Authority of Ghana, 3765 alcohol-related mental health cases were recorded in 2023. Kyei-Gyamfi et al [11] note that among adolescents, alcohol is the most commonly used substance. They link early alcohol use to peer pressure, curiosity, and coping motives. These trends occur alongside rapid urbanization and changing leisure cultures [12,13].

Alongside this epidemiological evidence of harm, alcohol is often represented as an ordinary and even beneficial part of life. Research from high-income settings shows that advertising, entertainment media, and digital communication frequently portray drinking as integral to sociability, emotional regulation, relaxation, and pleasure [14,15]. With the rise of social media, these representations have become increasingly participatory and peer-driven. Studies of user-generated content on platforms such as Facebook, X (formerly known as Twitter), Instagram, and TikTok suggest that alcohol is commonly depicted as humorous, socially expected, and emotionally rewarding [16,17]. Social media platforms, therefore, do not simply reflect attitudes toward alcohol. They also shape how drinking is normalized, justified, and embedded within social interactions.

Despite the global reach of social media platforms, most research on alcohol-related digital discourse remains concentrated in Western societies [18]. African contexts remain comparatively underexplored, even though social media is now deeply embedded in everyday communication and social interaction across the continent. This gap is significant because the meanings attached to alcohol and the discursive resources used to justify its consumption are shaped by local sociocultural contexts. However, there is limited research examining how Ghanaian social media users themselves construct alcohol in relation to emotional coping and social interaction. Understanding these constructions is important for digital health communication because social media increasingly shapes how health-related behaviors, coping practices, and everyday forms of self-care are interpreted. Without attention to these lived and digitally circulating meanings, public health communication risks becoming detached from the social realities through which alcohol use is legitimized and sustained. This study addresses this gap by asking: How is alcohol represented in Ghanaian social media discourse as a resource for emotional coping and social interaction?

Theoretical Grounding: Social Constructionism and Discourse Theory

This study is grounded in social constructionism, which provides its overarching theoretical orientation. Social constructionism posits that social realities, health behaviors, and risk perceptions are not objective, preexisting facts waiting to be discovered, but are produced, negotiated, and maintained through social interaction [19,20]. From this perspective, public understandings of alcohol use, including concepts of pleasure, intoxication, coping, and harm, are not simply fixed biological or psychological realities. Instead, they emerge through culturally and historically specific ways of talking about drinking, risk, friendship, and leisure [19]. The significance of alcohol, therefore, extends beyond its physiological effects to include the meanings attached to drinking through everyday communication. This perspective shifts the focus from measuring individual psychological motivations to examining how particular understandings of alcohol are constructed through language and interaction.

Discourse theory provides the broader analytical approach through which these constructions are examined. Within discourse-oriented research, language is understood as constitutive rather than descriptive; it does not simply reflect an external social reality but actively helps create, sustain, and contest meanings [4,21,22]. Discourses can be understood as broader ways of organizing language and meaning around particular subjects, shaping how those subjects can be understood and discussed within specific contexts [23]. Through discourse, behaviors may be normalized, justified, challenged, or stigmatized. Within public health, competing discourses therefore shape how health, responsibility, risk, and acceptable behavior are understood [24]. While clinical and epidemiological discourses often frame alcohol through pathology, harm, and

XSL•FO
RenderX

risk, popular and commercial discourses may construct drinking through narratives of vitality, emotional relief, social connection, and everyday usefulness. These competing ways of talking about alcohol help explain how drinking can remain socially meaningful even where its health risks are well documented.

In the Ghanaian context, these discursive processes intersect with local economic, cultural, and regulatory conditions. Alcohol has long been intertwined with social and ritual practices, ranging from traditional libation ceremonies to the widespread availability of commercial beverages, locally distilled spirit known as akpeteshie, and the marketing of products such as bitters and sachet spirits [12,25,26]. These cultural and commercial contexts are important to this study because they shape the meanings available for talking about alcohol. Examining alcohol-related discourse in Ghana, therefore, involves more than identifying individual attitudes toward drinking. It involves examining how alcohol is made meaningful and how some forms of consumption are presented as legitimate, desirable, or necessary within particular social contexts.

Interpretative repertoires provide the specific analytical tool used in this study to examine these recurring constructions. They are recurrent, flexible, and culturally familiar systems of terms, descriptions, and figures of speech used to characterize, evaluate, and explain social phenomena [4]. They provide users with shared linguistic resources for constructing versions of reality, explaining actions, and managing how their drinking is understood in interaction [27]. Previous studies demonstrate how people mobilize these repertoires to construct health behaviors, illness experiences, and moral identities [27]. Within discussions of alcohol use, interpretative repertoires can depict drinking as essential for relaxation, handling emotional stress, a sign of sociability, or part of daily coping habits. In this study, identifying these repertoires enables the analysis of recurring ways in which alcohol is constructed as useful, legitimate, and socially meaningful.

Social media provides an appropriate setting for examining these constructions because meanings are produced and negotiated through everyday interaction. Facebook and TikTok bring together peer commentary, commercial messages, humor, storytelling, and locally grounded forms of expression. In Ghana, users combine global social media trends with expressions in local languages and Ghanaian Pidgin English (GhPE), an English-lexified contact variety widely used in informal communication [28]. The analysis of interpretative repertoires in these spaces allows the study to examine recurring constructions of alcohol in naturally occurring digital communication. The focus is not only on user-generated content about alcohol but also on the discourse that makes drinking seem appropriate, beneficial, desirable, or socially expected.

Methods

Research Design

This study adopts a qualitative, exploratory design grounded in social constructionism and follows the discourse-analytic tradition associated with Potter and Wetherell [4] and Wetherell [22]. Within this tradition, discourse is understood as language-in-use that is action-oriented and constitutive: it participates in constructing, negotiating, and making meaningful social realities, identities, evaluations, and forms of action. The analysis is therefore interpretive rather than representative. Consistent with contemporary discourse studies of digital communication, discourse in this study is not restricted by text length or the presence of extended conversational exchange. Social media posts and comments are treated as discursive acts because they are produced for public audiences and perform social actions such as justifying behavior, expressing emotions, constructing identities, negotiating relationships, and circulating culturally shared understandings. Although many social media contributions are brief, they remain analytically valuable because they draw upon broader cultural resources and shared systems of meaning [29]. The study is therefore concerned not with whether alcohol actually produces the effects attributed to it by users, but with how alcohol is discursively constructed as meaningful, useful, desirable, therapeutic, or socially significant within everyday online interactions. It seeks to identify the shared interpretative repertoires through which alcohol is constructed as a resource for emotional coping and social interaction in Ghanaian social media.

Data Source and Platform Selection

Data consist of publicly available user-generated content, including posts and comments, from 2 widely used platforms in Ghana: Facebook and TikTok. Facebook was selected for its extensive user base across multiple demographic groups in Ghana, its support for longer-form messages and community group discussions, and its rich, multimodal content, including text, images, and video. Although the broader social media environment contained visual content such as photographs, videos, emojis, Graphics Interchange Format files, and stickers, this study focused exclusively on textual discourse. During data preparation, all nontextual elements were excluded from the analytical dataset. That is, the analysis was limited to posts and comments and did not examine visual representations of alcohol. This decision ensured analytic consistency and allowed the study to focus specifically on how alcohol was constructed through language. TikTok, which has rapidly grown in popularity in West Africa, was included to capture shorter, performative, and humor-driven content circulating through sounds, hashtags, and challenges. These 2 platforms offered complementary windows into different interactional styles. Facebook enabled analysis of more elaborate narrative accounts and exchanges, while TikTok provided access to condensed, affectively charged, and memetic expressions. These affordances shape how alcohol is represented and the intensity of affective expression in messages.

Only textual content that was publicly visible without login or membership was included. No attempts were made to access locked accounts, individual demographic parameters (such as age, gender, or social status of commenters) were not collected, and no interactions with users occurred. While using publicly available data does not require formal institutional ethics approval, ethical considerations remained central to the research process. In line with established guidelines for internet-mediated research [30,31], all usernames were deleted, and care was taken to minimize potential harm.

XSL•FO
RenderX

Sampling and Search Strategy

A purposive, keyword-driven sampling strategy was used to identify messages in which alcohol was explicitly linked to emotional coping, mood regulation, or social interaction. Prior to data collection, 2 bachelor-prepared research assistants conducted a month of preliminary ethnographic observation of Ghanaian social media spaces to become familiar with common phrasing patterns. This immersion informed the development of a search lexicon that included English terms (eg, alcohol, drink, drunk, beer, and wine), Ghana-specific beverages and brand names (eg, akpeteshie, pito, palm wine, Kasapreko, Adonko, and Alomo Bitters), and colloquial expressions (eg, "I need a drink," "make I drink," "broken heart medicine," "body dey pain me," and "man must chill"). Searches were conducted using the native search functions of Facebook and TikTok, supplemented by manual scrolling through relevant pages. To ensure comprehensive coverage while prioritizing recency, data collection began with the most recent year (2025) and moved backward to 2018. This backward sampling approach enabled the identification of current discursive patterns while also assessing their continuity over time.

The initial search yielded 266 messages that made some reference to alcohol. These were then screened against inclusion and exclusion criteria. For inclusion, messages were required to be in English, GhPE, or code-switched constructions involving a Ghanaian language and to make a direct discursive connection between alcohol and an emotion, mood state, or interpersonal context. Additionally, the messages had to be clearly produced by a user based in Ghana or by a user participating in a recognizably Ghanaian conversational community, a judgment made through indicators such as geotags, self-reported location, linguistic markers, and references to Ghana-specific places, events, and public figures. Messages were excluded if they were purely commercial promotions that lacked user commentary or interaction, duplicate content, or news headlines shared without personal commentary.

After this screening process, 189 messages were retained for analysis. The corpus comprised 153 Facebook messages and 36 TikTok messages, of which 41 were original posts and 148 were user comments. The predominance of comments reflects the study's interest in naturally occurring public discussions surrounding alcohol-related content. These comments typically appeared in response to birthday celebrations, party announcements, alcohol advertisements, public health messages, and other alcohol-related posts. Therefore, much of the dataset was situated within interactional contexts. As will be shown in the next section, examples included discussions surrounding palm wine consumption, the use of bitters such as Alomo Bitters, and references to alcohol as a remedy for physical or emotional distress. These discussions provided additional insight into how alcohol-related meanings are shared and reinforced on social media.

In instances where a message was lengthy or addressed multiple topics, only the portion that directly concerned alcohol in relation to emotion or social interaction was extracted, and the extraneous content was not coded. While the analysis dataset is modest in size, it is consistent with qualitative discourse-analytic research, where depth of interpretation is prioritized over breadth. Sampling continued until no substantively new discursive patterns were observed, suggesting that key repertoires had been sufficiently captured [32].

Analytical Procedure

The analysis followed an iterative, interpretive process informed by the discourse-analytic procedures outlined by Potter and Wetherell [4] and Wiggins [27]. The authors (RGAT, TT, and MKG) began by repeatedly reading the dataset to become familiar with its content. During this initial familiarization stage, each message was assigned a unique alphanumeric identifier consisting of the prefix "P" (denoting a public message) followed by a sequential number (eg, P1 and P2) to facilitate data management, coding, and retrieval. The authors paid attention to recurring phrases, metaphors, and the social actions performed in messages (eg, complaining, bonding, joking, and self-medicating) and recorded their initial observations. These identifiers were subsequently used throughout the analysis and the Results section to reference specific data excerpts while maintaining the anonymity of individual contributors.

The analysis moved from preliminary observations to the identification of interpretative repertoires. An interpretative repertoire is understood as a broadly shared, relatively coherent way of talking about a phenomenon that furnishes speakers with a range of linguistic resources for constructing it [22]. Through iterative grouping and refinement of codes, the team identified a set of repertoires that constructed alcohol, for example, as a therapeutic agent, as a social connector, and as a routine mood regulator. These repertoires were then compared across the dataset to identify similarities and differences in how alcohol was constructed. Table 1 illustrates this analytical progression using selected examples from the corpus, showing how recurring textual features were developed into descriptive codes and subsequently interpreted as elements of broader interpretative repertoires. The examples are illustrative rather than exhaustive and are intended to make transparent how the analytical categories were developed from the data.

XSL•FO
RenderX

Table 1. Illustrative progression from data extracts to interpretative repertoires.

Representative extract	Initial code	Interpretative repertoire	Analytical category
"You are not in Alcoholism. Your own na to cool your mind sharp"	Alcohol relieves stress	Alcohol as coping resource	Alcohol as emotional repair
"I'm drinking alcohol now because cocoa prices have fallen"	Drinking as response to economic distress	Alcohol as coping resource	Alcohol as emotional repair
"He was with me when I had a broken heart"	Alcohol as emotional companion	Alcohol as coping resource	Alcohol as emotional repair
"We club with club beer to socialise and brainstorm"	Alcohol facilitates interaction	Alcohol as social bonding	Alcohol as social connector
"No matter the fyt, we still sit and drink Guinness and Smirnoff together"	Drinking sustains friendship	Alcohol as social bonding	Alcohol as social connector
"Palm wine is the only authentic drink our gods approve"	Alcohol linked to tradition and spirituality	Alcohol as cultural belonging	Alcohol as social connector
"Alomo dey for every mood"	Alcohol improves mood	Alcohol as emotional regulation	Alcohol as everyday mood regulation
"One sip and you're already feeling like you're at a village festival, good vibes, sweet taste, zero stress"	Alcohol produces positive affect	Alcohol as emotional regulation	Alcohol as everyday mood regulation
"Obuase Bitters ... strengthens your immunity and powers your stamina"	Alcohol framed as health-enhancing	Commercial-health discourse	Commercial-lay hybrid discourse
"Palm wine contains yeast ... it relieves stress and is good for pregnant women"	Personal testimony used to support health claims	Commercial-health discourse	Commercial-lay hybrid discourse

Alongside the mapping of repertoires, the analysis examined how posts positioned speakers and audiences. For each repertoire, we examined the identities being claimed or ascribed, such as the person seeking emotional relief, the loyal friend for whom drinking together is a sacred ritual, and the hardworking individual who has earned the right to relax. We also considered how relationships with audiences were constructed through shared cultural knowledge, humor, and social expectations. Given that the data were drawn from 2 platforms with different affordances, the analysis remained sensitive to platform context without treating them as separate datasets. Facebook content often permitted more extended narrative development and was frequently accompanied by photographs of drinks or drinking scenes that reinforced their textual meanings. TikTok content, in contrast, tended toward brevity, irony, and memetic humor, with text overlaid on short video clips and the creative use of platform-specific features such as soundtracks and duets. While the study did not analyze visual content, it remained attentive to how each platform's communicative affordances shaped the style, tone, and structure of textual contributions. These variations were treated as complementary expressions within a shared digital environment in which alcohol-related meanings were produced and circulated.

The identified repertoires were then examined in relation to the broader functions they served within the dataset. These included alcohol as a social connector, covering friendship, leisure rituals, and communal identity; alcohol as emotional repair, encompassing references to stress, heartbreak, and bodily discomfort; and alcohol as everyday mood regulation, linked to routine leisure and the transition into weekend time. Commercial and promotional messages that had been shared, commented on, or repurposed by ordinary users were also analyzed for how they interfaced with and reinforced these lay repertoires. Throughout the process, the authors engaged in peer debriefing, maintained a reflexive audit trail, and actively tested interpretations against deviant cases in order to enhance the rigor and credibility of the analysis.

All authors were familiar with Ghanaian social media culture and understood local drinking practices. This helped them interpret the data more deeply and accurately. However, because of this familiarity, there was also a risk of overlooking assumptions or treating certain meanings as obvious. To address this, the authors continually reflected on how their backgrounds and perspectives might influence their interpretations throughout the analysis. The authors also remained attentive to the performative nature of social media discourse, where humor, exaggeration, and promotional language may shape how alcohol is represented. Particular care was taken not to infer audience beliefs or broader social consensus from individual messages. Instead, interpretations focused on how alcohol was presented, justified, normalized, or made meaningful within the messages themselves, rather than treating individual contributions as evidence of what social media users generally believed. Thus, the analysis focused on the social and rhetorical functions performed through constructions of alcohol use.

Ethical Considerations

All data were collected from publicly accessible social media spaces. Thus, ethical clearance was not obtained from any institution. However, in line with the Association of Internet Researchers' guidelines [30], all usernames, profile images, and other identifying markers have been removed from the excerpts reported in this paper to protect anonymity and minimize any risk of harm. No direct interaction with users occurred at any stage of the research.

XSL•FO
RenderX

Results

Overview

Of the 189 messages analyzed, 3 interpretative repertoires were identified through which alcohol was discursively constructed in Ghanaian social media discourse. These repertoires portrayed alcohol as a resource for emotional repair, a social connector, and a means of routine mood and bodily regulation. The analysis focused on how particular ways of talking about alcohol make drinking appear meaningful, legitimate, desirable, or socially appropriate, without assuming that individual messages directly reflect personal beliefs or behaviors. A fourth cross-cutting repertoire, termed commercial-lay hybrid discourse, shows how promotional language and everyday user narratives intersect to reinforce these constructions. In the excerpts that follow, the textual contents are reproduced verbatim except where usernames and other potentially identifying details have been removed or altered to protect privacy.

Alcohol as Emotional Repair

The first major repertoire positioned alcohol as an accessible, effective remedy for a range of emotional discomforts. Within this repertoire, several recurring discursive patterns were identified: everyday stress, heartbreak, and romantic loss.

Everyday Stress and Mental Pressure

In the following, alcohol is framed as a necessary tool for calming the mind.

> *Chairman. You are not in Alcoholism. Your own na to cool your mind sharp.* [P1]

P1 pushes back against the stigma of being labeled an "alcoholic" by drawing a line between problematic drinking and a purposeful act. It performs identity work by resisting the potentially stigmatizing label of "alcoholic." The commenter constructs drinking as a purposeful and rational response to emotional strain, positioning alcohol use as a form of active emotional management rather than as evidence of dependency or loss of control. The phrase "cool your mind" (figurative expression in GhPE) draws on a culturally familiar repertoire in which emotional distress is framed as something that can be actively managed through drinking.

Similarly, economic distress was explicitly named as a trigger for drinking: "I'm drinking alcohol now because cocoa prices have fallen" (P56).

The blunt causal structure presents alcohol use as an immediate and natural response to economic distress. The comment links an economic issue (falling cocoa prices) to an individual coping practice (drinking). By making this connection directly, the commenter constructs alcohol use as a commonsense response to economic disappointment, without presenting the drinking itself as requiring explanation or justification. The brevity of the statement reinforces this construction by treating the connection between economic hardship and alcohol consumption as self-evident within the communicative context. Alcohol is thus positioned as a way of responding to the pressures of economic hardship.

Heartbreak and Romantic Pain

The metaphor of alcohol as "medicine" for a broken heart was a highly visible repertoire. This trope drew on both global digital culture and local bitters brands that already carry medicinal connotations.

> *Please help me say thank you to this wonderful friend of mine. He was with me when I had a broken heart, when I lost my mum he was there and also when Business was like a Methodist hymn he was with me. To the Ultimate in beer GULDER aka Booooosu.* [P15]

In P15, a beer brand is anthropomorphized as a steadfast companion through bereavement, romantic pain, and business stagnation. Through personification, the beer brand is portrayed as a dependable companion that offers emotional support across various life difficulties. It is clear that the poster uses a range of symbols and portrays alcohol as a symbol of loyalty, companionship, and emotional connection.

> *... it's a good appetizer for people recovering from sickness and people who have lost appetite for food due to broken heart and love related complications ...* [P114]

Here, alcohol is described as helpful both for "people recovering from sickness" and for those who have "lost appetite for food due to broken heart and love-related complications." By placing emotional pain alongside physical illness, P114 treats heartbreak as a condition with bodily consequences, particularly loss of appetite.

Alcohol as a Social Connector

Another repertoire constructed alcohol as central to the formation and maintenance of social relationships. Drinking was associated with friendship, celebration, communal identity, and cultural belonging.

Drinking as the Substance of Friendship

> *No matter the fyt, we still sit and drink Guinness and Smirnoff together ... d love is deep ruff ... Guinness nd Smirnoff is bae ... try it.* [P45]

P45 constructs shared drinking as a relational practice through which friendship is maintained despite interpersonal conflict. The poster foregrounds the continued ability to drink together as evidence of an enduring social bond. The exaggerated affectionate expression "d love is deep ruff" (the love is very strong) and the description of Guinness and Smirnoff as "bae" (an affectionate term meaning beloved) further reinforce this construction by presenting the drinking relationship in terms of closeness, affection, and familiarity. Alcohol consumption, therefore, functions as a symbolic marker of solidarity and relational continuity.

> *The 3 things in life that are dearest to me: My wife, Star beer, and fufu!* [P38]

By placing beer alongside a spouse and fufu, a culturally valued staple food, the poster elevates alcohol to an important, emotionally meaningful component of everyday life. He

XSL•FO
RenderX

normalizes alcohol by embedding it within familiar narratives of comfort, pleasure, and fulfillment.

Drinking Rituals, Communal Identity, and Call-and-Response

> *In my community ... Sam1 will shout ULTIMATE nd de other will respond PARADISE ... This is written on this faithful companion We also got love 4 it too.* [P16]

Call-and-response exchanges are a familiar communicative practice in many Ghanaian social settings. The use of the call-and-response pattern "ULTIMATE" ... "PARADISE" around the alcohol brand transforms alcohol consumption into a participatory communal performance rather than an individual activity. Participation signals familiarity and belonging within the community.

> *Apobona Fatawu we club with club beer to socialise and brainstorm 1,2 and a mini for the road.* [P2]

Here, drinking is directly linked to social interaction and conversation. The reference to "socialize and brainstorm" constructs alcohol as a facilitator of interpersonal engagement and collective reflection. In Ghanaian colloquial use, "1,2" conveys the idea of activities unfolding continuously or one after another, while "a mini for the road" refers to a final small drink before departure. P2 thus normalizes drinking as an integral part of extended social interaction.

Celebrations and Life Course Events

> *Happy birthday to you my day one brother. May you live to see your great grandchildren ... 365 days nor bi easy my brother so enjoy till you booze.* [P37]

The directive "365 days nor bi easy my brother so enjoy till you booze" (365 days is not easy, my brother, so enjoy till you are drunk) portrays intoxication as an appropriate reward for enduring the difficulties of everyday life. Here, drinking is presented as part of marking survival, celebration, and personal milestones.

> *Your Valentine isn't complete without Kasapreko in the mix because, every great #love story deserves a bold taste, that lingers.* [P46]

This promotional post links alcohol to romantic fulfillment by suggesting that celebrations of love are incomplete without drinking Kasapreko. Alcohol is positioned as something that enhances intimacy and emotional experience.

Palm Wine, Tradition, and Cultural Authenticity

> *Palm wine is the only authentic drink our gods approve. It has to be taken fresh from the lips of the palm tree early in the morning. We travel from the cities to the villages just for Palm wine. The best we can do is freeze some quantities for family and friends in the city ...* [P28]

This comment portrays palm wine as more than an alcoholic beverage. It is presented as culturally authentic and closely associated with tradition and rural life. The phrase "our gods approve" invokes a cultural repertoire that links the drink to long-standing spiritual and communal practices, such as libations and rituals. The effort involved in transporting and freezing palm wine for relatives further frames it as a socially meaningful commodity through which care and social connection are expressed.

> *In Ghana, palm wine or nsafufuo is a traditional drink made from fermented palm sap. It's central to social and spiritual life, known as the drink of the gods.* [P19]

This comment constructs palm wine as more than an alcoholic beverage. By describing it as "the drink of the gods," the poster draws on a culturally recognizable discourse that links palm wine to spirituality and tradition, as shown in P28. The comment thus positions palm wine as a culturally valued substance whose significance extends beyond consumption.

Alcohol as a Resource for Everyday Mood and Bodily Regulation

Beyond emotional distress or social occasions, a third repertoire positioned alcohol as part of the routine management of everyday mood, especially in relation to leisure, relaxation, weekend enjoyment, and bodily well-being.

Leisure, Relaxation, and Emotional Enhancement

> *Alomo dey for every #mood ... Don't be left out this weekend.* [P3]

This post projects the alcohol brand as a flexible emotional resource capable of fitting any affective circumstance. By claiming that "Alomo dey for every #mood" (Alomo is available for every mood), the poster presents drinking as a broadly applicable strategy for managing everyday emotional states. The reference to the weekend also connects drinking to participation in shared leisure culture.

> *What better way to slide into the TGIF mood than with a bottle of Alomo Bitters and a fun emoji game?* [P10]

Here, alcohol is shown as a resource that aids the shift from a busy workweek into relaxation and enjoyment. The phrase "slide into the TGIF mood" depicts relaxation as a positive emotional change, with alcohol serving as a facilitator of this transition. Consequently, drinking is portrayed as a natural transition from work duties to leisure.

> *I kept it simple ... grilled a full chicken, paired it with an Alomo Bitters cocktail lol. Lazy day but still premium.* [P77]

In P77, alcohol is constructed as a resource for enhancing ordinary leisure experiences. The contrast between a "lazy day" and a day that is "still premium" positions the Alomo Bitters cocktail as the element that elevates an otherwise unremarkable occasion. Drinking is therefore presented not as a response to distress but as a means of adding value, enjoyment, and distinction to everyday life.

> *Obuasi bitter can make ur day.* [P13]
>
> *One sip and you're already feeling like you're at a village festival good vibes, sweet taste, zero stress.* [P49]

XSL•FO
RenderX

Through expressions such as "make ur day" and "one sip and you're already feeling," both P13 and P49 attribute some transformative powers to Obuase Bitters. They present positive emotional states as immediate and attainable through consumption. That is, they normalize drinking as an ordinary means of regulating affect and generating pleasurable experiences.

Some users also drew on the language of hormones and emotional balance to justify drinking as a form of self-regulation.

> *So no benefits from drinking responsibly? After all me here needs that happy hormones to release stress, wat do I gain from stress n sad hormones.* [P52]

The commenter rhetorically questions a public health message about "drinking responsibly" by emphasizing alcohol's perceived emotional benefits. The phrase "happy hormones" draws on quasi-biomedical language to construct drinking as a legitimate and even rational response to stress. By contrasting "happy hormones" with "stress" and "sad hormones," the commenter mobilizes a health-oriented vocabulary that frames alcohol consumption as beneficial to emotional well-being rather than as a potential health risk.

> *I will rather stay away from women than alcohol. Alcohol releases hormones and women impact pains.* [P63]

In this context, alcohol is compared to intimate relationships. While relationships are associated with "pains," alcohol is depicted as a more dependable source of emotional support and steadiness. Just as shown in P52, here, the commenter moves beyond pleasure-seeking reasons and reframes alcohol consumption as a reasonable strategy for maintaining emotional well-being.

Bodily Discomfort and Functional Use

Some messages presented alcohol, particularly herbal bitters and palm wine, as a practical remedy for bodily discomfort, especially appetite loss and digestive problems.

> *After taking Adonko bitters I'm able to finish the Ghana fufu with light soup.* [P42]
>
> *This is appetite stimulant.* [P5]
>
> *It helps my bowel movement.* [P22]
>
> *Yebu didi (Akan) "We get drunk to eat."* [P157]

These comments link drinking to eating, digestion, and bodily comfort. They present alcohol not mainly as recreation but as a tool for daily well-being. Drinking is portrayed as a practical way to cope with appetite loss and discomfort. It is seen as a normal part of daily health habits, which helps to normalize its use and makes it more acceptable.

A similar functional logic also appeared in relation to cognitive performance and studying.

> *But please what about we those will take before we can learn if i have exams i have to take before everything.* [P98]

In P98, alcohol is described as necessary for studying, especially during examinations. This challenges conventional assumptions that alcohol impairs cognitive performance. Through this reversal, alcohol is discursively reconstructed as a facilitator of concentration and mental readiness rather than a potential obstacle to learning.

Commercial-Lay Hybrid Discourse

The final repertoire captures the overlap between promotional messaging and everyday user narratives across the 3 repertoires presented earlier. In the examples that follow, it is clear that commercial actors and ordinary users are not engaged in separate meaning-making processes. Both contribute to a shared discursive environment in which alcohol is linked to vitality, immunity, and emotional fulfillment. As these meanings circulate across posts, comments, testimonials, and promotional messages, the distinction between advertising claims and everyday knowledge becomes increasingly blurred.

> *Obuase Bitters: crafted from pure herbs, the enduring alcoholic beverage that strengthens your immunity and powers your stamina. Try it out and thank Dr Amo Collins later.* [P14]

Here, the poster endorsed Obuase Bitters by using advertising language with the authority attached to the company founder, "Dr Amo Collins." The references to "pure herbs," "immunity," and "stamina" frame the product in therapeutic terms. It blurs the boundary between commercial promotion and health-oriented advice.

> *Paim[SIC] wine containes[SIC] yeast it even cure's fewer[SIC] for one to believe what am saying watch what you are eating or drinking, it relief you from stress and is good for pregnant women cause it increase breast milk, am witness to it.* [P61]

This relies on personal testimony to construct palm wine as medically beneficial. The commenter attributes multiple functions to palm wine, including fever relief, stress reduction, and increased breast milk production. The phrase "am witness to it" positions lived experience as evidence.

> *Get Smirnoff Ghana Smirnoff pineapple punch with a touch of chicken wings this hot afternoon and trust me ya will be in heaven.* [P94]

P94 frames alcohol as a source of pleasure and sensory enjoyment. The combination of the drink with food and the relaxed setting of a "hot afternoon" constructs drinking as part of an enjoyable leisure experience.

> *Opening Nkulenu Palm Wine may cause sudden happiness, loud laughter, and unnecessary enjoyment.* [P48]

This post playfully imitates the language of product warnings by using the phrase "may cause." However, instead of warning about harm, the post lists desirable emotional outcomes such as "happiness," "loud laughter," and "enjoyment." The reversal presents the pleasurable effects of alcohol as desirable and socially valued outcomes of drinking.

XSL•FO
RenderX

Discussion

Principal Findings

This study examined how alcohol is represented as a resource for emotional coping and social interaction in Ghanaian social media discourse. The findings reveal a rich discursive landscape in which alcohol is constructed as an accessible and legitimate resource for managing emotional, social, and bodily concerns. Within the identified repertoires, alcohol is seen not only as a beverage but also as a culturally significant means of managing stress, heartbreak, physical discomfort, social duties, and leisure activities. The analysis shows how social media users draw on shared interpretive tools to normalize drinking, justify its use, and weave it into everyday ideas of well-being, friendship, and cultural identity. A commercial-lay hybrid discourse illustrates how promotional claims and everyday user narratives converge to reinforce these constructions.

The findings extend existing scholarship on alcohol and emotion regulation, which has predominantly approached the issue through psychological or epidemiological lenses, by showing how coping itself is actively constructed in everyday digital discourse. In these messages, drinking is not simply reported as behavior. It is rhetorically organized to accomplish social actions such as justifying relaxation (You are not in Alcoholism), fostering resilience through humor (broken heart medicine), building solidarity through shared cultural language (ULTIMATE ... PARADISE), and claiming a culturally sanctioned form of self-care. More specifically, the findings demonstrate how discourse functions as a mechanism through which particular coping practices become intelligible and socially acceptable. Through humor, testimony, personification, and appeals to common sense, users construct alcohol consumption as an appropriate response to emotional difficulty. From this perspective, alcohol can be understood as a culturally available resource through which individuals make sense of and manage emotional experience. This aligns with social constructionist approaches that treat language as constitutive [4,33].

These coping constructions are intertwined with the ways alcohol is used to organize social life and relational identity. The findings confirm and culturally specify the well-documented role of alcohol in social bonding [2,18]. They also support existing work, suggesting that alcohol is frequently represented as a medium through which intimacy, celebration, and relational closeness are performed and maintained [34]. Messages that framed Valentine's celebrations or birthdays as incomplete without alcohol illustrate how drinking becomes symbolically tied to affection, solidarity, and shared emotional experience. However, the data highlight nuances that are less visible in Western contexts. The personification of beer brands as companions in moments of grief or business difficulty (P15) extends beyond conventional constructions of drinking as a shared social practice. Similarly, call-and-response drinking chants (P16) and the practice of traveling to rural areas to obtain fresh palm wine for sharing with family (P28) situate alcohol within broader traditions of orality, communal care, and rural-urban kinship. In this sense, drinking is not merely a leisure activity but an enactment of relational obligation and cultural belonging. From a discourse-theoretical perspective, these constructions do more than describe existing social relationships. They help reproduce and sustain particular understandings of friendship, loyalty, reciprocity, and communal participation. Alcohol becomes a symbolic resource through which social identities are performed and affirmed [4,22,34].

Alcohol is also consistently integrated into ordinary life as a routine means of mood regulation and transition into leisure. Messages that frame drinking as appropriate for "lazy days," weekends, or any emotional state illustrate how alcohol becomes normalized as a resource for managing everyday affect. In the Ghanaian context, the weekend discourse is particularly salient as it draws on both the cultural familiarity of communal end-of-week gatherings and the globalized TGIF (Thank God It's Friday) ethos. Within this framing, alcohol is often constructed as the means to "slide into" weekend mood, positioning it as a benign and, at times, necessary boundary marker between labor and restoration. This reflects broader processes of normalization, in which repeated associations render alcohol use routine, expected, and socially unremarkable, particularly within digitally mediated leisure cultures [17,35]. The significance of this repertoire lies in how these ordinary references make drinking part of everyday emotional life without requiring explicit advocacy for alcohol use.

The findings related to appetite stimulation, digestion, and studying further illustrate how alcohol is constructed as a functional resource embedded in everyday life. The repeated emphasis on appetite suggests that these messages construct eating ability as closely linked to bodily comfort and well-being. Similarly, messages describing alcohol as necessary for studying position drinking as cognitively enabling and supportive of concentration. These constructions challenge dominant public health narratives that primarily associate alcohol with impairment, irresponsibility, and loss of control [36]. They align more closely with culturally grounded understandings, in which alcohol is constructed as a functional resource capable of addressing practical bodily and emotional concerns [37]. The repertoire, therefore, expands the perceived usefulness of alcohol beyond recreation and positions drinking as compatible with productivity, health, and everyday competence.

A broader commercial-lay hybrid discourse emerges, in which promotional language, health claims, and everyday user talk become increasingly difficult to distinguish. Messages that frame bitters as immune boosters or palm wine as medically beneficial are not just treated as external marketing messages. They are taken up, circulated, and reformulated within peer communication. Through these processes, alcohol becomes associated with well-being, vitality, and emotional restoration. This convergence illustrates how commercial alcohol promotion draws legitimacy from locally meaningful health beliefs, while everyday users actively reproduce and adapt these claims in ways that blur boundaries between consumption, care, and cultural knowledge. It is noteworthy that alcohol promotion on social media operates through participatory, peer-driven forms of communication rather than solely through formal advertising. Users do not simply consume marketing messages; they actively reinterpret, personalize, and redistribute them through humor,

XSL•FO
RenderX

testimony, lifestyle performances, and everyday interaction. In this sense, social media users function simultaneously as consumers and informal promoters of alcohol-related meanings, reflecting broader dynamics of prosumer culture and participatory advertising within digital environments. The persuasive force of these messages may partly derive from their appearance as authentic peer-to-peer communication instead of institutional advertising. This finding highlights the need to reconsider how alcohol promotion is regulated in Ghanaian digital spaces, where existing advertising codes may not fully address the participatory, rapidly evolving nature of social media promotion [13,38]. Traditional regulatory approaches are typically designed to monitor formal advertising. However, the findings suggest that alcohol-related promotional messages increasingly circulate through user-generated content, testimonials, humor, and peer-to-peer sharing. This creates regulatory challenges because many such messages fall outside conventional definitions of advertising. Future policy discussions should therefore consider how public health protections can be extended to participatory digital spaces without undermining legitimate forms of online expression.

In all, the findings show that alcohol in Ghanaian social media discourse was more often constructed as a multifunctional resource embedded in emotional, social, and cultural life than as a risky substance. Across the repertoires, therapeutic, relational, and commercial meanings intersect to make drinking intelligible, useful, and socially acceptable in everyday life. These constructions have implications for health communication because they show that alcohol is normalized not only through promotion, but also through ordinary practices of coping, social connection, and leisure.

Implications

Comments referring to alcohol as “medicine” and as a resource for socialization and even “brainstorming” pose a significant challenge for health promotion. Such messages frame drinking as beneficial to health and socially normative, thereby establishing a discursive environment, in which messages emphasizing harm or promoting abstinence may encounter difficulties in gaining traction. Effective health communication in this context may therefore need to engage with, rather than dismiss, these everyday meanings. This includes developing alternative, culturally resonant scripts for coping and connection that acknowledge the emotional and social functions alcohol currently fulfills, while gradually reframing these functions in healthier directions.

Understanding diverse social media constructions of alcohol may also help public health practitioners and clinicians develop more targeted approaches to prevention and intervention [39,40]. The distinction observed in the present data between alcohol framed as a means of managing negative experiences, such as stress or heartbreak, and alcohol framed as a means of enhancing social experiences has parallels with established research on drinking motives, including relief and reward motives [39,40]. These motivational patterns are clinically relevant, as research suggests that relief-oriented drinking may be associated with greater difficulties during treatment than drinking motivated primarily by social or enhancement motives [39,40]. However, the present study does not establish that users who describe alcohol in these terms meet criteria for alcohol use disorder or correspond to particular clinical subtypes.

This distinction is important for interpreting the implications of the findings. Social media messages that construct alcohol as “medicine” may provide useful entry points for health communication, especially around coping with stress and emotional challenges. Conversely, messages that link drinking to friendship and social enjoyment may require distinct strategies. Instead of relying on generic warnings about alcohol-related harm, campaigns could co-design messages with communities that recognize the social and emotional roles attributed to drinking while also suggesting practical alternatives that can be integrated into the same digital environments. Brief motivational interventions that address specific reasons for drinking have been shown to be effective, particularly among individuals whose alcohol consumption is driven by social or enhancement motives [41]. These strategies highlight the importance of moving away from a one-size-fits-all approach toward health communication that considers the diverse meanings of alcohol use in everyday life.

Limitations

Several limitations should be noted. The dataset, while rich, is drawn from public messages and does not capture private messaging or offline talk, where different discursive conventions may apply. Since the messages were publicly accessible, users might also have filtered or curated what they posted in ways that differ from communication intended for smaller or more private audiences. The analysis was also restricted to textual posts and comments. Images, videos, emojis, Graphics Interchange Formats, and other visual elements were excluded from the analytical corpus. Consequently, the study does not capture the multimodal ways in which alcohol may be represented across social media platforms. In addition, the use of reverse chronological sampling, beginning with the most recent messages and moving backward, prioritized recency over equal temporal coverage. Although this approach enabled the identification of contemporary discursive patterns and an assessment of their continuity over time, it may have overlooked earlier or less persistent forms of discourse that were not visible in more recent periods. The dataset contained both original posts and user comments, but the study did not systematically compare discursive patterns across these 2 forms of participation. Future research could examine whether alcohol-related meanings are constructed differently by content creators and audience members.

The analysis did not differentiate between messages that are humorous, ironic, or sincere; while discourse analysis treats all such utterances as meaningful social actions, future research could usefully explore the specific rhetorical effects of humor in normalizing drinking. Some user-generated content also featured in recurring discussion threads on palm wine, bitters, and alcohol’s perceived health benefits, but the study did not conduct a detailed interactional analysis of conversational sequences. Future studies could examine how alcohol-related meanings are negotiated, challenged, or reinforced across extended comment exchanges. The dataset is limited to content

XSL•FO
RenderX

from Facebook and TikTok. Given that social media participation is shaped by access, digital literacy, and platform preferences, the data may overrepresent more digitally active populations and exclude interactions occurring in closed or less visible digital spaces. The study's focus on Ghana limits its generalizability, though the hybridity observed (global platforms combined with local cultural forms) may be relevant to other African social media contexts. Finally, the absence of user demographic data precludes analysis of how gender, age, and social status shape these discourses. The few messages referencing women (eg, P63) suggest gendered patterns that warrant further investigation. Consistent with a discourse analytic approach, the study does not seek to measure the prevalence of attitudes or behaviors, but to examine how meanings are constructed and circulated. The findings should therefore be understood as illustrative of dominant and recurring discursive patterns, not as representative of the Ghanaian population as a whole.

Conclusions

This study provides one of the first in-depth, discourse-analytic accounts of how alcohol is represented as a resource for emotional coping and social interaction in Ghanaian social media. It demonstrates that users draw on a set of interpretative repertoires—alcohol as emotional repair, as a social connector, and as a routine mood regulator—to construct drinking as a reasonable, effective, and culturally endorsed response to the demands of everyday life. These repertoires are further reinforced by a commercial-lay hybrid discourse that links alcohol products to health, vitality, and emotional well-being, while adapting broader narratives about drinking for self-care and leisure to Ghanaian expressive contexts. The findings show that alcohol-related discourse cannot be understood only through risk or harm. The meanings attached to drinking are also shaped by emotional needs, social relationships, cultural practices, and everyday understandings of well-being. These constructions may compete with risk-based health messages when such messages do not engage with the meanings that make alcohol appealing or useful in everyday life. Health communication interventions could therefore be tailored to address constructions of alcohol as emotional relief, social connection, or everyday mood regulation, while offering practical alternatives that acknowledge the emotional and social functions attributed to drinking. The study highlights the value of examining alcohol use as a discursive and culturally situated practice, particularly within participatory digital spaces where commercial and everyday forms of alcohol-related communication increasingly intersect.

Funding

No financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this study.

Conflicts of Interest

None declared.

References

1. Doyle A. Global burden of disease—what the data tell us and how to address it. Drugnet Ireland 2022(82):17.
2. Room R, Babor T, Rehm J. Alcohol and public health. Lancet 2005;365(9458):519-530. [doi: 10.1016/S0140-6736(05)17870-2] [Medline: 15705462]
3. Global status report on alcohol and health and treatment of substance use disorders. World Health Organization. 2024. URL: https://www.who.int/publications/i/item/9789240096745 [accessed 2026-08-26]
4. Potter J, Wetherell M. Discourse and Social Psychology: Beyond Attitudes and Behaviour. Thousand Oaks, CA: Sage Publisher; 1987.
5. Rehm J, Gmel GE, Gmel G, Hasan OSM, Imtiaz S, Popova S, et al. The relationship between different dimensions of alcohol use and the burden of disease—an update. Addiction 2017;112(6):968-1001 [FREE Full text] [doi: 10.1111/add.13757] [Medline: 28220587]
6. Rumgay H, Shield K, Charvat H, Ferrari P, Sornpaisarn B, Obot I, et al. Global burden of cancer in 2020 attributable to alcohol consumption: a population-based study. Lancet Oncol 2021;22(8):1071-1080 [FREE Full text] [doi: 10.1016/S1470-2045(21)00279-5] [Medline: 34270924]
7. Shield K, Manthey J, Rylett M, Probst C, Wettlaufer A, Parry CDH, et al. National, regional, and global burdens of disease from 2000 to 2016 attributable to alcohol use: a comparative risk assessment study. Lancet Public Health 2020;5(1):e51-e61 [FREE Full text] [doi: 10.1016/S2468-2667(19)30231-2] [Medline: 31910980]
8. Nadkarni A, Garber A, Costa S, Wood S, Kumar S, MacKinnon N, et al. Auditing the AUDIT: a systematic review of cut-off scores for the Alcohol Use Disorders Identification Test (AUDIT) in low- and middle-income countries. Drug Alcohol Depend 2019;202:123-133. [doi: 10.1016/j.drugalcdep.2019.04.031] [Medline: 31349205]
9. Puddephatt JA, Irizar P, Jones A, Gage SH, Goodwin L. Associations of common mental disorder with alcohol use in the adult general population: a systematic review and meta-analysis. Addiction 2022;117(6):1543-1572 [FREE Full text] [doi: 10.1111/add.15735] [Medline: 34729837]

XSL•FO
RenderX

10. 3,765 cases of mental disorders recorded in 2023 due to alcohol—Mental Health Authority. GhanaWeb. 2024. URL: https://www.ghanaweb.com/GhanaHomePage/NewsArchive/3-765-cases-of-mental-disorders-recorded-in-2023-due-to-alcohol-Mental-Health-Authority-1937422 [accessed 2026-08-21]
11. Kyei-Gyamfi S, Kyei-Arthur F, Alhassan N, Agyekum MW, Abrah PB, Kugbey N. Prevalence, correlates, and reasons for substance use among adolescents aged 10-17 in Ghana: a cross-sectional convergent parallel mixed-method study. Subst Abuse Treat Prev Policy 2024;19(1):17 [FREE Full text] [doi: 10.1186/s13011-024-00600-2] [Medline: 38424633]
12. Kabore A, Afriyie-Gyawu E, Awuah J, Hansen A, Walker A, Hester M, et al. Social ecological factors affecting substance abuse in Ghana (West Africa) using photovoice. Pan Afr Med J 2019;34:214 [FREE Full text] [doi: 10.11604/pamj.2019.34.214.12851] [Medline: 32180886]
13. Sweetland AC, Norcini Pala A, Mootz J, Kao JC, Carlson C, Oquendo MA, et al. Food insecurity, mental distress and suicidal ideation in rural Africa: evidence from Nigeria, Uganda and Ghana. Int J Soc Psychiatry 2019;65(1):20-27 [FREE Full text] [doi: 10.1177/0020764018814274] [Medline: 30479180]
14. Nicholls J. Everyday, everywhere: alcohol marketing and social media—current trends. Alcohol Alcohol 2012;47(4):486-493. [doi: 10.1093/alcalc/ags043] [Medline: 22532575]
15. Atkinson AM, Bellis M, Sumnall H. Young peoples' perspective on the portrayal of alcohol and drinking on television: findings of a focus group study. Addict Res Theory 2013;21(2):91-99. [doi: 10.3109/16066359.2012.687795]
16. McCreanor T, Lyons A, Griffin C, Goodwin I, Moewaka Barnes H, Hutton F. Youth drinking cultures, social networking and alcohol marketing: implications for public health. Crit Public Health 2013;23(1):110-120. [doi: 10.1080/09581596.2012.748883]
17. Lyons A, McCreanor T, Goodwin I, Moewaka Barnes H. Introduction to youth drinking cultures in a digital world. In: Lyons A, McCreanor T, Goodwin I, Moewaka Barnes H, editors. Youth Drinking Cultures in a Digital World: Alcohol, Social Media and Cultures of Intoxication. 1st Edition. New York: Routledge; 2017:1-11.
18. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci 2010;33(2-3):61-83 [FREE Full text] [doi: 10.1017/S0140525X0999152X] [Medline: 20550733]
19. Burr V. Social Constructionism. 3rd Edition. Abingdon, Oxfordshire: Routledge; 2015.
20. Gergen KJ. An Invitation to Social Construction. 3rd Edition. Thousand Oaks, CA: SAGE Publications; 2015.
21. Fairclough N. Discourse and Social Change. Cambridge: Polity Press; 1992.
22. Wetherell M. Positioning and interpretative repertoires: conversation analysis and post-structuralism in dialogue. Discourse Soc 1998;9(3):387-412. [doi: 10.1177/0957926598009003005]
23. Foucault M. The Archaeology of Knowledge. Abingdon, Oxfordshire: Routledge; 1972.
24. Lupton D. Risk. 2nd Edition. Abingdon, Oxfordshire: Routledge; 2012.
25. Akyeampong E. Drink, Power, and Cultural Change: A Social History of Alcohol in Ghana, c. 1800 to Recent Times. Portsmouth, NH: Heinemann; 1996.
26. Luginaah I, Dakubo C. Consumption and impacts of local brewed alcohol (akpeteshie) in the Upper West Region of Ghana: a public health tragedy. Soc Sci Med 2003;57(9):1747-1760. [doi: 10.1016/s0277-9536(03)00014-5]
27. Wiggins S. Discursive Psychology: Theory, Method and Applications. Thousands Oaks, CA: Sage; 2017.
28. Thompson RGA, Amuzu EK. Culturally bound abusive expressions in Ghanaian social media. Lingua 2025;328:104056. [doi: 10.1016/j.lingua.2025.104056]
29. Pond P, Lewis J. Riots and Twitter: connective politics, social media and framing discourses in the digital public sphere. Inform Commun Soc 2019;22(2):213-231. [doi: 10.1080/1369118X.2017.1366539]
30. Franzke AS, Bechmann A, Ess CM, Zimmer M. Internet Research: Ethical Guidelines 3.0. Association of Internet Researchers. 2020. URL: https://aoir.org/reports/ethics3.pdf [accessed 2026-08-21]
31. Markham A, Buchanan E. Ethical decision-making and internet research: recommendations from the AoIR Ethics Working Committee (Version 2.0). Association of Internet Researchers. 2012. URL: https://aoir.org/reports/ethics2.pdf [accessed 2026-08-21]
32. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
33. Willig C. Introducing Qualitative Research in Psychology. 3rd Edition. London: Open University Press; 2013.
34. Manton E, Pennay A, Savic M. Public drinking, social connection and social capital: a qualitative study. Addict Res Theory 2013;22(3):218-228. [doi: 10.3109/16066359.2013.812202]
35. Measham F, Brain K. 'Binge' drinking, British alcohol policy and the new culture of intoxication. Crime Media Cult 2005;1(3):262-283. [doi: 10.1177/1741659005057641]
36. Jernigan DH. Alcohol and public health: failure and opportunity. Milbank Q 2023;101(S1):552-578 [FREE Full text] [doi: 10.1111/1468-0009.12631] [Medline: 37096612]
37. Heath DB. Culture and substance abuse. Psychiatr Clin North Am 2001;24(3):479-496. [doi: 10.1016/s0193-953x(05)70242-2] [Medline: 11593858]
38. de Bruijn A, Ferreira-Borges C, Engels R, Bhavsar M. Monitoring outdoor alcohol advertising in developing countries: findings of a pilot study in five African countries. Afr J Drug Alcohol Stud 2014;13(1):13-29 [FREE Full text]

XSL•FO
RenderX

39. Grodin EN, Baskerville WA, Meredith LR, Nieto S, Ray LA. Reward, relief, and habit drinking profiles in treatment seeking individuals with an AUD. Alcohol Alcohol 2024;59(3):agae032 [FREE Full text] [doi: 10.1093/alcalc/agae032] [Medline: 38725398]
40. Kady A, Grodin E, Ray L. Characterizing reward and relief/habit drinking profiles in a study of naltrexone, varenicline, and placebo. Alcohol Alcohol 2024;59(4):agae044 [FREE Full text] [doi: 10.1093/alcalc/agae044] [Medline: 38953743]
41. LaBrie JW, Hummer JF, Pedersen ER. Reasons for drinking click: validation of a new component of brief motivational enhancement interventions for college students. Psychol Addict Behav 2008;22(1):149. [Medline: 18298242]

Abbreviations

GhPE: Ghanaian Pidgin English
WHO: World Health Organization

Edited by J Yang; submitted 27.May.2026; peer-reviewed by J Guendouzi, YM Yusoff; comments to author 28.Jul.2026; revised version received 13.Aug.2026; accepted 20.Aug.2026; published 16.Sep.2026.

Please cite as:
Thompson RGA, Amuzu EK, Thompson T, Gyasi MK, Harris O, Santos GM, Ouner JJ
Social Media Constructions of Alcohol as Emotional Coping and Social Connection in Ghana: Qualitative Study
JMIR Infodemiology 2026;6:e102546
URL: https://infodemiology.jmir.org/2026/1/e102546
doi:10.2196/102546
PMID:

©Rachel GA Thompson, Evershed Kwasi Amuzu, Tricia Thompson, Mark Kwaku Gyasi, Orlando Harris, Glenn-Milo Santos, Jerry John Ouner. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 16.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Social Media Reactions to Sex Toy Criminalization in Ghana: Cross-Sectional Sentiment and Text Network Analysis

Anthony Senanu Agbeve[1], MPhil; Celestin Niyomugabo[2], MSc; Celestin Hakizimana[2], MSc; Daniel Yaw Fiaveh[1], PhD; Amanda Yad-El Ugboji[3], MPH; Caroline Pukall[4], PhD; Martina Anto-Ocrah[3], MPH, PhD

[1]Department of Sociology and Anthropology, University of Cape Coast, Faculty of Social Sciences, Arku Korsah Road, Cape Coast, Central, Ghana

[2]Vonsung, Kigali, Rwanda

[3]Department of Medicine, Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[4]Department of Psychology, Queen's University, Kingston, ON, Canada

Corresponding Author:
Anthony Senanu Agbeve, MPhil
Department of Sociology and Anthropology, University of Cape Coast, Faculty of Social Sciences, Arku Korsah Road, Cape Coast, Central, Ghana

Abstract

Background: Sexuality discourse in Ghana has become more divided, especially around gender and lesbian, gay, bisexual, transgender, and queer rights, leading to new antigay laws criminalizing lesbian, gay, bisexual, transgender, and queer activities, and other forms of sexual practices. One of the most controversial pieces of legislation, previously titled “The Promotion of Proper Human Sexual Rights and Family Values Bill,” proposed in 2021, sought to criminalize the use of sex toys, sparking opposition from leading political figures. The then minister of communication, Ursula Owusu-Ekuful, a staunch women’s rights advocate, publicly opposed the clause criminalizing sex toys, stating that such initiatives infringed on women’s sexual autonomy. Her stance sparked heated debates on several social media platforms, reflecting the broader tensions surrounding conversations about sexuality in Ghana. To better understand the public’s reactions to the minister’s counterproposal, we analyzed comments from various social media platforms.

Objective: Specifically, we (1) investigated public perceptions and sentiments regarding the criminalization of sex toys as expressed on social media, and (2) identified whether sentiments differed across social media platforms (eg, are comments on X [X Corp] more negative compared to Instagram or Facebook [Meta Platforms, Inc]?).

Methods: Using Python (Python Software Foundation) as a web scraping tool, we extracted public comments from Instagram, Facebook, X, and YouTube (YouTube, LLC). Comments were analyzed using natural language processing to categorize sentiments (positive, negative, and neutral, indicating support for opposition to the bill, support for introducing the bill, and no clear stance in either direction, respectively), and text network analysis was used to identify keywords and their relationships to central themes. We used conditional probability analyses to determine the likelihood that a given comment was classified as negative, positive, or neutral, depending on its source (Instagram, X, Facebook, or YouTube).

Results: A total of 951 comments were retrieved: 62.7% (n=596) from Instagram, 28.2% (n=268) from X, 6.2% (n=59) from Facebook, and 2.9% (n=28) from YouTube. Negative sentiments dominated (n=631, 66.4%), with only 13.3% (n=126) positive and 20.4% (n=196) neutral sentiments. Strongly disapproving language—some directed towards the minister—was common. This included terms such as “lesbian,” “disgrace,” and “disgusting.” X showed the highest proportion of negative sentiment (190/268, 70.9%), followed by Facebook (41/59, 69.5%), YouTube (18/28, 64.3%), and Instagram (382/596, 64.1%). Conditional probability analysis showed that Facebook and X had the highest likelihood of negative comments (0.7), and sentiment distribution differed significantly across platforms (P<.001). Text network analysis revealed recurring themes of morality, politics, stigma, and gender.

Conclusions: Our findings highlight deep-rooted cultural resistance to non-heteronormative sexual behaviors in Ghana and also suggest that women’s advocacy for sexual autonomy may elicit patriarchal backlash. However, emerging positive voices suggest gradual shifts toward recognizing sexual pleasure as a right, pointing to the need for comprehensive sexuality education that emphasizes positive sexuality and the multifaceted nature of the human sexual experience.

(JMIR Infodemiology 2026;6:e91422) doi:10.2196/91422

KEYWORDS

sexuality; sexual pleasure; sex toy; LGBTQ+; Ghana; lesbian, gay, bisexual, transgender, and queer

XSL•FO
RenderX

Introduction

The Ghanaian understanding of sex and sexuality is deeply embedded in cultural and religious doctrines that shape societal perceptions, norms, and expectations about sex and reproduction [1,2]. Ghanaian culture largely emphasizes heterosexual relations as the foundation of kinship and lineage continuity. Christianity, Islam, and Ghanaian spirituality (also known as Ghanaian traditional belief systems), the dominant religions or spiritualities in Ghanaian kinship, uphold conservative views on sex and sexuality, often condemning publicly nonheteronormative sex and sexualities (largely because of the cultural values of reproduction as a product of sex) and reinforcing "traditional" gender roles of male dominance and the important role of the penis [3,4]. Clergy and leaders from religious faiths and social institutions (schools, markets, workplaces, and other public spaces) work together in endorsing the moral codes of sex through their teachings and practices, including influencing not only personal beliefs but also courting politicians [5] and their political parties as well as national debates and policies on sexual and reproductive rights [6,7].

Ghana's lesbian, gay, bisexual, transgender, and queer (LGBTQ) community has gained visibility by aligning with global movements for LGBTQ rights, such as Rightify Ghana, LGBT Rights Ghana, the OHF initiative, and Key Watch Ghana, which lobby through social groups and research and occasionally engage with the media [8]. However, the growing visibility of these LGBTQ groups and their advocacy for sexual rights and freedom has ignited public backlash from diverse stakeholders, including traditional leaders, parents, the clergy, academics, human rights advocates, the media, and religious groups, including the Christian Council of Ghana, the Ghana Catholic Bishops Conference, and the Coalition of Muslim Organizations Ghana [9,10]. Ghana's Supreme Court has also emerged as a key actor in shaping legal jurisprudence on sexual practices and LGBTQ-related matters. In its July 24, 2024, ruling on the constitutionality of Section 104(1)(b) of the Criminal Offenses Act (Act 29), the Supreme Court reaffirmed that "unnatural" carnal knowledge, and the use of sex toys such as dildos and vibrators, was unlawful [11]. This shows that the contestation over LGBTQ rights and sexual practices in Ghana extends beyond activism into legal, cultural, and religious arenas, setting the stage for deeper debates on sexuality-related issues in the country.

On August 2, 2021, a bill titled "The Promotion of Proper Human Sexual Rights and Ghanaian Family Values Bill" was laid before Ghana's Parliament. The purpose of the bill was to expand and strengthen legal prohibitions on LGBTQ-related activities, including advocacy, promoting or funding LGBTQ activities, and public display of same-sex-related activities. Even though the bill was passed by the Ghanaian parliament under the revised title "The Human Sexual Rights and Family Values Bill, 2025," amid stiff opposition from feminist rights activists in and out of the Ghanaian parliament, the then president, Nana Addo Dankwa Akufo-Addo, did not assent to the bill. There were several oppositions even within the then ruling party (The New Patriotic Party) and the opposition party (The National Democratic Congress) to aspects of the bill, which some described as draconian. One member of the opposition, prior to its passage, Madam Ursula Owusu-Ekuful, former Minister of Communication and also a known women's rights activist, raised concerns about the criminalization of sex toys included in the bill. According to her, the proposed Clause 3(c) of the bill, which seeks to criminalize sex with an inanimate object, infringes on the sexual rights of many Ghanaians, particularly women and heterosexual couples, who may use sex toys, vibrators, and other sexual aids to enhance their sexual pleasure [12]. Owusu-Ekuful called for the deletion of clause 3(c) of the proposed bill, stating that:

> *This may create unintended consequences because sexual intercourse between a man and an inanimate object or between a woman and an inanimate object could necessarily include sexual intercourse with all manner of aids, including sex toys that couples use to enhance their sexual experience.*

Public reactions to her comments reflected a deep societal divide. While some applauded her audacity in openly discussing a topic often considered taboo, others criticized her position and accused her of endorsing practices perceived as inconsistent with Ghanaian cultural and religious values. Given that her comments were made within debates surrounding the anti-LGBTQ bill, they could be interpreted in multiple ways. Some sections of the public may have viewed her remarks as a defense of sexual autonomy more broadly, including the use of sex toys by heterosexual couples, adults, and individuals engaged in nonheteronormative sexual relationships and practices. Others may have understood them through prevailing anxieties about LGBTQ rights and broader public narratives that portray Western governments and international actors as exerting pressure on Ghana to recognize the rights and freedoms of sexual minorities. However, the controversy also became distinctly gendered because the comments were made by a prominent female politician and women's rights advocate who has long been accused of being a member of the queer community. Consequently, many public responses focused on questions of women's sexual agency, pleasure, and bodily autonomy, issues that remain highly contested within Ghanaian society.

The reactions to Owusu-Ekuful's comments also highlight the growing role of social media in shaping and contesting public discourses on sexuality-related issues, including sexual rights and bodily autonomy. Social media has transformed narratives around sex and sexual pleasure seeking with sex toys and other pleasure aids, with platforms such as Instagram, Facebook (Meta Platforms, Inc), and TikTok (ByteDance Ltd) increasingly serving as spaces for promoting and fostering open discussions about sexual autonomy and pleasure seeking [8,13-15]. The growing visibility of these platforms, alongside the expansion of the sex market to online spaces, signals a significant cultural shift as individuals are increasingly finding alternative avenues to discussing and learning about sex. This development reflects the gradual normalization of conversations around sexual autonomy, beyond traditional heteronormative expectations. Scholarship on sexuality has long highlighted how sexual practices and identities are shaped through social norms, power relations, and mechanisms of regulation [16-18]. Within African

XSL•FO
RenderX

contexts, however, scholars (such as Stella Nyanzi, Sylvia Tamale, and Akosua Adomako Ampofo) have drawn attention to the specific ways gender, sexuality, and pleasure are negotiated within local cultural and political realities. The shifting landscape of discourse and resistance observed in digital spaces aligns with these African feminist perspectives, which view women not as passive recipients of cultural norms but as active agents capable of making self-reflective choices about their bodies and desires, even within highly regulated environments [19,20]. This perspective is particularly relevant in conservative heteropatriarchal contexts, where women's expressions of sexual autonomy are frequently contested and subjected to social regulation. Although the National Democratic Congress government led by President John Dramani Mahama has yet to assent to the bill, which has been reintroduced in the Ninth Parliament of the Fourth Republic, it remains unclear whether the concerns previously raised by Ursula Owusu-Ekuful have been adequately addressed. Both the president and the speaker of parliament, Alban Kingsford Sumana Bagbin, have called for broader consultations and a reconsideration of the bill. From this perspective, we argue that the concerns raised by Ursula Owusu-Ekuful resonate with broader feminist critiques and sections of Ghanaian public opinion that question state regulation and surveillance of sexuality and reproductive rights. These concerns underscore the importance of wider public engagement and continued debate on issues of autonomy, bodily integrity, and the limits of state intervention in intimate life.

Therefore, using publicly available data, this study analyzed social media comments regarding Ursula Owusu-Ekuful's stance on the criminalization of sex toys. We recognize that debates surrounding sex toys in Ghana may evoke broader concerns relating to sexuality, LGBTQ rights, sexual autonomy, and the legitimacy of nonreproductive sexual practices. Accordingly, public responses to Owusu-Ekuful's comments may reflect attitudes towards a range of sexual identities and practices. Nevertheless, we focus particular attention on women because the public controversy has become strongly gendered, with reactions frequently centering on the appropriateness of women discussing, seeking, or exercising sexual pleasure. Against this backdrop, our two aims were to (1) investigate public perceptions and sentiments regarding the criminalization of sex toys as expressed on social media, and (2) identify whether sentiments differed across social media platforms (eg, X [X Corp], Instagram, and Facebook). Our overarching goal was to explore the narratives surrounding sexuality and bodily autonomy and to examine how these discourses shape sexual agency in contemporary Ghana, particularly for women.

Methods

Research Design

This study uses a computational content analysis design to examine publicly available social media discourse on sex toy legislation in Ghana. The dataset consists entirely of textual data drawn from user comments posted on Instagram, X, Facebook, and YouTube. To analyze these comments, we combined natural language processing techniques with qualitative interpretation. Computational methods were used to identify sentiment patterns, keyword associations, and variations across social media platforms, generating outputs such as sentiment frequencies, conditional probabilities, and χ^2 statistics. These measures were used to describe and compare patterns within the data rather than to establish causal relationships [21,22]. The interpretation of results remained grounded in the content, context, and meanings expressed in the comments. This approach aligns with emerging traditions in computational social science that combine machine learning and natural language processing techniques with qualitative modes of inquiry to analyze large volumes of textual data.

Data Collection

Data was collected from the following social media platforms: Instagram, X, Facebook, and YouTube, capturing online engagements between December 17-31, 2023, using Python (Python Software Foundation), a programming language for data analysis, as a web scraping tool. Publicly available information was extracted from these platforms that focused on the minister's speech regarding the criminalization of sex toys in Ghana. Data cleaning was performed on all scraped data and involved removing duplicates, empty comments, and stop words (such as oh, hello, or, and is). Natural language processing methods, specifically lemmatization, were used to reduce words in the text to their root forms. Lemmatization served a critical function by standardizing texts, improving text categorization precision, and aiding feature extraction and semantic analysis [23]. We also removed emojis from the dataset during the data-cleaning phase to prevent misclassification of sentiment scores. For example, comments with laughing emojis might be classified as positive when they are actually attached to negative or mocking comments.

Data Analysis

The data were analyzed using a three-step process:

1. The first step was sentiment analysis. Sentiment analysis was used to identify the emotional tone underlying comments on these social media platforms. This helped us organize and categorize public reactions to Ghana's sex toy legislation. We used a pipeline sentiment analysis method, which sequentially preprocesses the text, extracts key linguistic features, and applies machine-learning models to classify each comment. Drawing on data mining, machine learning, AI, and computational linguistics techniques [24], this approach enabled us to determine whether sentiments were positive, negative, or neutral. Positive sentiments reflected support for opposition to the bill, negative sentiments indicated support for introducing the bill, and neutral sentiments showed no clear stance in either direction.
2. Second, text network analysis was used to visualize the structure and relationships of emerging keywords based on their usage and co-occurrence. In the visualization, each keyword is represented by a node, and lines (edges) are used to connect nodes that co-occur in similar contexts. The size of each node indicates how frequently the corresponding word appears in comments on all platforms. Larger nodes represent words that appear more frequently,

XSL•FO
RenderX

indicating they are central themes in the comments. The visualizations generated by the text network analyses reveal clusters, or "communities," of words, where nodes within the same community are more densely connected but have fewer connections to nodes in other groups.

3. Lastly, we conducted a comparative analysis of the comments to better understand which social media platforms had the most polarizing sentiments. This involved examining patterns and differences in reactions across the various social media platforms to identify platform-specific trends. Conditional probability analyses were conducted to determine the likelihood that a given comment was classified as negative, positive, or neutral based on its source being Instagram, X, Facebook, or YouTube. We used the following formulas where:
   - $P(AorB)$ is the conditional probability of event A given that event B has occurred.
   - $P(A \cap B)$ is the joint probability of events A and B occurring simultaneously.
   - $P(B)$ is the probability of event B occurring.

Ethical Considerations

This study involved the passive analysis of publicly available social media comments and did not require direct interaction with users. Consequently, ethical approval and informed consent were not required [25]. To protect users' privacy and minimize the risk of harm, usernames and links to the original posts were not reported, and quotations were presented without identifying information [26]. The study adhered to the ethical guidelines of the Association of Internet Researchers for research involving internet-based data [27].

Results

Overview

In total, 951 comments were scraped from four data sources (ie, Instagram, X, Facebook, and YouTube) after data cleaning. The majority of the comments, 62.7% (n=596), came from Instagram, while 28.2% (n=268) came from X, 6.2% (n=59) from Facebook, and 2.9% (n=28) from YouTube.

Aim 1: Public Perceptions and Sentiments Regarding the Criminalization of Sex Toys

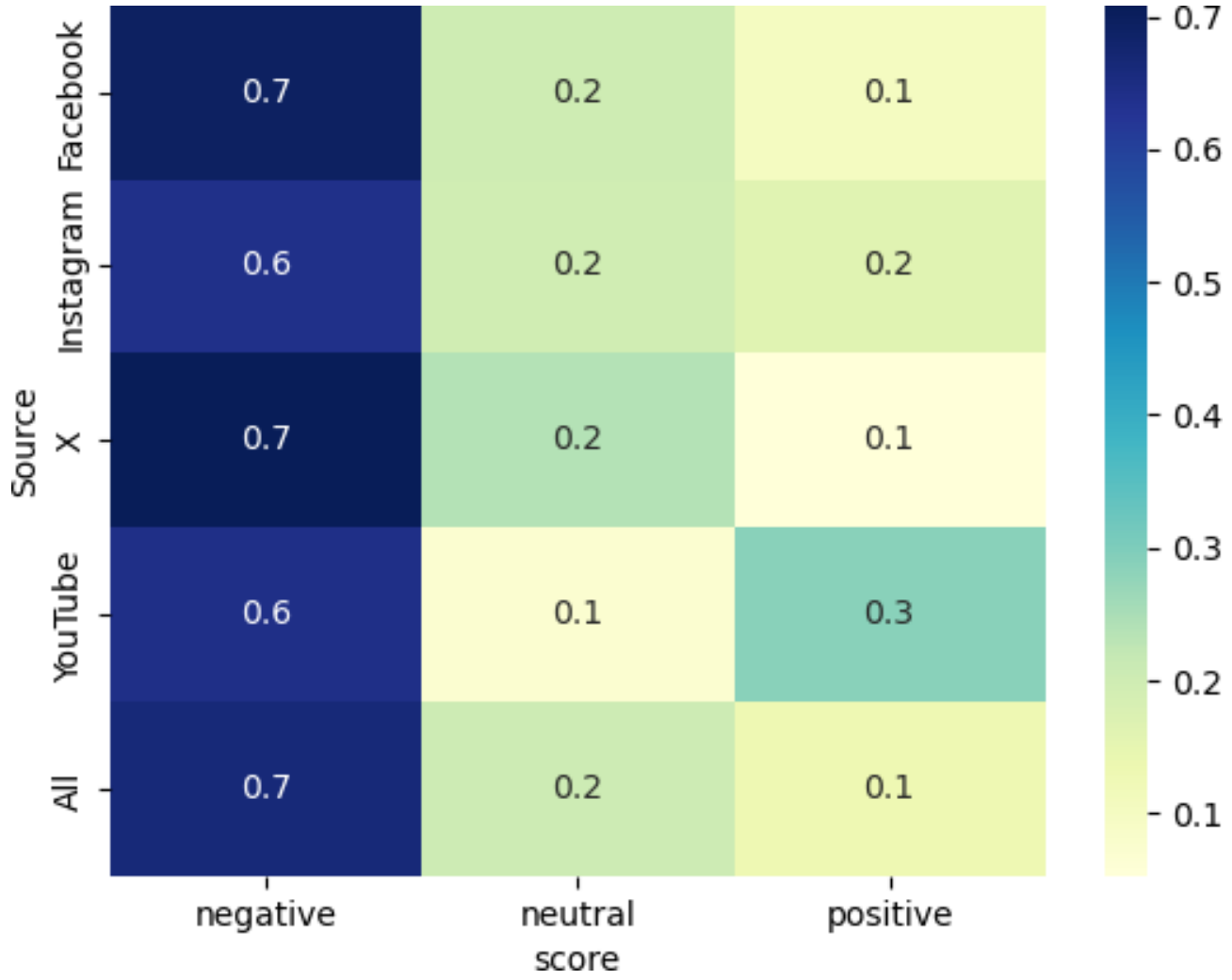
We analyzed social media comments to understand how members of the public responded to debates surrounding the proposed criminalization of sex toys. Specifically, we examined the distribution of positive, negative, and neutral sentiments expressed in relation to Ursula Owusu-Ekuful's comments and the broader issue of sex toy use. Representative excerpts are presented to illustrate the views, concerns, and positions reflected within each sentiment category.

Out of the 951 comments, the pipeline sentiment analysis classified 631 (66.4%) as negative sentiments, 126 (13.3%) were positive, and the remaining 194 comments (20.4%) were classified as neutral (Figure 1). Negative sentiments included keywords like "stupid," "useless," "foolish," "disgrace," and "disgusting." Phrases such as "She was a lesbian all through secondary school. I am not surprised. That's why she doesn't respect men…," "She is a disgrace to the Ghanaian culture," "this is exactly why our culture is losing its values. We should not be encouraging such things," and "people in leadership should promote morals, not defend things that bring shame to society" underscored the discontent and opposition to the minister's stance (Table 1).

Positive sentiments included words like "love," "joy," "better," and "enjoy." This was emphasized with phrases such as "Sex toys do not kill... they always bring joy to women who do not have boyfriends." and "She makes a very valid point. Heterosexual couples often employ aids to enhance their sexual experience; therefore, blanket criminalization of sex aids is rather misguided…," "adults should be free to make their own choices as long as nobody is harmed," "there is nothing wrong with couples using aids to improve their relationship," and "I support her on this because sex toys are for general use and not limited to only gays" (Table 1).

XSL•FO
RenderX

Figure 1. Probability distribution of scores by source.



XSL•FO
RenderX

Table . Sentiments and keywords.

Sentiment category	Comments	Keywords
Negative	• "I suspect this woman is going to be freaky in the bedroom a lot." • "She was a lesbian all through secondary school. I am not surprised. That's why she doesn't respect men…" • "If you give positions to gays and lesbians, these are the ideas they bring on board." • "Men can do the work, but she is here talking about sex toys. She is a disgrace to the Ghanaian culture." • "People in leadership should promote morals, not defend things that bring shame to society." • "This is exactly why our culture is losing its values. We should not be encouraging such things." • "People in leadership should promote morals, not defend things that bring shame to society."	Freaky, gay, lesbian, suspect, culture, respect, disgrace, morals, disgusting, and values
Neutral	• "So far as it doesn't harm anyone, why the need for such laws?" • "There are more pressing issues in this country than debating sex toys." • "It's better to focus on improving healthcare than criminalizing personal choices." • "Let's have a meaningful discussion about priorities in this country." • "Whether people use sex toys or not is their personal business." • "I do not see why Parliament should spend so much time debating this issue." • "Everyone is entitled to an opinion, but there are bigger challenges facing the country."	Harm, pressing issues, personal choices, personal business, law, and priorities
Positive	• "Sex toys do not kill... they always bring joy to women who do not have boyfriends." • "She has every right to use sex toys. Leave her alone…" • "She makes a very valid point. Heterosexual couples often employ aids to enhance their sexual experience; therefore, blanket criminalization of sex aids is rather misguided…" • "I support her on this because sex toys are for general use and not limited to only gays." • "Adults should be free to make their own choices as long as nobody is harmed." • "There is nothing wrong with couples using aids to improve their relationship." • "She is right. The government should focus on protecting rights rather than policing private lives."	Joy, sex toys, rights, support, autonomy, and freedom

Neutral comments reflected a segment of the population whose opinions were either ambivalent or not clearly articulated. This included phrases such as "So far as it doesn't harm anyone, why the need for such laws?", "There are more pressing issues in this country than debating sex toys," "I do not see why Parliament should spend so much time debating this issue," "Everyone is entitled to an opinion, but there are bigger challenges facing the country," and "It's better to focus on improving health care than criminalizing personal choices" (Table 1).

Overall, the sentiment analysis suggests that public reactions to Ursula Owusu-Ekuful's comments were predominantly negative, that is, they indicated support for the proposed criminalization of sex toys. While a smaller proportion of comments expressed support for her position or remained

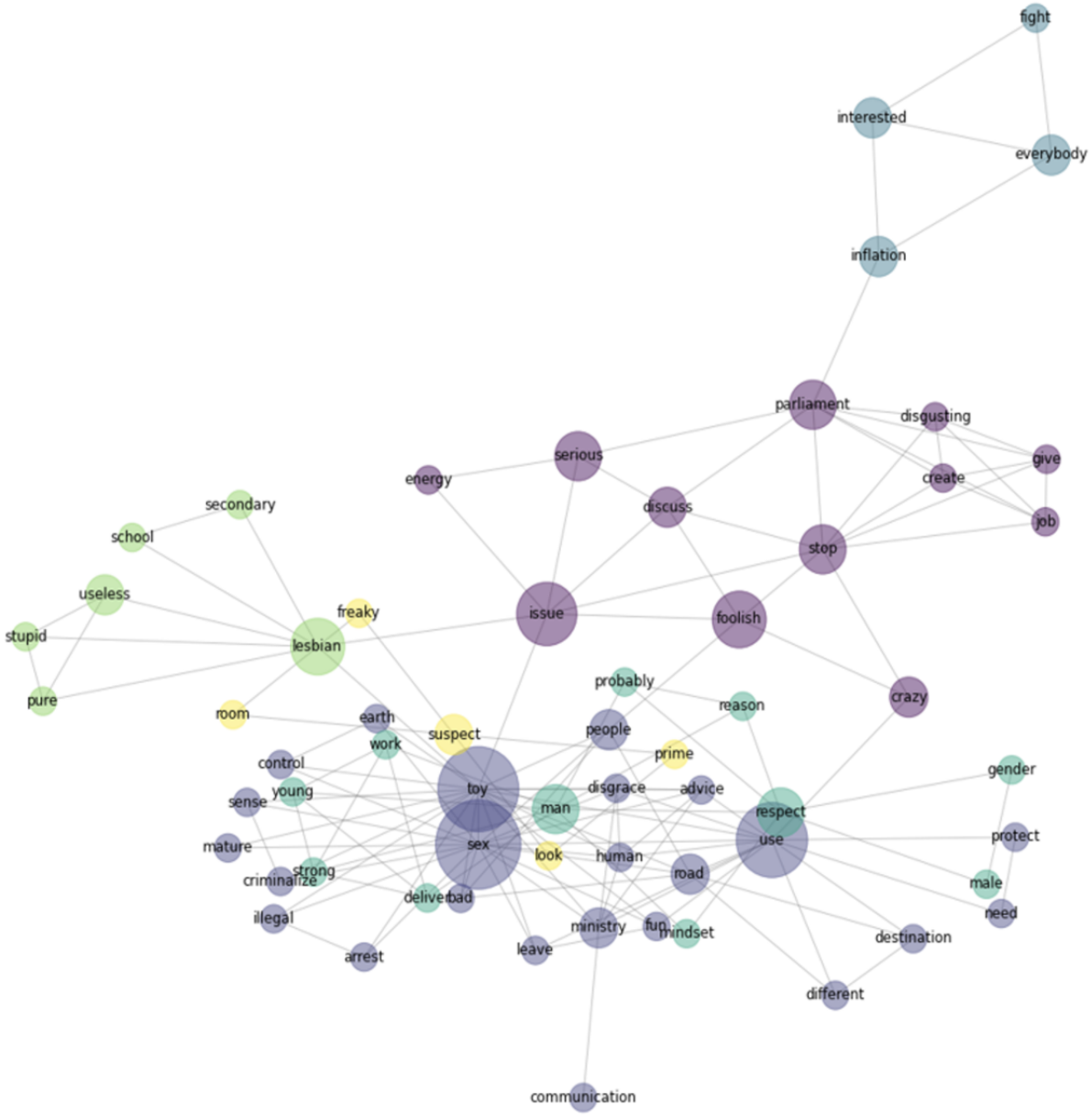
XSL•FO
RenderX

neutral, opposition to her stance was the dominant sentiment across the comments analyzed.

Text Network Analysis of Keywords

To complement the sentiment analysis, we conducted a text network analysis of keywords to identify dominant patterns of word co-occurrence and association within the social media discussions. This approach provides insight into how words co-occurred in the comments and highlights the different strands of discourse that shaped public reactions to the proposed criminalization of sex toys and the Minister's comments. In Figure 2, the results of the text network analyses are shown. Node sizes are determined based on the number of comments each word appears in. Larger nodes (words) indicate higher frequency or occurrence in the comments. These larger nodes may represent more significant topics or commonly discussed issues. The nodes were color-coded to differentiate the various communities identified in the analyses. For example, the purple community centers on words such as "issue," "foolish," "disgusting," "energy," and "jobs." These sentiments suggest that some individuals are unbothered by what people do in the private spaces of their bedrooms, as they do not consider sexuality-related issues a priority that warrants state intervention. The majority in this community are frustrated by what they perceive to be misplaced priorities by the government.

Figure 2. Text analysis of keywords based on community detection algorithms.



The mint green community features words such as "lesbian," "useless," and "stupid." This cluster mostly reflected negative attitudes towards the minister's statements. The sentiments expressed within this cluster suggest resistance to sexual diversity and opposition to the minister's position. The dark blue community, characterized by words such as "control," "illegal," "criminalize," and "disgrace," points to discussions focused on law, authority, and morality. Sentiments in this

XSL•FO
RenderX

community frame the discussion in terms of legality and morality, asserting that actions should be taken against sexual practices they see as incompatible with religious and cultural values. Unlike the previous clusters, the light green community connects words such as "respect," "mindset," and "gender," suggesting discussions on gender expectations and the importance of acknowledging and respecting sexual diversity.

Finally, the cyan community consists of words such as "inflation," "everybody," "fight," and "interested," which appear more isolated. Although this cluster does not directly engage in debates surrounding sexuality and governance, it suggests a broader concern for economic struggles and national priorities. The sentiments expressed by this group emphasize that economic hardship is a more pressing issue than the sexual affairs of Ghanaians. They are similar to the purple community in their concern about the government's misplaced priorities.

Overall, the text network analysis reveals that discussions surrounding sex toys and their proposed criminalization were shaped by multiple, interconnected concerns. While some communities expressed opposition rooted in moral, legal, and anti-LGBTQ sentiments, others emphasized sexual autonomy, gender equality, and respect for individual choice. Other discussions shifted attention away from sexuality altogether, highlighting economic hardship and questioning the prioritization of sexuality-related issues within the broader national context.

Aim 2: Differences in Sentiments Expressed Across Social Media Platforms

Sentiments expressed towards the proposed criminalization of sex toys varied across social media platforms. To examine these differences, conditional probability analysis and χ^2 tests were used to compare the distribution of positive, negative, and neutral sentiments across comments posted on X, Facebook, YouTube, and Instagram.

Conditional probability analysis showed that comments from Facebook and X had the highest likelihood of being negative (conditional probability of 0.7), followed by Instagram and YouTube (conditional probability of 0.6). Conversely, the most positive comments were posted on YouTube (conditional probability of 0.3), followed by Instagram (conditional probability of 0.2), while Facebook and X each had a conditional probability of 0.1. Neutral comments had a conditional probability of 0.2 for Facebook, Instagram, and X, compared with 0.1 for YouTube. The χ^2 test showed a statistically significant association (P<.001) between sentiment category (neutral, positive, or negative) and social media platform (Instagram, X, Facebook, or YouTube), indicating that the observed differences in sentiment distribution across platforms were unlikely to have occurred by chance.

The results show significant differences in sentiment across social media platforms. Comments posted on X and Facebook were more likely to express negative sentiments, while YouTube and Instagram contained relatively higher proportions of positive comments. These differences indicate that the distribution of sentiments varied by platform.

Discussion

Principal Findings

The study analyzed social media data to examine public sentiments surrounding state regulation and surveillance of sexuality in Ghana, with particular attention to the proposed criminalization of sex toys under the Anti-LGBTQ Bill. Drawing on public reactions to statements made by former minister of communications and member of parliament, Ursula Owusu-Ekuful, the study explored how social media users responded to provisions perceived as extending state control over intimate practices and sexual autonomy. In doing so, it examined the ways in which debates over sex toys became intertwined with broader concerns about bodily autonomy, privacy, and the governance of sexuality in contemporary Ghana.

The minister's vocal stance drew attention to the widespread policing and stigmatization of sexual agency and the broader pattern of how sexuality issues are tightly regulated in public discourse, particularly for Ghanaian women. From a Foucauldian perspective, these forms of regulation extend beyond formal laws or state control to the ways cultural, religious, and moral norms shape how women govern their own sexual conduct. While some women may be silenced or punished for asserting their sexual autonomy, others (such as Madam Ursula Owusu-Ekuful, who is highly educated and has legal training) may find spaces to navigate and even resist these restrictions because of their social positioning. In fact, studies show that Ghanaian women are increasingly turning to sex toys in response to sexual dissatisfaction, their partners' sexual weakness, infrequent sex, or a personal desire for additional pleasure [28]. For these women, owning and using a sex toy may be a form of sexual agency and a way of reclaiming bodily autonomy. This gendered resistance challenges the notion that sexual pleasure and fulfillment can be attained solely through penetrative sex. For married women, however, the opposite is observed. According to Fiaveh, married women are reserved about using sexual aids, even in the absence of a sexual partner, because they believe that such practices violate religious and cultural expectations [28]. This reflects how Foucault's idea of governmentality [29] operates through the internalization of dominant moral norms, leading individuals to regulate their own behavior in ways that align with socially accepted ideals of sexuality. To them, heterosexual, penetrative sex is the natural and morally sanctioned form of sexual expression. Thus, different women in Ghana face different kinds of vulnerabilities but also find different pathways through which they may explore and exercise agency in the face of societal constraints to assert and fulfill their sexual needs.

Our findings revealed predominantly negative attitudes and sentiments toward public discussions of sexuality, particularly regarding the use of sex toys and other pleasure-enhancing products on all social media platforms. The negative sentiments expressed were not merely expressions of personal discomfort but were deeply embedded in broader societal anxieties about shifting gender and sexual norms. Much of the backlash stems from growing fears over the visibility and activism of queer and

XSL•FO
RenderX

minority groups, whose demands for recognition and rights are perceived by many as a threat to the moral, religious, and cultural fabric of Ghanaian society [30,31]. However, beyond the general resistance to nonnormative sexualities, many of the negative sentiments revealed sexist undertones, targeting women who expressed or supported sexual autonomy. Women were often labeled as immoral, deviant, or "loose" simply for acknowledging or advocating for the right to sexual agency, whether through the use of sex toys or open discussion of desire and pleasure. Such views reflect a deep-seated discomfort with women asserting control over their sexual lives and reinforce the societal expectation that respectable womanhood is contingent upon sexual restraint, modesty, and subservience to male pleasure [28,32,33]. These notions highlight the ongoing policing of female sexuality and how patriarchal norms attempt to suppress women's sexual agency. It also shows how negative public reactions to women's sexual autonomy are not merely cultural expressions but active mechanisms of control that seek to limit women's bodily autonomy and right to pleasure.

The data also suggest that discussions about sex toys do not merely challenge acceptable forms of sexuality but also provoke discomfort because of their perceived association with same-sex sexual expression. In this context, sex toys were not viewed as legitimate pleasure-enhancing tools but as symbolic threats to the socially and religiously prescribed heteronormative order. Their connection to alternative sexual practices, especially those that center on female pleasure or queer sexualities, elicits strong resistance rooted in entrenched gender norms, conservative religious teachings, and moral expectations. Sexual expressions outside the framework of reproduction and male-dominated intimacy are often framed as immoral or unnatural. This reflects broader cultural discomfort with sexual diversity and sexual autonomy. Furthermore, the findings highlight the overconcentration of normativity, a phenomenon that feminist scholar Stella Nyanzi sharply critiques. Nyanzi argues that this obsession with what is deemed "normal" constrains public imagination and actively limits the recognition and legitimacy of other diverse forms of sexual expression [19]. In Ghanaian society, early socialization and dominant religious and cultural ideologies rigidly define what constitutes "normal" or "legitimate" sexuality. Society swiftly condemns deviations from dominant sexual norms, particularly sexualities that fall outside what is considered "normal" or "legitimate" [34,35]. At the same time, a culture of silence around sexuality limits opportunities for open discussion of sexual issues and diverse forms of sexual expression [36]. This normative framework creates significant barriers to open and inclusive conversations about sexuality in the public sphere [1]. Such cultural and moral policing reinforces the status quo, marginalizes those who do not conform, and curtails the possibilities for progressive discourse on sexual rights and autonomy.

The negative reactions to the minister's speech also underscore how women's sexual needs are routinely suppressed within patriarchal societies and deny them a voice and agency in sexual decision-making. The backlash, particularly the use of derogatory labels such as "freaky," "disgrace," and "lesbian," reveals not only discomfort with women asserting their sexual autonomy but also the societal mechanisms used to discipline and silence such vocal or assertive individuals. Women who openly discuss or advocate for sexual rights are often perceived as a threat to the masculine order and normative gender expectations, which dictate that women should be sexually passive, modest, and subordinate. Such labeling delegitimizes their sexual preferences and expressions and exposes them to heightened risks of verbal abuse, social ostracism, and physical violence. These derogatory terms and labels become ideological tools of control and deterrence, aimed at delegitimizing women's voices and reinforcing the idea that sexual autonomy or agency for women must be controlled. Nevertheless, in private spaces, some studies in Ghana have shown that heterosexual couples reportedly engage in diverse sexual practices, including the use of sexual aids such as sex toys to enhance their sexual experiences [28,32,33,37]. This underscores the double standards that characterize Ghanaian society, where public discourse fiercely defends cultural and moral norms by shaming women and others who advocate for sexual autonomy, while diverse sexual practices continue to exist, often away from public scrutiny. Despite widespread public opposition to LGBTQ issues and other nonheteronormative sexualities, such practices have long existed within African societies [38]. Yet, open discussion of these realities remains highly constrained. This disconnect reveals that, although Ghanaians may publicly express resentment toward nonnormative sexualities and pleasure-seeking practices, such behaviors are not absent in private life. However, the public backlash and associated moral condemnation may push women and other marginalized individuals into silence. This stance may stifle broader conversations about sexual rights and autonomy. The implications are profound, not just for women's sexual autonomy but also for their ability to define and experience pleasure on their terms.

In addition, comments such as "men can do the work" or claims that women who use sex toys "have no respect for men" reflect deeply entrenched cultural beliefs that prioritize male sexual performance and pleasure while relegating women's sexual needs as secondary or unnecessary [32]. These perspectives are indicative of a patriarchal, heteronormative sexual framework in which women's sexual needs are regarded as optional or contingent upon male involvement. This framing of female sexuality raises critical questions: are women only expected to participate in sex for reproduction or to satisfy men? Is sexual pleasure viewed as primarily for men? These presumptions contribute to a socio-cultural environment that actively denies women the right to pursue their sexual inclinations and make decisions about their bodies. The consequences will be far-reaching as women's sexual autonomy is systematically restricted, and their agency is undermined when male pleasure is normalized as primary and female pleasure is devalued or stigmatized. This cultural logic not only restricts women's capacity to explore their sexual preferences but also undermines the legitimacy of women's sexual rights in a broader context. The normalization of male-centered sexual authority perpetuates power imbalances, which may even create environments that are conducive to gender-based violence or coercion. By framing sex toys as a rejection of male authority rather than as tools for individual or shared pleasure, such discourse reinforces the idea that women's sexuality must remain tied to male control.

XSL•FO
RenderX

Ultimately, these dynamics reproduce a social order in which women's sexual needs are not only subordinated but also surveilled, restricted, and silenced.

Despite the associated negativities regarding female sexual expressions and autonomy, we identified some positive comments that countered the overly negative sentiments expressed about the sex toy discourse. The recognition that sex toys are intended to improve and enhance the sexual experiences of individuals, including those in heterosexual relationships, points to a gradual departure from the conventional normative ideals. For instance, comments such as "…heterosexual couples often employ aids to enhance their sexual experience; therefore, blanket criminalization of sex aids is rather misguided…" illustrate how heteronormative sexual norms are being contested. Narratives of this nature challenge conservative ideologies that seek to confine sexual expressions within the normative framework of penovaginal sex. This positive outlook signals a gradual shift away from the rigid structures that have long marginalized female sexual agency. In a cultural context where patriarchal and religious norms tightly regulate sexuality, there are dissenting voices advocating autonomy and the right to make self-reflective choices. This suggests that with time and continued advocacy, women and other marginalized groups will be empowered to resist restrictive cultural norms and narratives shaping sexuality discourses. This could, to a large extent, alter the cultural landscape and public perceptions of sexual rights and autonomy in Ghana.

Strengths and Limitations

To the best of our knowledge, this is the first and only study to analyze social media comments on sex toy legislation in Ghana and to do so using AI. Our study integrates insights from media studies, law, gender and sexuality studies, and computational social science to offer a thorough analysis of public discourse on this emerging issue. In doing so, we not only pioneer a forward-looking methodology that anticipates how intelligent technologies can be used to better understand social media debates in Ghana and beyond, but we also break new ground in the intersection of legal media analysis and AI. This study has several limitations. Although the analysis focuses on public discourses surrounding women's sexual autonomy and pleasure, the gender identities of social media commenters could not be verified. Consequently, interpretations are based on the content and context of comments, as well as the gendered debates prompted by Ursula Owusu-Ekuful's remarks. Furthermore, because the commenter's gender could not be established, views interpreted as supporting or opposing women's sexual autonomy may have been expressed by men rather than women. The findings should therefore be understood as reflecting broader public discourses on women's sexuality rather than the perspectives of women alone. This study also excludes the perspectives of offline, nonsocial-media users, potentially overlooking significant viewpoints held by less-connected or rural populations. Prior research has shown that online public discourse can diverge substantially from offline sentiment due to representativeness biases and echo chamber effects, in which people are repeatedly exposed to information and opinions that closely align with their views, reinforcing those beliefs and shielding them from opposing perspectives. This often happens in algorithm-driven platforms that prioritize familiar content [39]. In order to facilitate a more holistic understanding of the Ghanaian public opinion on sex toy legislation, future studies should incorporate offline surveys or community interviews to compare offline views with online sentiment.

Conclusion

This study reveals entrenched societal discomfort and cultural unease over nonheteronormative sexualities and expressions in Ghana. We found that cultural and religious norms largely shaped public discussions, especially those on social media. The opposition to the sex toy debate reflects not merely moral condemnation but also the aversion to public display and support for sexual expressions that contest the established social order. We infer from the narratives and the analysis that the use of sex toys in the context of solo or partnered sex may not be the primary source of contention, but the public clamor for its decriminalization and the fact that a woman was making this call is what appears to invite greater scrutiny. This stems from perceptions regarding marital fidelity, where the use of sex toys could be considered as undermining male sexual hegemony or dislocating masculinity. The sentiments did not differ from these notions, as many expressed fear, ridicule, or moral outrage at the thought of women seeking pleasure outside the direct involvement of men, reinforcing the belief that sexual fulfillment should remain tied to male authority and control. This reflects broader patriarchal concerns about men maintaining power within intimate relationships and preserving traditional gender hierarchies, which have been part of the major concerns of both feminists and other activist groups. While a section of Ghanaians expressed optimism in an attempt to challenge these notions and cultural scripting, the overwhelmingly negative public backlash reveals the powerful and resilient nature of structural constraints. However, the positive sentiments identified in this study indicate a gradual shift toward recognizing individuals' rights to make autonomous decisions regarding their sexual well-being. These emerging perspectives, although limited, suggest the possibility of more inclusive discussions about sexuality issues in Ghana, and this requires both cultural and structural shifts through inclusive sexuality education that recognizes sexual pleasure as a right, thereby challenging the conception that African societies limit women's sexual options. Also, the findings point to the need for sexuality education and sexual and reproductive health policies to move beyond a primary focus on risk and reproduction and engage more directly with issues of sexual autonomy, pleasure, and bodily agency in culturally appropriate ways. Indeed, we agree with views such as those of Daniel Fiaveh, especially in his 2019 study on sex toys among women and other gender activists who suggest that women, like their male counterparts, exercise sexual choices in the "bedroom" so long as sexual pleasure is a "thing" they want to seek. So, so long as Madam Ursula Owusu, the Ghanaian women's rights activist, is concerned, anti-LGBTQ legislation is not only targeted at gender and sexual minorities but also a threat to women's sexual autonomy, and that should not be countenanced. We agree with Ursula's feminist stance in that regard.

XSL•FO
RenderX

Acknowledgments

The authors acknowledge the role of Nabeeha Affan in the initial literature review for the paper. The authors cite remarks made by Hon. Ursula Owusu-Ekuful during the parliamentary debate on Ghana's anti-gay bill. These remarks inform the discussion but do not imply her endorsement of the article's arguments; the authors are solely responsible for the analysis and conclusions.

Funding

Funding support for this work was provided by the final author, MA-O, via the University of Pittsburgh Department of Medicine, Division of General Internal Medicine.

Data Availability

The data that support the findings of this study are available from the corresponding author upon request.

Authors' Contributions

ASA, CN, CH, DYF, and MA-O made substantial contributions to conception and design, and acquisition of data, and analysis and interpretation of data. ASA, DYF, AYU, CP, and MA-O drafted the article or revised it critically for important intellectual content. ASA, CN, CH, DYF, AYU, CP, and MO-A had granted final approval of the version to be published.

Conflicts of Interest

None declared.

References

1. Agbeve AS, Fiaveh DY, Anto-Ocrah M. Parent-adolescent sexuality communication in the African context: a scoping review of the literature. Sex Med Rev 2022 Oct;10(4):554-566. [doi: 10.1016/j.sxmr.2022.04.001] [Medline: 36210094]
2. Agbeve AS, Fiaveh DY, Anto-Ocrah M. A Qualitative assessment of adolescent-parent sex talk in Ghana. Afr J Reprod Health 2022 Dec;26(12s):146-160. [doi: 10.29063/ajrh2022/v26i12s.16] [Medline: 37585170]
3. Anarfi JK, Owusu AY. The making of a sexual being in Ghana: the state, religion and the influence of society as agents of sexual socialization. Sex Cult 2011 Mar;15(1):1-18. [doi: 10.1007/s12119-010-9078-6]
4. Anarfi JK, Gyasi-Gyamerah AA. Religiosity and attitudes toward homosexuality: views of Ghanaian university students. Res Soc Sci Study Relig 2014;25:173-201. [doi: 10.1163/9789004272385_011]
5. Okyerefo MPK, Fiaveh DY, Asante KT. Religion as a tool in strengthening the democratic process in Ghana. J Afr Stud Dev 2011;3(6):124-130 [FREE Full text]
6. Amo-Adjei J. Resistances to the implementation of comprehensive sexuality education curriculum in Ghana's educational institutions. Int J Adolesc Youth 2024 Dec 31;29(1). [doi: 10.1080/02673843.2024.2398038]
7. Fiaveh DY. Sexuality, gender, and spiritual interventions: reflections on sexual world-making in Ghana. J Africana Relig 2025 Sep 26;13(1):134-140. [doi: 10.5325/jafrireli.13.1.0134]
8. Asante GA, Mohammed WF, Appiah-Kubi AB. LGBT activism, social media, and the politics of queer visibility in ghana. In: The Oxford Handbook of Media and Social Justice 2024:268. [doi: 10.1093/oxfordhb/9780197744345.013.29]
9. Catholic Bishops back anti-gay bill but want rights of homosexuals respected. Citi Newsroom. 2023. URL: https://www.citinewsroom.com/2023/11/catholic-bishops-back-anti-gay-bill-but-want-rights-of-homosexuals-respected/ [accessed 2026-07-24]
10. Fiaveh DY. LGBQ+ in Ghana. Sociolinguist Stud 2023 Aug 7;17(1–3):21-43. [doi: 10.1558/sols.24050]
11. Using dildos and vibrators to please yourself, anal sex and other forms of unnatural carnal knowledge is still unlawful - Supreme Court. Graphic Online. 2024. URL: https://www.graphic.com.gh/news/general-news/ghana-news-unnatural-carnal-knowledge-unlawful-anal-sex-use-of-vibrators-others-still-offences-supreme-court.html [accessed 2026-07-29]
12. Anti-LGBTQ bill: We shouldn't criminalise sex toys – Ursula Owusu-Ekuful. MyJoyOnline. 2023. URL: https://www.myjoyonline.com/anti-lgbt-bill-we-shouldnt-criminalise-sex-toys-ursula-owusu-ekuful/#google_vignette [accessed 2026-07-29]
13. Döring N. Sex education on social media. In: Encyclopedia of Sexuality and Gender 2021:1-12. [doi: 10.1007/978-3-319-59531-3_64-1]
14. Suenzo FN, Pinch A, Cruz IF, Liang CA, Ross Arguedas A. Knowledge and expertise in the digital age: how people engage with sex advice on TikTok. Int J Commun 2026;20:523-544. [doi: 10.65476/rqpmyd95]
15. Manduley AE, Mertens A, Plante I, Sultana A. The role of social media in sex education: dispatches from queer, trans, and racialized communities. Fem Psychol 2018 Feb;28(1):152-170. [doi: 10.1177/0959353517717751]

XSL•FO
RenderX

16. Foucault M. The History of Sexuality: Pantheon Books; 1978, Vol. 1. URL: https://monoskop.org/images/4/40/Foucault_Michel_The_History_of_Sexuality_1_An_Introduction.pdf [accessed 2026-07-24]
17. Butler J. Subversive bodily acts. In: Gender Trouble: Feminism and the Subversion of Identity: Routledge; 2011:143-240. [doi: 10.4324/9780203824979-8]
18. Butler J. Bodies That Matter: On the Discursive Limits of “sex": Routledge; 1993:1-27. [doi: 10.4324/9780203760079]
19. Nyanzi S. Queering queer Africa. In: Reclaiming Afrikan: Queer Perspectives on Sexual and Gender Identities: Modjaji Books; 2014:61-66 URL: https://books.google.es/books?id=enchBQAAQBAJ&pg=PA61&hl=es&source=gbs_toc_r&cad=2#v=onepage&q&f=false [accessed 2026-07-24]
20. Tamale S. Exploring the contours of African sexualities: religion, law and power. Afr Hum Rights Law J 2014;14(1):150-177 [FREE Full text] [doi: 10.2307/j.ctvh8r1ds.6]
21. Munnes S, Harsch C, Knobloch M, Vogel JS, Hipp L, Schilling E. Examining sentiment in complex texts. A comparison of different computational approaches. Front Big Data 2022;5:886362. [doi: 10.3389/fdata.2022.886362] [Medline: 35600329]
22. Nielbo KL, Karsdorp F, Wevers M, et al. Quantitative text analysis. Nat Rev Methods Primers 2024;4(1):25. [doi: 10.1038/s43586-024-00302-w]
23. Bird S, Klein E, Loper E. Natural Language Processing with Python: O’Reilly Media; 2009. URL: https://doi.org/10.1162/coli_r_00022 [accessed 2026-07-24] [doi: 10.1162/coli_r_00022]
24. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 2014 Dec;5(4):1093-1113. [doi: 10.1016/j.asej.2014.04.011]
25. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105. [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
26. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002;4(2):E13. [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
27. Franzke AS, Bechmann A, Ess CM, Zimmer M. Internet research: ethical guidelines 30. Association of Internet Researchers. 2020. URL: https://aoir.org/reports/ethics3.pdf [accessed 2026-07-24]
28. Fiaveh DY. Phallocentricism, female penile choices, and the use of sex toys in Ghana. Sexualities 2019 Oct;22(7-8):1127-1144. [doi: 10.1177/1363460718781975]
29. Finnane M. In: Burchell G, Gordon C, Miller P, editors. The Foucault Effect: Studies in Governmentality with Two Lectures by and an Interview with Michel Foucault: University of Chicago Press & Hemel Hempstead; 1992, Vol. 7:309-312. [doi: 10.1017/S0829320100002507]
30. Asante GA. Anti-LGBT violence and the ambivalent (colonial) discourses of Ghanaian Pentecostalist-Charismatic church leaders. Howard J Commun 2020 Jan 1;31(1):20-34. [doi: 10.1080/10646175.2019.1590255]
31. Otu KE. Queer slacktivism as silent activism? the contested politics of queer subjectivities on GhanaWeb. Sexualities 2021 Feb;24(1-2):46-66. [doi: 10.1177/1363460719893620]
32. Fiaveh DY. Masculinity, male sexual virility, and use of aphrodisiacs in Ghana. J Mens Stud 2020 Jun;28(2):165-182. [doi: 10.1177/1060826519887510]
33. Fiaveh DY, Okyerefo MP. Femininity, sexual positions and choice. Sexualities 2019 Feb;22(1-2):131-147. [doi: 10.1177/1363460716677281]
34. Fiaveh DY. Gendered scripts and young adults’ sexual practices on a first date in urban Ghana. In: Donkor M, Gooden A, editors. Gender and Sexuality in Ghanaian Societies: Lexington Books; 2021:75-100. [doi: 10.5040/9781666989670]
35. Dery I, Fiaveh DY, Apusigah AA. “You Cannot Be Like That Here”: discourses of sexual identities among urban Ghanaian families. Gend Issues 2019 Dec;36(4):342-356. [doi: 10.1007/s12147-019-09230-6]
36. Fiaveh DY, Mensah E. Gender and sexuality in African discourses. Socioling Stud 2023 Aug 7;17(1-3):7-19. [doi: 10.1558/sols.24323]
37. Fiaveh DY. Scoping review of discourses on masturbation in Africa: an Afrofeminist analysis. J Gend Stud 2024 Nov 16;33(8):1021-1038. [doi: 10.1080/09589236.2023.2300420]
38. Signorini I. Ag nwole agyal : the marriage between two persons of the same sex among the Nzema of southwestern Ghana. J Africanistes 1973;43(2):221-234. [doi: 10.3406/jafr.1973.1713]
39. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 2;118(9):e2023301118. [doi: 10.1073/pnas.2023301118] [Medline: 33622786]

Abbreviations

LGBTQ: lesbian, gay, bisexual, transgender, and queer

XSL•FO
RenderX

Edited by T Nguyen; submitted 14.Jan.2026; peer-reviewed by B Ocran, S Buckler; revised version received 02.Jul.2026; accepted 05.Jul.2026; published 11.Aug.2026.

Please cite as:
Agbeve AS, Niyomugabo C, Hakizimana C, Fiaveh DY, Ugboji AYE, Pukall C, Anto-Ocrah M
Social Media Reactions to Sex Toy Criminalization in Ghana: Cross-Sectional Sentiment and Text Network Analysis
JMIR Infodemiology 2026;6:e91422
URL: https://infodemiology.jmir.org/2026/1/e91422
doi:10.2196/91422

© Anthony Senanu Agbeve, Celestin Niyomugabo, Celestin Hakizimana, Daniel Yaw Fiaveh, Amanda Yad-El Ugboji, Caroline Pukall, Martina Anto-Ocrah. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 11.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Social Media Perspectives on a Future HIV Vaccine: Mixed Methods Analysis

Megan A Rabin[1], MPH; Sarah Penuela-Wermers[2], MPH; Neil K R Sehgal[3,4], PhD, MPH; Teniola I Egbe[1], MPH, MBE; Criswell L M Lavery[5], MPH; Sharath Chandra Guntuku[3,4], PhD; Alison M Buttenheim[2,4], MBA, PhD

[1]Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Blockley Hall, 423 Guardian Drive, Philadelphia, PA, United States

[2]Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[3]Computer and Information Science Department, University of Pennsylvania, Philadelphia, PA, United States

[4]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States

[5]Philadelphia VA Medical Center, Philadelphia, PA, United States

Corresponding Author:
Megan A Rabin, MPH
Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Blockley Hall, 423 Guardian Drive, Philadelphia, PA, United States

Abstract

Background: As the prospect of an HIV vaccine nears reality, understanding public discourse around the vaccine is essential for informing communication strategies and addressing misinformation. Social media platforms are influential spaces where public narratives form, yet little research has examined discourse around an HIV vaccine, especially on TikTok.

Objective: This study aims to compare and characterize public discourse about a future HIV vaccine across Twitter (subsequently rebranded X) and TikTok, identifying prevailing themes, sentiments, and rhetorical strategies to inform public health communication.

Methods: From over 400,000 tweets and 65,000 TikTok comments, we analyzed the 1000 most-liked posts on each platform using natural language processing and coded the top 500 most-liked posts for rhetorical strategies, sentiment, and themes.

Results: Our findings reveal expressions of hope and trust in science on both platforms, as well as concerns about institutional corruption and conspiracy theories, such as the belief that the HIV vaccine responds to harm caused by the COVID-19 vaccine. Tweets tended to be more linguistically complex and yielded richer insights, while TikTok comments were shorter and more difficult to interpret without video context. Key rhetorical strategies included conspiracy theories, post hoc reasoning, and emotional appeals.

Conclusions: This study underscores the need for platform-specific communication strategies to address misinformation and build public trust. The findings offer timely insights into emerging HIV vaccine discourse and highlight actionable opportunities for public health stakeholders to build trust and combat misinformation in advance of the vaccine rollout.

(JMIR Infodemiology 2026;6:e82917) doi:10.2196/82917

KEYWORDS

HIV vaccines; vaccine hesitancy; social media; misinformation; health communication; public opinion; conspiracy theories; attitude to health; qualitative research; natural language processing

Introduction

Worldwide, “infodemics,” or an oversaturation of information regardless of accuracy, have been well-documented as a threat to public health [1]. In particular, the spread of vaccine-related misinformation has emerged as a barrier to vaccine demand, acceptance, and uptake [2-4]. For example, study participants in the United States and the United Kingdom who were exposed to online COVID-19 misinformation reported decreased intention to receive a COVID-19 vaccination, including a 6 percentage point decrease among those who had previously reported they would “definitely” get the vaccine relative to those in the control group [2].

Given the volume of influential narratives and misinformation circulating on social media, it is valuable for scientists, government bodies, and public health professionals to understand public attitudes, beliefs, and concerns regarding existing and potential vaccines [1,2,4-7]. Prior studies have sought to characterize prevailing vaccine sentiment in online discourse about COVID-19, human papillomavirus, influenza, HIV, and measles vaccines, across multiple social media platforms [5-8].

XSL•FO
RenderX

Notably sparse in the online vaccine discourse literature is the HIV vaccine. This emerging vaccine, which has been the subject of recent online discussion, has over 20 clinical trials currently underway [9]. A future HIV vaccine has the potential to help end the epidemic in highly impacted regions such as Sub-Saharan Africa, home to two-thirds of the world's population living with HIV [10]. The vaccine could also help prevent HIV in groups at the highest risk of infection, including injection drug users, sex workers, men who have sex with men, and vulnerable groups like young people and women [11].

Various studies have explored social media discourse around other aspects of HIV, namely, pre-exposure prophylaxis (PrEP) across platforms including Instagram, TikTok, Twitter (subsequently rebranded X), and Reddit [12-15]. Social media has been a place where individuals seek and share information regarding PrEP, including costs, availability, and resources [12,13,15]. Post-COVID-19, social media revealed how the pandemic disrupted various HIV prevention services, especially for racial and sexual minority groups [15]. However, social media has also served to perpetuate misinformation related to HIV [14,15]. For example, a study examining TikTok found themes about PrEP encouraging risky sex, reiterating outdated ideas about gay men as "unsafe," and discussing HIV as a disease with poor prospects for individuals once they contract it [14].

On social media, vaccine misinformation can be especially pervasive, as messaging and communities devoted to antivaccine beliefs can rapidly grow and spread across geographic and cultural boundaries [16]. One study tracking misinformation during the COVID-19 infodemic categorized upwards of 2000 reports of "rumors" across media and social media sites, as well as another 200 reports of conspiracy theories or stigma [1]. Another study identified the increased polarization of vaccination content on Facebook, resulting in "echo chambers" in which a user might be surrounded by exclusively positive or negative vaccine attitudes depending on their interactions [17].

Despite the growing body of research studying social media attitudes regarding vaccines, only 1 study of which we are aware has specifically examined attitudes on social media regarding a future HIV vaccine, and this study excluded the social media platform TikTok [7]. Given TikTok's rapid rise as the fastest-growing social media app in the world, with 1.5 billion active monthly users who spend an average of 52 minutes per day on the app, this is an important gap [18].

Nested within a larger trial investigating the use of messaging to "inoculate" adolescent girls and young women from HIV vaccine misinformation, this study seeks to illuminate the information ecology and characterize the discourse surrounding a future HIV vaccine across the platforms Twitter and TikTok. Both platforms offer fast-moving, public-facing communication among users, with algorithms predicting what users will want to see in an endless stream of content [19,20]. To compare the nature of the discourse across these platforms, we selected comparable, short-text posts available on the platforms: tweets on Twitter, which have a limit of 280 characters, and comments on TikTok videos, which online reports suggest had a limit of 150 characters during the time of the scrape [21-23]. We aim to highlight key insights and opportunities for action *before* the HIV vaccine is approved and available, in the hope of supporting future public health and demand creation communications.

Methods

Data

We scraped tweets and TikTok video comments using platform-specific protocols, using slightly different search terms due to differences in audience, purpose, and tone of content across the 2 platforms. In March 2023, we used the Twitter application programming interface and conducted a keyword-based search of relevant terms ("hiv vaccine," "hiv vax," "hivvaccine," and "aids vaccine") to collect tweets posted between January 2022 and March 2023 (a window during which 3 National Institutes of Health–funded HIV vaccine trials were in progress) [24].

Comments on TikTok videos were collected via a 2-step process. We first searched TikTok videos on October 27, 2023, using the terms "HIV Vaccine," "AIDS Vaccine," and "HIV Jab," yielding a total of 777 videos. The TikTok-provided research application programming interface was used to retrieve all comments posted to those videos on or before October 31, 2023.

Quantitative Analysis

We used the Python-based (Python Software Foundation) natural language processing software "TextAnalyzer" for lexical analysis of a subset of the high-engagement tweets and comments comprising the top-liked 1000 TikToks and 1000 tweets [25]. Posts in this dataset had a minimum of 75 likes for a given TikTok comment, or 77 likes for a tweet. TextAnalyzer provides scores on the following metrics for each unit of text input: *Flesch-Kincaid Grade Level* for sentence complexity, *Emotionality*, *Emotion* for a variety of different emotions, *Positive* and *Negative Sentiment*, for the degree of positive versus negative emotions separately, and *Emotional Valence* for a composite score indicating a positive, negative, or neutral attitude [26-30]. To understand differences in the discourse about the HIV vaccine on TikTok versus Twitter, we compared these metrics across the 2 platforms using the Welch 2-sample 2-tailed *t* tests for mean scores (conducted using R version 4.5.1; R Foundation for Statistical Computing).

Qualitative Analysis

We coded a further subset of the top 500 most highly-engaged posts by like count, in which TikTok comments had 206 or more likes and tweets had 173 or more likes. The coders first familiarized themselves with the data and then developed a codebook including both *a priori* and emergent codes related to comment or tweet content. Based on the goals of the parent study, *a priori* codes included rhetorical and persuasion strategies from the social media and communications literatures; the veracity or validity (ie, "truth") of any claims or information presented about the HIV vaccine, and the sentiment of the comment or tweet, that is, whether it was generally provaccine, antivaccine, or neutral. *A priori* rhetorical strategies consisted of the invocation of conspiracy theories, humor, fake experts, or highly emotional language (pathos), as well as using only selective data (cherry picking), attacking the person making an

XSL•FO
RenderX

argument (ad hominem), and misattributing events happening in sequence to being cause-and-effect (post hoc propter hoc) [31,32]. We also included the strategy of "clickbait" after seeing how many posts used inflammatory and exaggerated language to fuel their argument, and "red herring" for posters who made irrelevant comments to distract the audience from the issue at hand [33]. Other sections of the codebook included "vaccine sentiment" to capture attitudes toward vaccines, "topic" for broader categories such as HIV vaccine- or COVID-19–related content, and "content themes" for more specific ideas expressed about the HIV vaccine and other emerging ideas (Multimedia Appendix 1).

This first codebook was applied to a 20% sample (100 TikTok comments and 100 tweets) of the dataset by 2 coders (MAR and SPW) to assess codebook appropriateness and establish intercoder reliability. After the initial sample coding, the coders met to iteratively revise the draft codebook, coding an additional 10% sample of the final dataset (50 TikTok comments and 50 tweets) before finalizing the codebook. Once this final codebook was agreed upon by both coders, it was applied to the entire dataset of the most-liked 500 comments and 500 tweets. The coders checked in multiple times throughout this process to resolve uncertainties in coding and ensure that they were in agreement about code applications.

After coding was complete, we analyzed a subset of 232 posts (215 tweets and 17 TikTok comments) deemed the "high-relevance" dataset because we could confidently code them to the topic area of "HIV Vaccine" across the thematic areas and rhetorical strategies outlined in the codebook. These posts were those that referenced an HIV vaccine either directly or indirectly and left minimal ambiguity or question of whether they referenced another topic of vaccine. For example, a tweet like "This has been in the works for ages but the covid stuff allowed them to accelerate mRNA research" was determined to be referencing an mRNA HIV vaccine. However, a comment reading "They've reached a new level of crazy…it's actually dangerous that they're spreading stuff like this" may have been in response to HIV vaccine discourse, but it also could have been referencing any number of other related topics such as HIV, COVID-19, or other vaccines—themes that came up in other excluded posts. Within this dataset, we identified the most prevalent topics referenced and rhetorical strategies used, and used inductive analysis to describe the themes that emerged in these posts.

Ethical Considerations

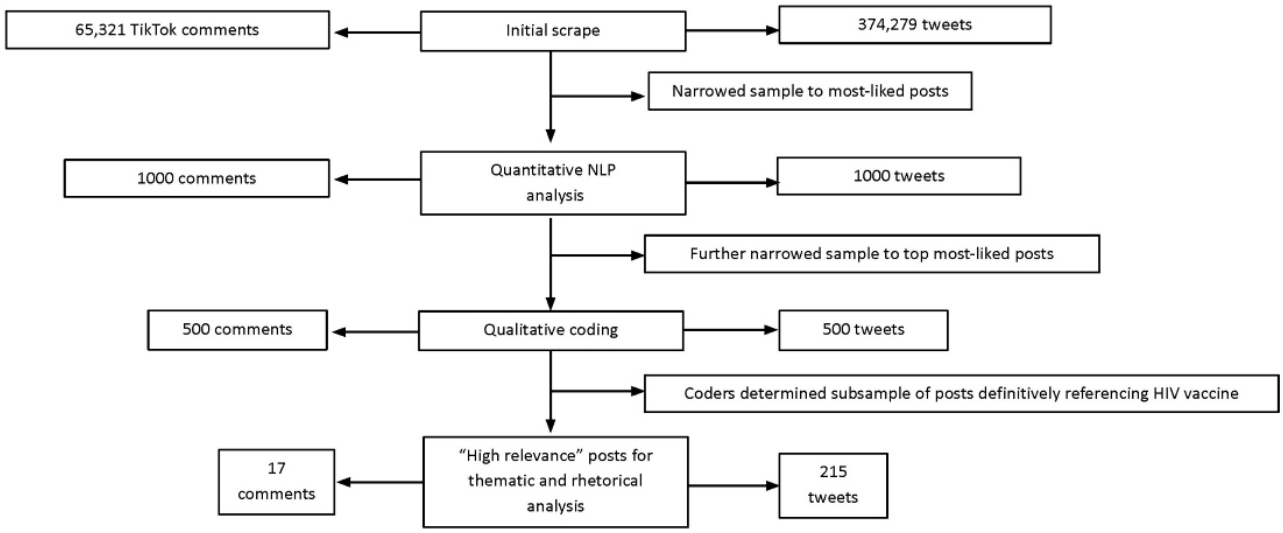
The study was approved by the ethics committees at the University of the Witwatersrand (reference: 230904), the University of Pennsylvania (854420), the University of Cape Town (666/2023), and Boston University (H-44422). The study used only secondary data and removed usernames and identifying details from posts to preserve individual privacy.

Results

Dataset Description

The data scrape yielded a dataset of 374,279 tweets and 65,321 TikTok comments, along with accompanying metadata. Notably, in the case of the TikTok videos, 150 out of 459 videos were posted from a single account, although we were not able to trace comments back to their original video. From these data, we conducted a quantitative analysis on the 1000 highest-liked posts from each dataset and a qualitative analysis of the top 500 posts (see Figure 1 for a diagram of the study outputs). The sections below outline the results of the lexical analysis, as well as the rhetorical and thematic analysis of the high-relevance subset of 232 posts referencing an HIV vaccine.

Figure 1. Outputs at different stages of analysis. NLP: natural language processing.



Quantitative Analysis

Lexical analysis offered insight into areas in which the datasets converged and diverged (Table 1). Tweets tended to be longer than TikTok comments (41.2, SD 16.2 vs 15.7, SD 9.6 words on average; $P<.001$), with more complex sentence structure and language (16.4 vs 5.3 Flesch-Kincaid grade level score; $P<.001$).

XSL•FO
RenderX

Table . Comparative lexical analysis of tweets and TikTok comments.

Category	Range	TikTok comments (n=1000), mean (SD)	Tweets (n=1000), mean (SD)	*t* test (*P* value)
Word count	—[a]	15.8 (9.6)	41.2 (16.2)	42.5[b] (<.001)
Flesch-Kincaid Grade level	0+	5.3 (4.7)	16.4 (5.2)	50.2[b] (<.001)
Emotionality	0 to 9	1.8 (2.7)	1.5 (2.3)	–2.3[b] (.02)
Emotional valence	0 (highly negative attitude) to 9 (highly positive attitude)	1.8 (3)	1.6 (2.7)	–2.0 (.05)
Anger	0 to 1	0.14 (0.12)	0.19 (0.1)	11.4[b] (<.001)
Disgust	0 to 1	0.25 (0.23)	0.44 (0.22)	18.9[b] (<.001)
Fear	0 to 1	0.17 (0.2)	0.29 (0.21)	13.1[b] (<.001)
Joy	0 to 1	0.25 (0.29)	0.30 (0.23)	4.4[b] (<.001)
Anticipation	0 to 1	0.19 (0.22)	0.32 (0.21)	13.2[b] (<.001)
Trust	0 to 1	0.26 (0.35)	0.37 (0.26)	7.9[b] (<.001)
Negative sentiment	−1 to 0	−0.37 (0.22)	−0.42 (0.17)	−5.0[b] (<.001)
Positive sentiment	0 to 1	0.29 (0.26)	0.3 (0.12)	−0.4 (.72)

[a]Not applicable.

[b]significant at $P<.05$.

Tweets showed higher mean scores than comments for all 6 measured emotions, with markedly greater differences for disgust (0.44, SD 0.22 in tweets; 0.25, SD 0.23 in comments; $P<.001$), anticipation (0.32, SD 0.21 in tweets; 0.19, SD 0.22 in comments; $P=<.001$) and fear (0.29, SD 0.21 in tweets; 0.17, SD 0.2 in comments; $P<.001$). In terms of emotionality, TikTok comment scores were only slightly higher (0.25 difference on the 0 - 9 score; $P=.02$). This indicates that the sentiments expressed in TikTok comments were more heavily based in emotion than tweets, though this did not necessarily map onto specific emotions generated by TextAnalyzer. Additionally, the two were similar in terms of mean emotional valence (1.8, SD 3 for comments and 1.6, SD 2.7 for tweets; $P=.05$), tending toward negative attitudes, though comments were slightly more positive. This was also evident in sentiment; the platforms had somewhat similar mean negative sentiment scores (−0.42, SD 0.17 for tweets; −0.37, SD 0.22 for comments; $P<.001$) slightly greater than their mean positive sentiment score (0.3, SD 0.12 for tweets; 0.29, SD 0.26 for comments; $P=.72$). This showed that posts exhibited moderate average positive and negative associations, skewing slightly higher toward negative sentiment, and with incrementally greater averages for tweets versus comments.

Qualitative Analysis

Coders examined the discourse surrounding HIV vaccine development on TikTok and Twitter. The codes fell into 4 categories: topic, vaccine sentiment, content themes, and rhetorical strategies. Out of 500 tweets and 500 TikTok comments, 215 tweets and 17 TikTok comments were coded under the topic code “HIV vaccines.” The themes and rhetorical strategies most prevalent in this highly relevant dataset of 232 total posts are outlined below. See Table 2 for a full list of key themes and rhetorical strategies.

XSL•FO
RenderX

Table . Key themes and rhetorical strategies.

	Definition	Example
Key theme		
Scientific and technological progress	Displays of trust and celebration of scientific research and technological progress, as evidenced by HIV vaccine development.	"...This is really exciting news I wanted to share. The first participant has been dosed in the phase I study of Moderna's HIV vaccine candidate, mRNA-1644, which uses the same mRNA technology as our COVID-19 vaccines!" [Tweet, record #7]
Hope and excitement for HIV vaccine	Enthusiasm voiced for an HIV vaccine and the impact it may hold.	"Wow—as a queer elder, this news made me tear up: An HIV vaccine is in phase 1 trials!!!" [Tweet, record #184]
Distrust in pharmaceutical companies and government	Suspicion of the government's and pharmaceutical companies' motives as fueled by news of HIV vaccine development	"Any guess on how long it will be before your rights of free movement are restricted until you have taken the new HIV vaccine? I mean, we can't be too careful 'for the common good…'" [Tweet, record #272]
Belief that COVID-19 vaccine causes HIV	Ideas that recent progress in HIV vaccine development confirms the conspiracy theory that the COVID-19 vaccine causes HIV	"So when the shot causes VAIDS and you test positive for HIV don't panic!! They've got you covered with an experimental mRNA vax to make you all better " [Tweet, record #462]
Key rhetorical strategy		
Pathos	Use of vivid, emotional language in describing excitement and hope regarding an HIV vaccine	"If this came 6 yrs ago I might still have [name]. He is loved and missed. There is so much hope, I'm crying happy tears" [TikTok comment, record #380]
Conspiracy	Depicting HIV vaccine development as an indicator of a secret, often malevolent plot in order to evoke mistrust in an entity (ie, the government, pharmaceutical industry)	"There is no vaccine for HIV after 40 years of research. No cure for cancer after 100 years of research. The flu spreads like wildfire each year without any concern. Yet a virus appears out of nowhere, a vaccine magically appears from 4 different pharma and forced on us?." [Tweet, record #10]
Key figures	Invoking prominent figures to garner support for or against HIV vaccines	"Groundhog Day: Nearly every year since 1983, Fauci extracted another billion+from Congress promising AIDS Vax just as the press dolefully announces failure of his most recent candidate. His Rasputin-like hold on press+politicians means game never ends." [Tweet, record #23]
Post hoc propter hoc	Fallacy that events happening in succession must have a causal relationship, namely with regard to fears that the COVID-19 vaccine causes HIV, and subsequent strides in HIV vaccine development	"Last week it was announced there was a new aggressive HIV going around in Neatherlands [sic], the next day [it] was announced there is a new MRNA HIV shot. Create a problem (HIV in vax) then create the solution (new HIV vax). Just our world leaders doing what they do" [Tweet, record #137]

Content Themes

Scientific and Technological Progress (57 Posts)

Content demonstrated trust in scientific research and celebration of technological advancements, particularly in the realm of vaccine development, across a total of 57 posts. Fifteen posters specifically highlighted advancements in mRNA technology during COVID-19 vaccine development, and how these advancements have helped bring the HIV vaccine closer to reality:

> *...This is really exciting news I wanted to share. The first participant has been dosed in the phase I study of Moderna's HIV vaccine candidate, mRNA-1644, which uses the same mRNA technology as our COVID-19 vaccines!* [Tweet, record #7]

Hope and Excitement for the HIV Vaccine (40 Posts)

Posters expressed optimism regarding the development of an HIV vaccine in 40 posts, and even described it as a monumental breakthrough. The vaccine was seen as offering hope, especially to those who have been personally impacted by HIV, and as a critical step forward in the HIV response:

> *Wow—as a queer elder, this news made me tear up: An HIV vaccine is in phase 1 trials!!!* [Tweet, record #184]

XSL•FO
RenderX

Distrust in Pharmaceutical Companies and Government (34 Posts)

Many posters expressed a lack of trust in pharmaceutical companies and the government. There was speculation that the "sudden" announcement of HIV vaccine trials after mass COVID-19 vaccinations was part of a larger conspiracy—whether to cover up COVID-19 vaccine–induced HIV or to exploit and control the public. Posters also voiced concerns about government support for an HIV vaccine. They feared that they would be forced to take the vaccine against their will, as they felt had happened with COVID-19 vaccine mandates:

> *Any guess on how long it will be before your rights of free movement are restricted until you have taken the new HIV vaccine? I mean, we can't be too careful 'for the common good…'* [Tweet, record #272]

Belief That COVID-19 Vaccine Causes HIV (20 Posts)

Various tweets and comments used HIV vaccine development announcements to amplify posters' theories that the COVID-19 vaccine causes HIV. These posters viewed the HIV vaccine development as proof of a conspiracy, claiming that pharmaceutical companies developed COVID-19 vaccines to infect recipients with HIV and thereby increase demand for their HIV vaccines.

> *So when the shot causes VAIDS and you test positive for HIV don't panic!! They've got you covered with an experimental mRNA vax to make you all better* [Tweet, record #462]

Common Rhetorical Strategies

Overview

Below are the most common rhetorical strategies, listed in order of frequency. We included the 4 most frequent rhetorical strategies because the second to fourth most-coded strategies appeared nearly the same number of times in the sample, with 20 to 22 references, after which point rhetorical strategies dropped to 15 or fewer codes in the high-relevance sample.

Pathos (27 Posts)

Some posters expressed heightened emotion, or pathos, in their posts. These 27 posts were overwhelmingly pro–HIV vaccine; for example, posters shared their joy for a vaccine and the sadness about the toll of HIV so far. In describing their personal hopes, struggles, and celebrations, these posts draw posters into more personal narratives supporting HIV vaccines and invite them to feel delight in the progress or grief for lives lost alongside the original poster.

> *If this came 6 yrs ago I might still have [name]. He is loved and missed. There is so much hope, I'm crying happy tears* [TikTok comment, record #380]

Conspiracy (22 Posts)

Conspiracy theories emerged as a predominant rhetorical strategy used by anti–HIV vaccine posters. In these comments and tweets, posters conjured up and repeated extreme and dire circumstances as proof that the HIV vaccine cannot be trusted. By making users fear the most horrific of circumstances, posters generated doubt in the safety of the HIV vaccine and those producing it. As described in the "Pathos" section, various posters parroted the conspiracy theory that the COVID-19 vaccine causes HIV. They were convinced that the government and pharmaceutical companies were being duplicitous in hiding this and trying to promote an HIV vaccine in response to "vaccine-induced AIDS." Some were also convinced that the speed with which the COVID-19 vaccine was developed indicated a kind of government and pharmaceutical conspiracy.

> *There is no vaccine for HIV after 40yrs of research. No cure for cancer after 100yrs of research. The flu spreads like wildfire each year without any concern. Yet a virus appears out of nowhere, a vaccine magically appears from 4 different pharma and forced on us?....* [Tweet, record #10]

Key Figures (21 Posts)

Various posters mentioned key figures in their posts regarding an HIV vaccine. In this strategy, posters referenced well-known individuals to support their (largely anti-HIV vaccine) arguments—figures whom users would easily recognize. By naming these individuals, most of whom are powerful and evoke strong opinions, posters attach a face to the corruption or mistrust they are trying to incite.

One figure who was often mentioned to support users' distrust of the government and vaccines was Anthony Fauci, former director of the National Institute for Allergy and Infectious Diseases, who has been a central figure in the HIV and COVID-19 responses [34]. Twelve posters pointed to past statements by Fauci about HIV vaccine development, using them to foster distrust in current vaccine development practices. Other key figures mentioned include Prince Harry, Elon Musk, and Luc Montagnier, who was awarded the Nobel Prize for his discovery of HIV and later perpetuated antivaccine conspiracy theories in his later years [35].

> *Groundhog Day: Nearly every year since 1983, Fauci extracted another billion+from Congress promising AIDS Vax just as the press dolefully announces failure of his most recent candidate. His Rasputin-like hold on press+politicians means game never ends.* [Tweets, record #23]

Post Hoc Propter Hoc (20 Posts)

Twitter and TikTok posters often use post hoc propter hoc reasoning (assuming that because one event happened after another, they must have a causal relationship) to convince readers that the HIV vaccine cannot be trusted. By pairing the 2 events together (eg, mass COVID vaccination and announcements of new HIV vaccine clinical trials), the posters suggest or explicitly tell readers that the two must be related, and that there is a sort of causal relationship between them.

> *Last week it was announced there was a new aggressive HIV going around in Neatherlands [sic], the next day [it] was announced there is a new MRNA HIV shot. Create a problem (HIV in vax) then create*

XSL•FO
RenderX

the solution (new HIV vax). Just our world leaders doing what they do. [Tweet, record #137]

Discussion

Principal Findings

In this study, we conducted a mixed-methods comparative, thematic, and rhetorical analysis of public discourse surrounding the HIV vaccine on Twitter and TikTok. We focused on the most highly engaged posts, first selecting the 1000 most-liked for quantitative analysis, then the 500 most-liked for qualitative analysis. We found that the majority of content in direct reference to the HIV vaccine came from Twitter, and that tweets tended to be longer and more linguistically complex than TikTok comments. This may in part be due to the different character limits between the apps—280 characters for tweets versus 150 characters for TikTok comments [21-23]. Major thematic areas included support for scientific progress and excitement around vaccine development, as well as distrust in government and scientific entities, and beliefs that the HIV vaccine is being introduced to combat COVID-19 vaccine–induced HIV. Across both platforms, there was a mix of positive and negative sentiment regarding HIV vaccines, and a pattern of rhetorical strategies used to convey these ideas. The dichotomy of emotions toward the HIV vaccine was evident; emotions most seen in tweets were disgust and trust, and in TikTok comments were trust, joy, and disgust.

We aimed to analyze not only what people were saying about an HIV vaccine online, but also how they were saying it. Using *a priori coding* informed by misinformation research, we identified and refined four common rhetorical strategies: conspiracy theories, post hoc propter hoc, references to key figures, and pathos. The first three strategies (conspiracy theories, post hoc reasoning, and references to key figures) were primarily associated with anti-HIV-vaccine sentiment. For example, some posters claimed that the HIV vaccine was being pushed as a response to the COVID-19 vaccine supposedly infecting people with HIV, suggesting government or pharmaceutical conspiracies and citing figures like Anthony Fauci as evidence of corruption. In contrast, pathos was more common in provaccine messaging. Supporters shared personal stories of loss during the HIV pandemic and expressed hope and excitement at the prospect of a vaccine. These emotionally resonant narratives appeared to foster more positive perceptions of the HIV vaccine. Understanding both negative and positive rhetorical strategies provides insight into how vaccine discourse unfolds online and highlights potential pathways for encouraging public support.

In our literature search, we found only 1 study pertaining to social media discourse on HIV vaccines specifically. In this 2024 study, Zhang et al [7] used machine learning algorithms to analyze tweets posted in 2022 about HIV vaccines, then manually coded the most highly engaged tweets about HIV and COVID-19. This study situated the discourse surrounding the HIV vaccine largely within the context of the COVID-19 vaccine, similarly revealing both positive sentiments as posters compared the progress of one vaccine to another, and negative sentiments as they parroted the conspiracy theory that the COVID-19 vaccine causes HIV, and the introduction of an HIV vaccine as part of a larger government conspiracy.

Other studies in related topic areas indicate similar thematic trends. One study of antivaccine attitudes on social media found an emphasis on conspiracy theories in messaging, complete with "secret, sinister organizations and manipulative government bodies causing harm" [36]. In this sense, the perpetuation of conspiracy theories is not unique to the topic of HIV vaccines, although the topic of those theories may vary from vaccine to vaccine. A study of COVID-19 vaccine social media discourse also showed a mixture of positive sentiment about vaccine development and availability, combined with negative sentiments about vaccine safety, conspiracies, and the impact of the government [5].

The themes uncovered in our HIV vaccine discourse analysis are distinct from those in past studies of PrEP communication on social media. Broadly, users sought to share information about both topics, and topics in both areas included misinformation about HIV and its prevention [12,13,15]. However, there were different conversations about who is most impacted by HIV. A TikTok study of PrEP communication found that homophobic and heteronormative misconceptions influenced much of the dialogue [14]. This theme was not present in our data, where conversations focused more on the hope of an HIV vaccine after losing so many people in the queer community to HIV. This is interesting, considering that both the HIV vaccine and PrEP are designed to prevent HIV. However, the differences in discussion may be due to the fact that there is far less concrete information about the HIV vaccine at this stage than there is about PrEP.

During the qualitative analysis, it became clear that most data confidently categorized as related to an HIV vaccine came from Twitter. A total of 248 TikTok comments, nearly half the sample, were coded as "unable to ascertain," compared to only 10 tweets, indicating difficulty in determining the subject matter of many TikTok comments. Vague remarks like "This is the kind of history I want to live through" or "my jaw dropped" lacked clear topic indicators and offered limited qualitative insight. One reason for this may be that tweets often stand alone and convey meaning independently, whereas TikTok comments rely heavily on the accompanying video, making interpretation difficult without that context. This issue was compounded by different scraping methods: tweets entered the dataset only if they included specific search terms, while TikTok comments were collected if they responded to videos that matched search criteria, regardless of their own content.

Furthermore, the average TikTok comment contained 15.8 (SD 9.6) words, nearly a third of the tweets' average of 41.2 (SD 16.2) words. In addition, the comments' Flesch-Kincaid grade level was an average of 5.3 (SD 4.7), much lower than the tweets' estimated average of 16.4 (SD 5.2). In other words, one dataset is composed of fairly short, simplistic sentiments, while the other contains much longer ideas written at a graduate level. As such, it is perhaps unsurprising that the tweets yielded richer data regarding an HIV vaccine.

This study's findings can inform public health messaging by highlighting the distinct communication needs of social media

XSL•FO
RenderX

audiences across platforms. Murthy et al's [37] "3 Rs" framework—reviewing the audience, recognizing communication needs, and responding appropriately—offers a useful model for time-sensitive health topics like the HIV vaccine. Our analysis supports the first 2 "Rs" by identifying audience communication patterns, prevalent themes, and rhetorical strategies. We found that TikTok and Twitter posters' most prevalent barriers to HIV vaccine acceptance are a lack of trust in the government and scientific institutions, as well as the belief in the conspiracy theory that the HIV vaccine was introduced as a response to the COVID-19 vaccine–induced HIV. Additionally, many people *do* feel excited about the idea of future HIV vaccines, and many cited personal stories and emotions about the lifesaving potential of an HIV vaccine. Should future public health campaigns arise to combat HIV vaccine misinformation, they might focus on rebuilding trust in the scientific and governmental systems, dispelling misinformation about COVID-19 vaccine–induced HIV, and using personal stories from real people to show the benefits an HIV vaccine offers. Effective messaging should also consider platform differences, using more scientific, complete sentences on Twitter and shorter, simpler language on TikTok to better engage each audience.

Limitations

This study offers a snapshot of evolving social media discourse on the HIV vaccine and focuses only on English-language posts. Since dialogue on social media moves much faster than the research following it, these data do not capture trends or events occurring from 2024 to the present. However, we feel confident that the core themes extracted from TikTok and Twitter remain relevant and beneficial to researchers and policymakers in the HIV vaccine space. In addition, TikTok and Twitter are 2 distinct platforms with different post offerings, which pose a challenge in comparing their content. Although the study considers short text posts on both platforms, the tweets analyzed were a product of direct keyword filtering, whereas TikTok comments came from videos matching the search terms. As such, stand-alone TikTok comments largely offered less insight than tweets and made up a much smaller proportion of the high-relevance dataset. For this reason, we also extended the window of time over which TikTok comments could be posted to include in the study, ranging until October 2023, about 7 months later than the tweets we included. While the results give new insights into an uncharted area of HIV vaccine discourse on social media, the limitations in the comparative aspect of the qualitative analysis are important to consider. It should also be noted that nearly one-third of TikTok videos from which comments were scraped came from a single account that reposted content from others. Since there was no significant alteration in these reposts, we find that the comments scraped still reflect prevailing TikTok narratives about the HIV vaccine. However, we acknowledge that the followers of this account may be more heavily represented in our findings.

Directions for Further Research

At the time of publication, there is little research analyzing TikTok comments or comparing TikTok comments to other social media posts. Further research should aim to illuminate best practices in analyzing this type of information and extracting themes from TikTok comments. Future research should also continue to examine the discourse surrounding the HIV vaccine, especially as trials progress, on different social media platforms. Other platforms such as Instagram, or the newly popular platform BlueSky—a more liberal substitute for Twitter for many—may offer a new lens on HIV vaccine discourse to compare to other platforms [38]. Researchers should also consider piloting messaging or interventions to curb misinformation surrounding the vaccine, especially if it involves penetrating online groups that are echo chambers for vaccine distrust and misinformation.

Conclusions

As medical technology advances, the prospect of a successful HIV vaccine feels closer than ever. Thus, it is essential to understand the nature of public discourse regarding the vaccine. By conducting a mixed methods comparative analysis of Twitter and TikTok comments, we offer further context and additional themes in this exceedingly small body of literature. These insights are valuable to public health, medical, and governmental actors seeking to promote accurate information and foster trust in the HIV vaccine. Our findings show that while both platforms feature expressions of hope and trust in science, they also contain concerns about institutional corruption and conspiracy theories, such as the belief that the HIV vaccine responds to harm caused by the COVID-19 vaccine. Tweets tended to be more linguistically complex and yielded richer insights, while TikTok comments were shorter and more difficult to interpret without video context. Key rhetorical strategies included conspiracy theories, post hoc reasoning, and emotional appeals. These findings underscore the need for platform-specific communication strategies to address misinformation and build public trust. As conversations on social media continue to rapidly evolve, future research should continue to monitor and analyze new trends in sentiment toward HIV vaccines. Researchers should also broaden the scope to include additional text-based platforms such as Threads, as well as other languages, and explore tailored interventions that align with each platform's communication style. By understanding how users express themselves in the midst of misinformation and an ever-changing social media landscape, we can craft more empathetic, accurate, and effective public health messaging to support confidence in an HIV vaccine.

Acknowledgments

During the preparation of this work, the authors used ChatGPT to refine wording and ensure clarity in certain sections. After using this tool or service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

XSL•FO
RenderX

Funding

This research was supported by the National Institutes of Health (R01MH132401). The funder played no role in study design, data collection, analysis, and interpretation of data, or the writing of this manuscript.

Data Availability

The data that support the findings of this study are available from the corresponding author (MAR) upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook developed by coders to qualitatively analyze tweets and comments.
[DOCX File, 18 KB - infodemiology_v6i1e82917_app1.docx]

References

1. Islam MS, Sarkar T, Khan SH, et al. COVID-19–related infodemic and its impact on public health: a global social media analysis. Am J Trop Med Hyg 2020 Oct;103(4):1621-1629. [doi: 10.4269/ajtmh.20-0812] [Medline: 32783794]
2. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
3. Burki T. Vaccine misinformation and social media. Lancet Digit Health 2019 Oct;1(6):e258-e259. [doi: 10.1016/S2589-7500(19)30136-0]
4. Roozenbeek J, Schneider CR, Dryhurst S, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci 2020 Oct;7(10):201199. [doi: 10.1098/rsos.201199] [Medline: 33204475]
5. Hussain A, Tahir A, Hussain Z, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. J Med Internet Res 2021 Apr 5;23(4):e26627. [doi: 10.2196/26627] [Medline: 33724919]
6. Wawrzuta D, Jaworski M, Gotlib J, Panczyk M. Characteristics of antivaccine messages on social media: systematic review. J Med Internet Res 2021 Jun 4;23(6):e24564. [doi: 10.2196/24564] [Medline: 34085943]
7. Zhang JM, Wang Y, Mouton M, Zhang J, Shi M. Public discourse, user reactions, and conspiracy theories on the X platform about HIV vaccines: data mining and content analysis. J Med Internet Res 2024 Apr 3;26:e53375. [doi: 10.2196/53375] [Medline: 38568723]
8. Basch CH, MacLean SA. A content analysis of HPV related posts on Instagram. Hum Vaccin Immunother 2019;15(7-8):1476-1478. [doi: 10.1080/21645515.2018.1560774] [Medline: 30570379]
9. HIV vaccines. International AIDS Vaccine Initiative. 2025. URL: https://www.iavi.org/our-work/hiv-vaccines/ [accessed 2025-07-31]
10. Moyo E, Moyo P, Murewanhema G, Mhango M, Chitungo I, Dzinamarira T. Key populations and Sub-Saharan Africa's HIV response. Front Public Health 2023;11:1079990. [doi: 10.3389/fpubh.2023.1079990] [Medline: 37261232]
11. Vulnerable groups and key populations at increased risk of HIV. World Health Organization Regional Office for the Eastern Mediterranean. 2025. URL: https://www.emro.who.int/asd/health-topics/vulnerable-groups-and-key-populations-at-increased-risk-of-hiv.html [accessed 2025-07-31]
12. Edinger A, Valdez D, Walsh-Buhi E, Bollen J. Deep learning for topical trend discovery in online discourse about pre-exposure prophylaxis (PrEP). AIDS Behav 2023 Feb;27(2):443-453. [doi: 10.1007/s10461-022-03779-2] [Medline: 35916950]
13. Walsh-Buhi E, Houghton RF, Lange C, Hockensmith R, Ferrand J, Martinez L. Pre-exposure prophylaxis (PrEP) information on Instagram: content analysis. JMIR Public Health Surveill 2021 Jul 27;7(7):e23876. [doi: 10.2196/23876] [Medline: 34061759]
14. Lewis J, Melendez-Torres GJ. Prep-Tok: a queer critical discourse analysis of TikToks regarding HIV-related pre-exposure prophylaxis. Cult Health Sex 2024 Apr;26(4):449-465. [doi: 10.1080/13691058.2023.2224414] [Medline: 37435990]
15. Xu Q, McMann T, Godinez H, et al. Impact of COVID-19 on HIV prevention access: a multi-platform social media infodemiology study. AIDS Behav 2023 Jun;27(6):1886-1896. [doi: 10.1007/s10461-022-03922-z] [Medline: 36471205]
16. Muric G, Wu Y, Ferrara E. COVID-19 vaccine hesitancy on social media: building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. JMIR Public Health Surveill 2021 Nov 17;7(11):e30642. [doi: 10.2196/30642] [Medline: 34653016]
17. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine (Auckl) 2018 Jun;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040]
18. Woodward M. TikTok user statistics 2025: everything you need to know. SearchLogistics. 2021. URL: https://www.searchlogistics.com/learn/statistics/tiktok-user-statistics/ [accessed 2025-07-31]

XSL•FO
RenderX

19. Hern A. How TikTok's algorithm made it a success: 'it pushes the boundaries'. The Guardian. 2022 Oct 24. URL: https://www.theguardian.com/technology/2022/oct/23/tiktok-rise-algorithm-popularity [accessed 2026-01-12]
20. Twitter's recommendation algorithm. X. 2023 Mar 31. URL: https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm [accessed 2026-01-12]
21. Counting characters. X. URL: https://docs.x.com/fundamentals/counting-characters [accessed 2025-12-16]
22. TikTok business capabilities and limitations. Sprinklr Help Center. 2025. URL: https://www.sprinklr.com/help/articles/capabilities-and-limitations-tiktok/tiktok-business-capabilities-and-limitations/68317e3ca6fe4f29b3d5fd4e [accessed 2025-12-16]
23. TikTok character counter. CharacterCounter.com. URL: https://charactercounter.com/tiktok [accessed 2025-12-19]
24. NIH launches clinical trial of three mRNA HIV vaccines. National Institutes of Health. 2022 Mar 14. URL: https://www.nih.gov/news-events/news-releases/nih-launches-clinical-trial-three-mrna-hiv-vaccines [accessed 2025-07-31]
25. Berger J, Sherman G, Ungar L. TextAnalyzer. 2020. URL: https://textanalyzer.org/ [accessed 2025-07-31]
26. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel. : Defense Technical Information Center, US Department of Defense; 1975 Feb URL: https://apps.dtic.mil/sti/pdfs/ADA006655.pdf [accessed 2026-02-27]
27. Kiritchenko S, Zhu X, Mohammad SM. Sentiment analysis of short informal texts. J Artif Intell Res 2014 Aug 20;50:723-762. [doi: 10.1613/jair.4272]
28. Mohammad SM. #Emotional tweets. Presented at: Proceedings of the First Joint Conference on Lexical and Computational Semantics; Jun 7-8, 2012 URL: https://aclanthology.org/S12-1033/ [accessed 2026-02-27]
29. Mohammad SM, Kiritchenko S. Using hashtags to capture fine emotion categories from tweets. Comput Intell 2015 May;31(2):301-326. [doi: 10.1111/coin.12024]
30. Rocklage MD, Rucker DD, Nordgren LF. The Evaluative Lexicon 2.0: the measurement of emotionality, extremity, and valence in language. Behav Res Methods 2018 Aug;50(4):1327-1344. [doi: 10.3758/s13428-017-0975-6] [Medline: 29052167]
31. Hopkins KL, Lepage C, Cook W, et al. Co-designing a mobile-based game to improve misinformation resistance and vaccine knowledge in Uganda, Kenya, and Rwanda. J Health Commun 2023 Oct 31;28(sup2):49-60. [doi: 10.1080/10810730.2023.2231377] [Medline: 38146161]
32. Roozenbeek J, van der Linden S. Fake news game confers psychological resistance against online misinformation. Palgrave Commun 2019 Jun 25;5(1):65. [doi: 10.1057/s41599-019-0279-9]
33. Da San Martino G, Barrón-Cedeño A, Wachsmuth H, Petrov R, Nakov P. SemEval-2020 task 11: detection of propaganda techniques in news articles. Presented at: Proceedings of the Fourteenth Workshop on Semantic Evaluation; Dec 12-13, 2020 p. 1377-1414. [doi: 10.18653/v1/2020.semeval-1.186]
34. Anthony S. Fauci, M.D. National Institute of Allergy and Infectious Diseases. 2025. URL: https://www.niaid.nih.gov/about/anthony-s-fauci-md [accessed 2025-07-31]
35. Kazi F, Mushtaq A. Luc Montagnier. Lancet Infect Dis 2022 Apr;22(4):458. [doi: 10.1016/S1473-3099(22)00170-0]
36. Mitra T, Counts S, Pennebaker J. Understanding anti-vaccination attitudes in social media. Presented at: Proceedings of the International AAAI Conference on Web and Social Media; Dec 12-13, 2020. [doi: 10.1609/icwsm.v10i1.14729]
37. Murthy BP, LeBlanc TT, Vagi SJ, Avchen RN. Going viral: the 3 Rs of social media messaging during public health emergencies. Health Secur 2021;19(1):75-81. [doi: 10.1089/hs.2020.0157] [Medline: 33372847]
38. Chayka K. Bluesky's quest to build nontoxic social media. New Yorker. 2025 Apr 7. URL: https://www.newyorker.com/magazine/2025/04/14/blueskys-quest-to-build-nontoxic-social-media [accessed 2025-07-31]

Abbreviations

PrEP: pre-exposure prophylaxis

Edited by T Nguyen; submitted 24.Aug.2025; peer-reviewed by Sunny, CL Au, T Mcmann; revised version received 09.Feb.2026; accepted 09.Feb.2026; published 19.Mar.2026.

Please cite as:
Rabin MA, Penuela-Wermers S, Sehgal NKR, Egbe TI, Lavery CLM, Guntuku SC, Buttenheim AM
Social Media Perspectives on a Future HIV Vaccine: Mixed Methods Analysis
JMIR Infodemiology 2026;6:e82917
URL: https://infodemiology.jmir.org/2026/1/e82917
doi:10.2196/82917

XSL•FO
RenderX

© Megan A Rabin, Sarah Penuela-Wermers, Neil K R Sehgal, Teniola I Egbe, Criswell L M Lavery, Sharath Chandra Guntuku, Alison M Buttenheim. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 19.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Online Information Behavior Regarding COVID-19 Vaccination and Its Association With Vaccination Behavior Based on Cluster Analysis of User Groups: Cross-Sectional Study

Lea Liebner[1], MEd; Birgit Babitsch[1], MPH, PhD, Prof Dr; Lisa Schmidt[1], MPH

Department of New Public Health, School of Human Sciences, Osnabrück University, Osnabrück, Germany

Corresponding Author:
Lisa Schmidt, MPH
Department of New Public Health
School of Human Sciences
Osnabrück University
Nelson-Mandela-Str. 13
Osnabrück, 49076
Germany
Phone: 49 541 969 7016
Email: lisa.schmidt@uni-osnabrueck.de

Abstract

Background: The COVID-19 pandemic highlighted the importance of effective health communication and reliable information for crisis management, particularly following the introduction of vaccinations. Varied attitudes toward COVID-19 vaccination and an overwhelming amount of online information complicated communication and pandemic management. Previous studies have often focused on general vaccination behavior and its correlation with vaccination attitudes, establishing a link between information-seeking and vaccination decisions. However, there is insufficient analysis distinguishing specific user groups based on their actual online information behavior regarding COVID-19 vaccination and examining its correlation with vaccination behavior.

Objective: This study aims to fill this research gap by identifying user groups based on their information behavior and investigating its influence on vaccination uptake.

Methods: As part of the "Internetnutzung zur COVID-19-Impfung" (INCOVI) study, 1000 individuals in Germany were surveyed online (November 26 to December 8, 2021) regarding their internet usage related to COVID-19 vaccination. A hierarchical cluster analysis was conducted to identify user groups. Logistic regression analyses were then used to explore correlations among the user groups and their demographic characteristics, readiness to vaccinate, knowledge of vaccination, and health literacy. Additionally, a logistic regression analysis was performed to identify the influence of user groups and other factors on vaccination behavior.

Results: A total of 3 user groups were identified: frequent and critical information evaluators (454/778, 58.4%), who primarily relied on official information sources, exhibited a higher level of health literacy, and were older than the other groups; infrequent and passive recipients (222/778, 28.5%), who rarely sought information actively and were younger than the other groups; and frequent and multichannel, interaction-focused users (102/778, 13.1%), who actively searched across multiple channels and engaged in information exchange. Notably, the user groups did not significantly differ in knowledge or willingness to vaccinate. User group affiliation, knowledge, and health literacy did not significantly influence vaccination behavior. The strongest predictor of vaccination was preexisting willingness to vaccinate. Additionally, women were more likely to be vaccinated than men, and individuals with medium or higher education levels were 6-11 times more likely to be vaccinated compared to those with only a basic level of education.

Conclusions: Segmenting the population into different user groups allows for more targeted communication tailored to the specific needs and beliefs of each group. Because these groups stem from observable usage patterns, they constitute a transferable framework for other health topics. For frequent and critical information evaluators, providing well-founded and detailed information on public channels is important. Infrequent and passive recipients benefit from straightforward formats, such as short explanatory videos, while frequent and multichannel, interaction-focused users are better reached through interactive offerings on social media. By specifically targeting these groups, informed decision-making about vaccinations can be supported.

XSL•FO
RenderX

(JMIR Infodemiology 2026;6:e82221) doi:10.2196/82221

KEYWORDS

cluster analysis; COVID-19 vaccination; information-seeking behavior; internet use; user groups

Introduction

In the context of the COVID-19 pandemic, the importance of effective health communication and informed information behavior became evident worldwide, and in particular in Germany [1]. The rapid introduction of vaccines against SARS-CoV-2 and the overwhelming amount of information necessitated comprehensive and targeted communication to build trust and encourage vaccination. In Germany, vaccination against different diseases is a regular part of health care and is financed by health insurance [2]. However, vaccination rates are below expectations. The SARS-CoV-2 vaccination was the subject of critical discussion, which intensified when it became an obligatory part of the measures to combat the COVID-19 pandemic [3].

The effects of this political shift were also reflected in the vaccination rates. After the introduction of the first COVID-19 vaccines in Germany, the vaccination readiness score (on a scale from 1 to 7) significantly increased from a mean value of 4.14 in December 2020 to 6.45 in December 2021, and then remained consistently high throughout the pandemic [4]. Although willingness to be vaccinated was high, the actual vaccination rate was lower than expected. As of November and December 2021, the basic vaccination rate in Germany was only 69.8%, while the Robert Koch Institute (Germany's public health institute) aimed for a target vaccination rate of 85% for those aged 12-59 years and 90% for individuals aged 60 years and older by autumn 2021 [5,6]. Consequently, factors other than willingness to be vaccinated influence actual vaccination behavior.

Different models have been developed to explain vaccination behavior. One such is the 7C model by Geiger et al [7], which addresses the mechanism of the preliminary stage of vaccination behavior. It captures the vaccination decisions for or against the COVID-19 vaccination. The 7 "Cs" of the model are: confidence, complacency, constraints, calculation, collective responsibility, compliance, and conspiracy. Confidence refers to an individual's trust in the efficacy and safety of vaccinations, as well as in the health care system and the responsible organizations. People with high confidence levels are generally more willing to be vaccinated. Complacency is the subjective assessment of one's own risk of disease. People who perceive their risk of contracting a disease as low often show a lower willingness to be vaccinated. Structural, psychological, or everyday barriers that can make vaccination difficult are summarized in the determinant constraints. If vaccination is not considered particularly important in overcoming barriers, willingness to be vaccinated is low. The determinant calculation refers to the subjective weighing of the risks and benefits or advantages and disadvantages of vaccination for individuals. Collective responsibility describes the willingness to be vaccinated not only to protect oneself, but also to protect others. People with a high sense of responsibility toward the community recognize the indirect protection that their own vaccination offers others and are more willing to be vaccinated. Geiger et al [7] describe the willingness to comply with social recommendations, such as vaccination and monitoring, as compliance [7]. The final determinant of the 7C model, conspiracy, considers the tendency to believe in conspiracy theories and fake news. The 7C model includes psychological, social, and structural factors, making it possible to investigate and understand the multifaceted and complex nature of vaccination behavior. In this study, readiness to vaccinate, as measured by the 7C scale, is considered one of several determinants of vaccination behavior. Another contributing factor could be health information behavior, that is, the way in which people obtain information.

The components of health information behavior can be described by Baumann's [8] model of information behavior. Baumann [8] assumes that information strategies always exist as multidimensional constructs consisting of different attitudes and behaviors. Individual factors influencing information behavior provide only moderate validity and should not be considered in isolation from one another, as they ultimately interact with each other. At the center of the model is the information strategy, which outlines the specific information behavior of individuals. This behavior is shaped by three dimensions: (1) the aim of the information search—why the individual seeks information, (2) the type of information sought—what topics the person is interested in, and (3) the sources of information—where the person searches. The aims guide how information is handled. Baumann [8] distinguishes between three aims: surveillance, empowerment, and autonomy. The purpose of surveillance is to gain a comprehensive overview of health issues, be informed about risks and diseases, and better understand the health care system. The aim of empowerment includes that patients specifically seek information so that they can actively participate in their medical care; strengthening their own competence should enable them to better classify and evaluate advice and decisions made by health experts. The aim of autonomy goes one step further: here, the patient strives to take on the role of the therapist themselves, that is, to make health-related decisions competently and independently of medical experts, and also to advise others on health issues.

Baumann [8] summarized the topics related to health and illness into four areas: healthy lifestyle/well-being, diseases and treatment options, the health care system, and health care providers/institutions. The sources of information are grouped into 5 categories: general mass media, magazines, academic media, health experts, and the private social environment. Factors influencing these 3 dimensions are the intensity and variety of health-related information searches, the attention given to health topics, and the evaluation of the relevance of the information received for action [8]. In addition, sociodemographic, socioeconomic, and health-related characteristics of individuals impact their information strategy.

XSL•FO
RenderX

Baumann's [8] components of health information behavior are used to form user groups in this study.

When it comes to vaccination behavior, all 3 dimensions of information behavior play an important role: the frequency of information searches, the information channels used, and the reasons for searching for information. Analyses of Google search volume relating to COVID-19 vaccination intentions found that search queries were higher in regions with a large number of people who were still undecided compared to regions with a higher proportion of vaccination enthusiasts [9]. In addition, people who felt poorly informed about COVID-19 vaccines were more likely to refuse vaccination or be undecided [10]. Increased knowledge was associated with targeted independent research and a preference for information from government institutions and health care professionals. At the same time, not using social media was associated with a lower rate of vaccine refusal [10].

So far, most research findings have examined influencing variables separately, with the complex pattern of information seeking, information retrieval, and their impact on behavior being poorly understood in the context of COVID-19 vaccination. User group analyses can be used to make the complex phenomenon of information retrieval comprehensible, segmentable, and manageable. Two studies, Piltch-Loeb et al [11] and Tan et al [12], analyzed user groups concerning health information behavior related to COVID-19 vaccination. Both studies focused on the third dimension of Baumann's [8] model, the source of information and its relationship with vaccine intentions or status. Using a sample of 3014 Americans, Piltch-Loeb et al [11] defined 6 types based on their use of traditional media, social networks, and vaccination attitudes. For example, users who primarily relied on Facebook and Instagram, as well as nonseekers, were less likely to be vaccinated and held more negative attitudes toward vaccination. Tan et al [12] identified 4 types based on trust in information sources of Singaporeans aged 56-75 years (N=6094). High trust in official sources, such as health authorities and scientific institutions, correlated with a higher likelihood of vaccination, while distrust in all sources was associated with a lower likelihood of vaccination. These findings underscore the importance of trust in the credibility of information sources, indicating that health information behavior can influence vaccination behavior [13,14].

A nuanced understanding of specific user groups—defined by their online information-seeking behavior (frequency of information searches, information channels used, and the reasons for searching for information)—remains lacking. Identifying these groups and their characteristics is crucial for tailoring effective public health communication strategies. Therefore, this study seeks to identify distinct user groups based on their online information behavior regarding COVID-19 vaccination and to investigate the key factors known to influence vaccination uptake in Germany, such as demographics, health literacy, and vaccination readiness (7Cs). The following research questions are explored:

1. Which user groups can be identified concerning information behavior regarding COVID-19 vaccination on the internet within the German population?
2. What relationships exist among sociodemographic characteristics, vaccination readiness, knowledge about COVID-19 vaccination, health literacy, and membership in a user group?
3. To what extent does membership in a user group correlate with vaccination status?

In addition to vaccination readiness and information behavior, other factors can influence vaccination behavior, such as health literacy. Good health literacy enables individuals to find and understand information and make decisions accordingly [15,16]. Furthermore, sociodemographic characteristics such as age, gender, education, income, and immigration history have been highlighted as significant influencing factors [11,15-17]. In particular, a higher level of education and greater health literacy have been associated with more positive attitudes toward vaccination [11,12]. Additionally, specific knowledge about COVID-19 vaccination can also affect vaccination behavior [18]. According to Tan et al [12], factors such as place of residence and apartment size also correlate with vaccination behavior. Self-efficacy expectations can also play a role, as individuals with high self-efficacy are more likely to overcome barriers and choose preventive measures such as vaccination [19]. Conversely, those who feel disproportionately high external pressure to be vaccinated exhibit lower willingness to vaccinate against COVID-19 [20]. This study also aims to examine which of the aforementioned factors influenced vaccination behavior, in particular membership of a user group. Therefore, the third research question is: "To what extent does membership in a user group correlate with vaccination status?" Accordingly, the following hypotheses are proposed:

Hypothesis 1: There is a correlation between identified user groups and sociodemographic characteristics, willingness to vaccinate, knowledge about COVID-19 vaccination, and health literacy.

Hypothesis 2: Belonging to an identified user group predicts vaccination status.

Methods

Study Design

Data were collected using an online questionnaire from November 26 to December 8, 2021, by an external provider of online surveys (Bilendi GmbH). This period coincided with increased media coverage of the first booster vaccinations for COVID-19. The sample was a quota sample, representative of age, gender, and place of residence distribution within a single federal state according to Eurostat 2018. Bilendi invited registered users to participate, offering remuneration for completion. The invitation included information on survey duration, compensation, and a link to the questionnaire. Eligible participants were German-speaking individuals aged 18-75 years.

The questionnaire was divided into three sections: (1) internet use for COVID-19 vaccination information, knowledge about

XSL•FO
RenderX

the vaccination, and health literacy; (2) attitudes toward vaccination, vaccination behavior, and determinants of vaccine willingness; and (3) sociodemographic characteristics.

Ethical Considerations

This research project was approved by the Ethics Committee of Osnabrück University (reference Ethik-46/2021 and Ethik-68/2021). Bilendi ensured data protection and anonymity in accordance with the German Data Protection Act. The participants were informed about the method of data collection, anonymized data transfer, and compensation during panel registration, and were assured that their responses would remain anonymous and not be shared with third parties. Participation was voluntary and could be terminated at any time without penalty. The participants accessed the survey at Respondi Mingle [21] after accepting the terms and conditions and data protection guidelines.

Online Usage Behavior for COVID-19 Vaccination

The frequency of using the internet to search for information regarding COVID-19 vaccination was recorded using the following items: daily, several times a week, once a week, less than once a week, not at all, and use of other sources.

Channels used to search for information: the participants were asked about the most common online information channels. The responses were initially coded as follows: perceived fleetingly, consciously searched for information, no contact, and do not use. Since the options “no contact” and “I don’t use” were similar, these were combined into “consciously searched for information,” “fleetingly perceived,” and “no contact/I don’t use.”

Formats: the participants were asked which formats related to COVID-19 vaccination they encountered online (text articles, audio articles, videos, posts on social networks, posts via messenger services, discussion forums), using a scale of frequently, sometimes, rarely, and never.

Reasons for information-seeking: the reasons for seeking information about COVID-19 vaccination online included obtaining in-depth information about the vaccination, researching information, or exchanging views with like-minded individuals about the benefits and risks of the vaccination. The participants confirmed these statements with “yes” or rejected them with “no.”

Sociodemographic Characteristics

Age was categorized into 4 groups: 18-29 years, 30-44 years, 45-64 years, and 65-75 years. Education was recorded based on school education and professional qualifications according to the demographic standard of the Federal Statistical Office of Germany [22]. Education levels were then clustered into basic, medium, and higher educational attainment using the Comparative Analysis of Social Mobility in Industrial Nations classification [23]. Immigration history was assessed according to the recommended standards in Germany [24,25]. A distinction was made between two categories: (1) immigrants and their (direct) descendants (ie, participants not born in Germany or participants born in Germany with both parents born outside of Germany) and (2) individuals without an immigration history (ie, participants and both parents born in Germany) [26]. Place of residence was categorized into rural areas (up to 5000 inhabitants), small towns (5001 to under 20,000 inhabitants), medium-sized cities (20,000 to 100,000 inhabitants), and large cities (more than 100,000 inhabitants) [27].

Health Literacy

A total of 6 items on general health literacy were developed based on Schaeffer et al’s [28] study. Participants answered the following questions: When you think back to the time before the coronavirus pandemic, how easy was it for you to (1) obtain health-related information, (2) comprehend health-related information, (3) evaluate the content of health-related information, (4) assess the credibility of health-related information, (5) critically question health-related information, and (6) use health-related information as a basis for decision-making (translation from German). The answer options from the Likert scale were coded as follows: 1=very difficult, 2=difficult, 3=neither, 4=easy, 5=very easy. By summing the results, a new variable, “health literacy score,” was created (range from 6 [minimum] to 30 [maximum]). The interitem correlation for this survey is α=.86.

Knowledge

For the knowledge question on COVID-19 vaccination, 5 statements were developed based on information or vaccination myths published on the Bundesministerium für Gesundheit website [29].

Please evaluate the following statements: (1) An mRNA vaccine can alter human DNA. (2) Vaccine reactions such as pain at the injection site, fatigue, headache, and muscle pain are signs that the immune system is responding to the vaccine as intended. (3) Depending on the vaccine, vaccinated individuals are 65%-95% less likely to contract COVID-19 than those who are not vaccinated. (4) After vaccination, I no longer have to follow the rules of conduct (keeping distance, wearing a mask). (5) Full immunity is achieved immediately after vaccination.

A total score (knowledge score from 0 to a maximum of 5 points) was calculated with values for each item as either “not known” (0) or “known” (1).

External Pressure on Vaccination Behavior

The participants were asked whether they felt external pressure to be vaccinated, for example, from their employer or the general population, or whether they felt free to decide for or against vaccination (1 item, possible answers: yes/no).

Self-Efficacy Expectations

Self-efficacy was measured using the current version of the validated General Self-Efficacy Scale by Jerusalem and Schwarzer [30]. The self-assessment procedure comprises 10 items and is used to record general optimistic self-beliefs. The interitem correlation for this survey is α=.92.

Vaccination Readiness and Vaccination Status

The validated 7C Vaccination Readiness Scale by Geiger et al [7] was used to assess readiness to vaccinate, evaluated as recommended by the authors. The scale contains 7 components of readiness to vaccinate: confidence, complacency, constraints,

XSL•FO
RenderX

calculation, collective responsibility, compliance, and conspiracy, each with 3 associated items (range: 1=strongly disagree to 7=strongly agree). The instrument is publicly available [31]. Readiness to vaccinate is represented as the average value across all items of the 7C scale. The internal consistency of the scale was examined using McDonald's omega (ω). The value of ω=0.96 indicates very good internal consistency.

For vaccination status, it was first determined whether the individuals had received 1 or 2 vaccine doses (from BioNTech/Pfizer, Moderna, or AstraZeneca) or 1 vaccine dose from Johnson & Johnson or none to date, which corresponded to the vaccination recommendation in Germany. This information was then summarized into 2 categories—vaccinated or unvaccinated—to represent the participants' vaccination status.

Statistical Analyses

Analyses on online information behavior were only conducted with a group of 889 participants who searched online for information about the COVID-19 vaccination. Absolute and relative frequencies were calculated for the formats and channels used, the reasons for searching for information about COVID-19, and the frequency of searching for information on the internet over the last 4 weeks from completing the questionnaire. A cluster analysis was conducted to determine user groups regarding information behavior about COVID-19 vaccination online. The channels, formats, frequency of information behavior, and reasons for searching for information about the vaccination were used as potential variables for determining user groups through a cluster analysis.

Before the cluster analysis, a Spearman correlation analysis was conducted between the ordinal-scaled variables. Significant, strong correlations were identified between the variables channels and formats (podcast channel and audio contributions: Spearman ρ=0.58, P<.001; WhatsApp and posts via messenger services: Spearman ρ=0.67, P<.001; Facebook and posts in social networks: Spearman ρ=0.63; P<.001). Consequently, the format variables were excluded, which can also be justified because they lacked variation. For the cluster analysis, a hierarchical cluster analysis using complete linkage and Gower distance was applied in RStudio (Posit PBS). After removing cases with missing values and outliers (identified through single-linkage analysis) from the group of 889 participants searching for information online, 778 cases were included in the cluster analysis. The optimal number of clusters was determined using the elbow method and the silhouette method, both of which indicated that 2- and 3-cluster solutions were appropriate. Based on content-related considerations, the 3-cluster solution was selected. A subsequent discriminant analysis showed that 86.4% (672/778) of the originally grouped cases were correctly classified, indicating high quality and accuracy of the cluster analysis and supporting the decision to adopt the 3-cluster model.

Bivariate analyses (chi-square test and ANOVA) were conducted to identify differences between user groups. To analyze the correlations between the user groups and personal characteristics, a binary logistic regression was performed with 3 models for each user group. Model 1 included sociodemographic characteristics (gender, age, education, immigration history, and size of place of residence). Model 2 added willingness to vaccinate, and model 3 incorporated knowledge about vaccination and health literacy.

Additionally, a binary logistic regression analysis was conducted to predict vaccination status. In the first model, the correlation with user groups was tested. The second model included sociodemographic variables, while the third model added potential influencing factors such as vaccination readiness score, knowledge, health literacy, external pressure, and self-efficacy expectation. These analyses were performed using IBM SPSS Statistics, version 29.

Results

Description of the Sample

In total, 1000 people living in Germany completed the questionnaire. Among the participants, 50% (n=500) identified themselves as male and 50% (n=500) as female. The average age was 46.4 (SD 15.5) years, with ages ranging from 18 to 75 years. Further sociodemographic characteristics of the study population revealed that 8.6% (n=86) had a history of immigration. In terms of educational attainment, 12.9% (n=129) had a basic level, 57.8% (n=578) had a medium level, and 27.1% (n=271) had a higher level. More than half of the respondents lived in a medium-sized town or a large city. Overall, 86% (n=860) of the participants had been vaccinated at least once, while 13.3% (n=133) remained unvaccinated at the time of the survey. Attitudes toward vaccination were as follows: 13.7% (n=137) were (somewhat) opposed, 6.9% (n=69) were undecided, and 78.7% (n=787) were (somewhat) in favor. An overview is provided in Table 1.

XSL•FO
RenderX

Table 1. Description of the sample (N=1000).

Characteristic	Value
Age (years), n (%)	
18-29	188 (18.8)
30-44	271 (27.1)
45-64	403 (40.3)
65-75	138 (13.8)
Gender, n (%)	
Male	500 (50)
Female	500 (50)
Immigration history, n (%)	
Individuals without an immigration history	914 (91.4)
Immigrants and their (direct) descendants	86 (8.6)
Education attainment, n (%)	
Basic	129 (12.9)
Medium	578 (57.8)
Higher	271 (27.1)
Prefer not to answer	22 (2.2)
Residence, n (%)	
Rural area	202 (20.2)
Small town	191 (19.1)
Medium-sized city	247 (24.7)
Big city	353 (35.3)
Prefer not to answer	7 (0.7)
Vaccination status, n (%)	
Vaccinated at least once	860 (86)
Not vaccinated	133 (13.3)
Prefer not to answer	7 (0.7)

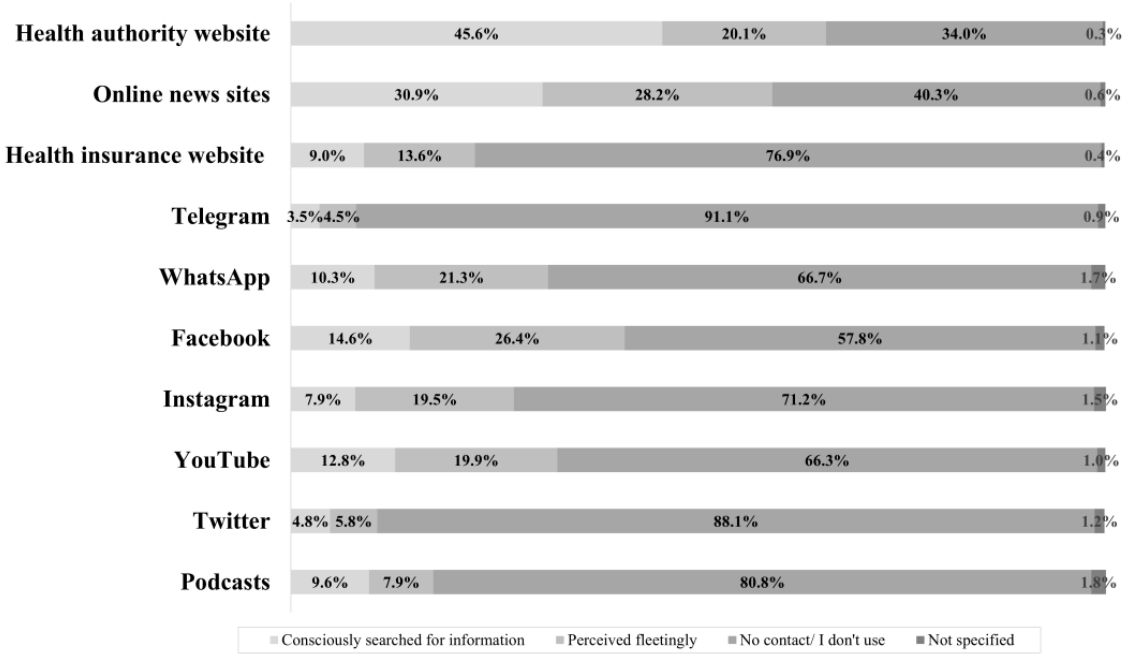
Online Information Behavior

In the past 4 weeks, most respondents (N=1000) reported searching the internet for information on COVID-19 vaccination daily (n=266, 26.6%) or several times a week (n=356, 35.6%). Conversely, 11% (n=110) did not search online at all for the information or used other sources.

Further analyses of information behavior were conducted only with the subsample of 889 online users. Websites of health authorities (n=405, 45.6% consciously searched) and online news sites (n=275, 30.9% consciously searched) were used more frequently than health insurance company websites (80/889, 9%). Social media and messenger services were used less intentionally for information gathering. Facebook, WhatsApp, and Instagram were perceived as less reliable sources, while Telegram and Twitter were used by only a few online users (Figure 1).

XSL•FO
RenderX

Figure 1. Channels used to search for information (n=889).



Text content was the most frequently selected method for learning about COVID-19 vaccination (405/889, 45.6% often and 312/889, 35.1% sometimes). Posts on social media and audio and video content were used less frequently, while discussion forums and messages via messenger services were rarely selected (Multimedia Appendix 1).

Of the 889 online users, 85.9% (n=764) stated that their reason for searching for information on COVID-19 vaccination was to stay updated on new developments. Additionally, 59.4% (n=528) sought to learn more about the vaccination, and 69.4% (n=617) wanted to verify information and research claims. A less substantial reason for using the internet was the intention to persuade others of their views on vaccination through social media posts (131/889, 15.1%). Furthermore, 43.6% (388/889) of respondents reported regularly discussing the advantages and disadvantages of vaccination with others, and half sought information outside of public institutions.

Identified User Groups on Information Behavior Regarding COVID-19 Vaccination on the Internet

Overview

Due to missing values, only 778 participants could be included in the cluster analysis. We identified 3 distinct user groups: “frequent and critical information evaluators,” “*infrequent* and passive recipients,” and “frequent and multichannel, interaction-focused users.” The distribution of user behavior, personal characteristics, and attitudes among these groups is detailed below (Table 2). The bivariate analyses revealed significant differences between the user groups in terms of age, self-efficacy, knowledge, and health literacy.

XSL•FO
RenderX

Table 2. Distribution of sociodemographic characteristics, vaccination behavior, vaccination attitudes, health literacy, and knowledge among the user groups and differences between groups (n=778).

Variable	User group 1 (n=454)	User group 2 (n=222)	User group 3 (n=102)	Differences between groups	*P* value
Gender, n (%)				4.51 (2)[a]	.11
Male	220 (48.5)	125 (56.3)	57 (55.9)		
Female	234 (51.5)	97 (43.7)	45 (44.1)		
Age groups (years), n (%)				12.71 (6)[a]	.048
18-29	85 (18.7)	49 (22.1)	21 (20.6)		
30-44	110 (24.2)	62 (27.9)	35 (34.3)		
45-64	182 (40.1)	91 (41)	35 (34.3)		
65-75	77 (17)	20 (9)	11 (10.8)		
Educational attainment, n (%)				3.53 (4)[a]	.47
Basic	48 (10.7)	25 (11.5)	13 (13.1)		
Medium	250 (55.7)	133 (61)	59 (59.6)		
Higher	151 (33.6)	60 (27.5)	27 (27.3)		
Immigration history, n (%)				3.63 (2)[a]	.16
Immigrants and their (direct) descendants	47 (10.4)	17 (7.7)	5 (4.9)		
Individuals without an immigration history	407 (89.6)	205 (92.3)	97 (95.1)		
Residence, n (%)				5.49 (6)[a]	.48
Rural area	83 (18.3)	49 (22.1)	14 (13.7)		
Small town	88 (19.4)	37 (16.7)	22 (21.6)		
Medium-sized city	122 (24.7)	61 (27.5)	24 (23.5)		
Big city	170 (37.5)	75 (33.8)	42 (41.2)		
Vaccination status, n (%)				3.31 (2)[a]	.19
Vaccinated at least once	410 (90.3)	195 (87.8)	86 (84.3)		
Not vaccinated	44 (9.7)	27 (12.2)	16 (15.7)		
Attitude toward vaccination, n (%)				6.19 (4)[a]	.19
(Rather) negative	42 (9.3)	28 (12.6)	15 (14.7)		
Undecisive	26 (5.8)	17 (7.7)	3 (2.9)		
(Rather) favorable	384 (85)	177 (79.7)	84 (82.4)		
Decision to vaccinate without external pressure, n (%)				1.77 (2)[a]	.41
Yes	382 (85.1)	181 (81.5)	83 (83.6)		
No	67 (14.9)	41 (18.5)	19 (18.6)		
Self-efficacy, mean (SD)	30.44 (5.56)	29.85 (5.06)	32.12 (5.87)	5.64 (2, 721)[b]	.004
Knowledge on COVID-19 vaccination, mean (SD)	4.34 (0.99)	4.09 (1.03)	4.05 (1.10)	6.35 (2, 767)[b]	.002
Health literacy, mean (SD)	21.57 (4.18)	20.29 (4.13)	22.66 (4.64)	12.13 (2, 757)[b]	<.001

[a]Chi-square (*df*).

[b]*F* test (*df*).

Frequent and Critical Evaluators (User Group 1)

A total of 58.4% (n=454) of the participants were categorized into this group. They searched the internet for information about COVID-19 vaccination several times a week (n=199, 43.8%) or daily (n=158, 34.8%) during the 4 weeks prior to completing the questionnaire. They also deliberately used public channels, such as health authority websites (n=240, 52.9%). Social media platforms—such as Facebook (n=46, 10.1%), Instagram (n=24, 5.3%), and Twitter (n=26, 3.5%), as well as messenger services

XSL•FO
RenderX

such as WhatsApp (n=31, 6.8%) and Telegram (n=6, 1.3%)—were rarely used for information. YouTube and podcasts were also infrequently used (n=39, 8.6% and n=40, 8.8%, respectively). Key reasons for internet use regarding COVID-19 vaccination for this group included obtaining detailed information (n=360, 79.3%), verifying information and research claims (n=360, 79.3%), and staying informed on current developments (n=445, 98%).

In the logistic regression analysis, participants older than 45 years with higher education levels were more likely to belong to this user group. Overall, 85% (n=384) of this group were (somewhat) in favor of vaccination, and 90.3% (n=410) had been vaccinated at least once. On average, user group 1 scored higher in knowledge than the overall participant group (mean 4.34, SD 0.99 compared with mean 4.23, SD 1.02) and rated their health literacy higher (mean 21.57, SD 4.18 compared to mean 21.37, SD 4.29). The probability of being assigned to user group 1 was 1.6 times higher for women than for men (*P*=.007). No further significant correlations could be identified with other sociodemographic characteristics, willingness to be vaccinated, knowledge, or health literacy.

Infrequent and Passive Recipients (User Group 2)

This group comprised 28.5% (n=222) of the participants. They searched for information on the internet less frequently than the other 2 user groups, with 20.7% (n=46) searching once a week and 36% (n=80) doing so less than once a week. Members of user group 2 intentionally searched for information less frequently, regardless of the channel. Regarding their reasons for using the internet in relation to COVID-19 vaccination, the only reason that garnered agreement from more than half of the respondents was the desire to stay informed about the latest developments regarding the vaccination (n=136, 61.3%). This user group was characterized by occasional information searches and a passive attitude toward seeking information online about COVID-19 vaccination.

Individuals aged 18-29 years and 30-44 years, particularly those with a medium level of education, were more likely to belong to this user group. Additionally, this group contained a higher proportion of unvaccinated individuals and those undecided about vaccination. Overall, 79.7% (n=177) of the participants in user group 2 were (somewhat) in favor of vaccination, and 87.8% (n=195) had been vaccinated at least once.

Members of this user group rated their health literacy lower than those in the other 2 groups (mean 20.29, SD 4.13), which was also reflected in their knowledge of COVID-19 vaccination (mean 4.09, SD 1.03). In the logistic regression analysis, a significant correlation was found only for user group 2 concerning health literacy and age groups. Higher health literacy was associated with a lower probability of being assigned to user group 2 (odds ratio [OR] 0.93, 95% CI 0.89-0.96; *P*<.001). People aged 65 years and older were less likely to be in user group 2 than 18- to 29-year-olds (OR 0.41, 95% CI 0.21-0.78; *P*=.01).

Frequent and Multichannel, Interaction-Focused Users (User Group 3)

This group accounted for 13.1% (n=102) of the participants. During the 4 weeks prior to completing the questionnaire, 48% (n=49) of individuals searched for information about COVID-19 vaccination on the internet daily, while 43.1% (n=44) did so several times a week. Members of this user group actively sought information from online news sites (n=66, 64.7%) and public channels, such as health authority websites (n=62, 60.8%), as well as social media platforms such as Facebook (n=43, 42.2%), YouTube (n=46, 45.1%), and the WhatsApp messaging service (n=38, 37.3%). They rarely used channels such as Telegram, Twitter, and Instagram. The most common reasons for using the internet regarding COVID-19 vaccination included following current developments (n=102, 100%), obtaining in-depth information about the vaccination (n=98, 96.1%), researching statements and verifying information (n=99, 97.1%), regularly exchanging information with like-minded individuals about the benefits and risks of vaccination (n=91, 89.2%), and seeking information outside of public institutions (n=81, 79.4%). The information behavior of this user group overlaps with that of the frequent and critical information evaluators; however, their choice of diverse channels and intention to exchange information with others was particularly notable.

Individuals aged 30-44 years were increasingly represented in this group. Overall, 82.4% (n=84) of participants in this user group were (rather) in favor of vaccination, and 84.3% (n=86) had been vaccinated at least once. Although members of this group rated their health literacy higher than those in the other two groups (mean 22.66, SD 4.64), they achieved the lowest average score on the knowledge questions (mean 4.05, SD 1.10).

In the logistic regression analysis, a modest association was observed between health literacy and assignment to this user group. As health literacy increased, the probability of being assigned to user group 3 slightly increased (OR 1.08, 95% CI 1.02-1.13; *P*=.01). No statistically significant correlation was found between other characteristics and the vaccination variables.

The regression models can be found in the Multimedia Appendix 2.

Predictors of Vaccination Status

A logistic regression was performed to identify the factors influencing vaccination status (Table 3). Model 1 indicated that the user group is not related to vaccination status. When personal characteristics were added in model 2, education and place of residence emerged as significant influencing factors. Model 3 demonstrated that vaccination readiness—defined as personal attitudes toward COVID-19 vaccination—was the most important predictor of vaccination status. For each additional point on the Vaccination Readiness Scale, there is a 9.72 (95% CI 5.68-16.66; *P*<.001) times higher chance of being vaccinated. Another statistically significant influencing factor remains education; compared to people with a basic level of education, those with a medium level of education have a 6.78 (95% CI 1.76-26.15; *P*=.005) times higher chance, and people with a

XSL•FO
RenderX

higher level of education have an 11.90 (95% CI 2.39-59.18; *P*=.002) times higher chance of being vaccinated. In addition, men have a 65% (95% CI 14%-86%; *P*=.02) lower chance of being vaccinated than women. Knowledge about vaccination, health literacy, external pressure, and self-efficacy did not play a significant role in the overall model.

Table 3. Logistic regression to predict vaccination status (reference value: vaccinated at least once; N=1000).

Variable	Model 1		Model 2		Model 3	
	P value	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)
User group						
User group 1	—[b]	1	—	1	—	1
User group 2	.31	0.75 (0.44-1.30)	.61	0.86 (0.49-1.52)	.56	1.34 (0.49-3.67)
User group 3	.27	0.67 (0.34-1.35)	.44	0.75 (0.37-1.55)	.77	1.21 (0.34-4.23)
Age group (years)						
18-29	—	—	—	1	—	1
30-44	—	—	.93	1.03 (0.52-2.03)	.54	1.44 (0.45-4.65)
45-64	—	—	.14	1.67 (0.84-3.33)	.07	2.99 (0.92-9.73)
65-75	—	—	.06	2.65 (0.95-7.39)	.62	1.68 (0.21-13.36)
Gender						
Female	—	—	—	1	—	1
Male	—	—	.32	0.77 (0.46-1.29)	.02	0.35 (0.14-0.86)
Education attainment						
Basic	—	—	—	1	—	1
Medium	—	—	.06	1.97 (0.97-4.01)	.01	6.78 (1.76-26.15)
Higher	—	—	.002	4.01 (1.69-9.53)	.002	11.9 (2.39-59.18)
Immigration history						
Immigrants and their (direct) descendants	—	—	—	1	—	1
Individuals without immigration history	—	—	.61	0.8 (0.34-1.88)	.21	2.46 (0.60-10.07)
Residence						
Rural area	—	—	—	1	—	1
Small town	—	—	.54	1.25 (0.61-2.58)	.10	0.32 (0.08-1.24)
Medium-sized city	—	—	.33	1.4 (0.71-2.78)	.81	0.86 (0.24-3.06)
Big city	—	—	.01	2.47 (1.23-4.95)	.80	1.17 (0.34-3.98)
Vaccination readiness (scale)	—	—	—	—	<.001	9.72 (5.68-16.66)
Knowledge score	—	—	—	—	.92	0.98 (0.60-1.58)
Health literacy score	—	—	—	—	.16	0.93 (0.83-1.03)
Decision to vaccinate without external pressure						
Yes	—	—	—	—	—	1
No	—	—	—	—	.41	1.43 (0.61-3.38)
Self-efficacy (score)	—	—	—	—	.79	0.99 (0.92-1.07)

[a]OR: odds ratio.

[b]Not available.

XSL•FO
RenderX

Discussion

Principal Findings and Comparison With Prior Work

In our study, we observed different information behavior regarding the COVID-19 vaccination, indicating not only the use and nonuse of such information but also the role of internet- and social media–based information. According to Baumann's [8] model of information behavior, the groups were formed based on search frequency, motivation for searching, and channels used. The 3 user groups identified through cluster analyses are clear examples of cumulative strategies. The frequent and critical information evaluators (user group 1) rely on internet-based information of public authorities, whereas for the interaction-focused users (user group 3), social media–based information plays an important role. Furthermore, the infrequent and passive recipients reflect population groups (user group 2) who use information on COVID-19 vaccination less frequently. The observed types were similar to those described by Tan et al [12] and Piltch-Loeb et al [11], suggesting a degree of consistency in the results. Note that different variables were used in these studies to form the clusters and identify influencing factors. Frequent and critical information evaluators exhibited similar information behavior to the user groups "legacy" and "traditional omnivores" in Piltch-Loeb et al's [11] study and "pro-formal selective" and "broad-trust" in Tan et al's [12] study. Each of these user groups increasingly relied on local, traditional, and public channels from authorities to search for information, while making less use of social media and messaging services. In terms of vaccination behavior and attitudes, individuals in these groups were generally more willing to be vaccinated and demonstrated higher vaccination rates [11]. People with a strong intention to vaccinate and a positive attitude toward vaccination tend to use official websites more frequently, engage in more thorough information-seeking, and conduct searches more often [15,16].

There are parallels between the infrequent and passive recipients and the "nonseekers" in Piltch-Loeb et al's [11] study or "broad distrust" in Tan et al's [12] study. These groups rarely, if ever, actively searched for information and either did not use any channels or lacked trust in them. Consequently, these groups exhibited a low probability of vaccination and an ambivalent attitude toward COVID-19 vaccination.

The frequent and multichannel, interaction-focused users shared similarities with the "omnivore" + "broad social media users" and the "legacy and Facebook/Instagram users" [11]. Individuals in these user groups sought information about COVID-19 vaccination from both traditional public sources, such as government authorities, and increasingly from social media platforms such as Facebook and Instagram, as well as messaging services like WhatsApp [11]. Additionally, these user types tended to exhibit a higher likelihood of being unvaccinated and often held an undecided or negative attitude toward vaccination [11,12]. This trend may be linked to the observation that individuals who rely more on social media platforms for information rather than official websites generally have lower vaccination rates and a more negative view of vaccination [15,16]. The quality of information on these social media sources is often heterogeneous, and misinformation can spread rapidly, leading to uncertainty, skepticism, and rejection [32,33].

Frequent and critical information evaluators predominantly consisted of participants older than 44 years, who increasingly sought information about COVID-19 vaccination from traditional public and official sources [11]. In contrast, respondents aged 18-29 years were more likely to belong to the infrequent and passive recipients group. Individuals in this age bracket tend to engage in minimal or no active searching for information; when they do seek information, they adopt a targeted and specific approach [8,11].

The results of the analysis indicate that people with higher health literacy (user groups 1 and 3) searched for information on the internet more frequently than people with lower health literacy (user group 2). These are particularly groups who state that the reason for their search is to find out about current developments and to verify information. The study by Moffett et al [9] showed that people who are undecided about vaccination are more likely to conduct online research. This could not be confirmed in our study. However, a kind of uncertainty aspect may also have played a role among the frequent users identified here.

Notably, no clear correlation was identified between the user groups and vaccination status. Consequently, the determinants of information seeking identified by Baumann [8]—frequency, channels, and motivations underlying information-seeking activities—do not seem to be the primary determinants in this context. Instead, preliminary evidence suggests that the quality of the information accessed and the trust placed in it may influence vaccination behavior—variables that were not addressed in this investigation. Furthermore, only a few social differences could be observed, namely in gender and education. The vaccination readiness emerged as the strongest predictor of vaccination behavior in the analysis presented, highlighting the significance of individual beliefs and attitudes. These findings are consistent with previous studies that emphasized risk perception and social influence on vaccination behavior, which are components of the 7C Vaccination Readiness Scale used in this study [7,20].

Limitations

Overall, when interpreting the results, it is important to consider that the study's sample exhibited a higher vaccination rate of 86% (860/1000) compared to the 70% vaccination rate reported by the Robert Koch Institute for all citizens in Germany at the time of the survey [5]. Furthermore, the method of participant selection—the sample's distribution—can also explain the discrepancies between the study's findings and existing theories regarding the factors influencing information-seeking and vaccination behavior. The survey was conducted online by an external provider, which limited participation to individuals with internet access and registration with that provider. This limitation may impact representativeness, as those who are online are more likely to search for health information. Additionally, the health literacy score was based on subjective assessments, potentially leading to bias due to overestimation by the respondents. Moreover, no causal conclusions can be drawn from the cross-sectional design.

XSL•FO
RenderX

Recommendations

Several recommendations can be made for health communication regarding COVID-19 vaccination that may also apply to future novel infectious diseases. Identifying distinct types of online information use allows for a more nuanced understanding of information behavior. Targeted health communication to different user groups may not influence the vaccination behavior, but could lead to an informed vaccination decision, for example, understanding the risks and benefits. The following recommendations support a targeted approach for providing relevant health information to the identified user groups. For frequent and critical information evaluators, more evidence-based and detailed information could be provided on government and scientific platforms. Infrequent and passive recipients could be more effectively reached through low-threshold, easily understandable formats. In contrast, interactive formats, such as those found in social media or discussion forums, could better engage frequent and multichannel, interaction-focused users. Because this group obtains information from platforms that historically exhibit lower information accuracy, public health agencies should implement targeted interventions. For example, collaboration with trusted microinfluencers and platform-native content creators could help coproduce short, evidence-based messages that match the stylistic norms of each channel.

The 3 information-seeking groups identified in our study reflect fundamental behavioral patterns in how individuals obtain and process health information. Because these clusters were derived from usage-based variables, they have the potential to serve as a generalizable framework for other health-related issues. However, the transferability of the cluster structure is not automatic. For each new health topic, researchers should collect the same behavioral variables, conduct a confirmatory cluster or latent-class analysis to test whether the original 3-class solution fits the new data, and examine how cluster membership relates to the specific health behavior of interest. When the clusters are confirmed, targeted communication strategies that proved effective for COVID-19 vaccination can be adapted to the new context.

Longitudinal studies are required to assess whether communication strategies that are tailored to each group produce sustained changes in knowledge, attitudes, and subsequent health actions. Finally, the efficacy of the proposed communication strategies should be evaluated through randomized controlled trials, allowing for evidence-based optimization of future communication campaigns.

Data Availability

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

Funding

Publication fees were supported by the publication fund of the Deutsche Forschungsgemeinschaft.

Authors' Contributions

Conceptualization: LS
Data curation: LS
Formal analysis: LL (lead), LS (supporting)
Investigation: LS
Methodology: LS (lead), BB (supporting), LL (supporting)
Project administration: LS (lead), BB (supporting)
Validation: LS
Writing – original draft: LL (lead), LS (supporting)
Writing – review & editing: LS (lead), BB (equal), LL (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Formats used for information (N=889).
[PNG File , 97 KB - infodemiology_v6i1e82221_app1.png]

Multimedia Appendix 2
Logistic regression analyses on the characteristics of the user groups.
[DOCX File , 18 KB - infodemiology_v6i1e82221_app2.docx]

References

XSL•FO
RenderX

1. Schmidt L, Götz NA, Hannemann N, Babitsch B. Disparities in COVID-19 related health literacy, knowledge, and the assessment of the measures taken in Germany: a cross-sectional study. Z Gesundh Wiss 2023:1-12 [FREE Full text] [doi: 10.1007/s10389-023-01827-2] [Medline: 36714071]
2. Federal Ministry of Health. Vaccinations. 2025. URL: https://www.bundesgesundheitsministerium.de/en/topics/vaccinations.html
3. Rieger T, Schmidt-Petri C, Schröder C. Attitudes toward mandatory COVID-19 vaccination in Germany. Dtsch Arztebl Int 2022;119(19):335-341 [FREE Full text] [doi: 10.3238/arztebl.m2022.0174] [Medline: 35384836]
4. Robert Koch-Institut, Bundeszentrale für gesundheitliche Aufklärung, Leibniz-Zentrum für Psychologische Information und Dokumentation, Science Media Center, Bernhard Nocht Institute for Tropical Medicine, Yale Institute for Global Health. COSMO explorer: impfbereitschaft (mit Geimpften). COSMO. URL: https://projekte.uni-erfurt.de/cosmo2020/web/explorer/
5. Robert Koch Institut. COVID-19-impfungen in Deutschland. URL: https://robert-koch-institut.github.io/COVID-19-Impfungen_in_Deutschland/
6. Maier BF, Wiedermann M, Burdinski A, Klamser PP, Jenny MA, Betsch C, et al. Germany's fourth COVID-19 wave was mainly driven by the unvaccinated. Commun Med (Lond) 2022;2:116 [FREE Full text] [doi: 10.1038/s43856-022-00176-7] [Medline: 36124059]
7. Geiger M, Rees F, Lilleholt L, Santana AP, Zettler I, Wilhelm O. Measuring the 7Cs of vaccination readiness. Eur J Psychol Assess 2022;38(4):261-269. [doi: 10.1027/1015-5759/a000663]
8. Baumann E. Auf der Suche nach der Zielgruppe? Das Informationsverhalten über Gesundheit und Krankheit als Grundlage erfolgreicher Gesundheitskommunikation. In: Böcken J, Braun B, Amhof R, Schnee M, editors. Gesundheitsmonitor 2006. Gesundheitsversorgung und Gestaltungsoptionen aus der Perspektive von Bevölkerung und Ärzten. Gütersloh: Bertelsmann Stiftung; 2006:117-153.
9. Moffett KW, Marshall MC, Kim JEC, Dahlen H, Denison B, Kranzler EC. Analyzing google COVID-19 vaccine intent search trends and vaccine readiness in the United States: panel data study. Online J Public Health Inform 2024;16:e55422 [FREE Full text] [doi: 10.2196/55422] [Medline: 39073868]
10. Zhelyazkova A, Kim S, Klein M, Prueckner S, Horster S, Kressirer P. COVID-19 vaccination intent, barriers and facilitators in healthcare workers: insights from a cross-sectional study on 2500 employees at LMU University Hospital in Munich, Germany. Vaccines (Basel) 2022;10(8):1231. [doi: 10.3390/vaccines10081231] [Medline: 36016119]
11. Piltch-Loeb R, Silver D, Kim Y, Abramson D. COVID-19 vaccine information seeking patterns and vaccine hesitancy: a latent class analysis to inform practice. J Public Health Manag Pract 2024;30(2):183-194. [doi: 10.1097/PHH.0000000000001834] [Medline: 38038353]
12. Tan M, Straughan PT, Cheong G. Information trust and COVID-19 vaccine hesitancy amongst middle-aged and older adults in Singapore: a latent class analysis approach. Soc Sci Med 2022;296:114767 [FREE Full text] [doi: 10.1016/j.socscimed.2022.114767] [Medline: 35144226]
13. Ozawa S, Paina L, Qiu M. Exploring pathways for building trust in vaccination and strengthening health system resilience. BMC Health Serv Res 2016;16(Suppl 7):639 [FREE Full text] [doi: 10.1186/s12913-016-1867-7] [Medline: 28185595]
14. Pennisi F, Genovese C, Gianfredi V. Lessons from the COVID-19 pandemic: promoting vaccination and public health resilience, a narrative review. Vaccines (Basel) 2024;12(8):891 [FREE Full text] [doi: 10.3390/vaccines12080891] [Medline: 39204017]
15. Bosle C, Orth B, Reibling N, Merkel C, Muschalik C, Von Rüden U. Gesundheitsinformationsverhalten und Gesundheitskompetenzen zur COVID-19-Schutzimpfung von Menschen in Deutschland ? Befunde der CoSiD-Studie. [Health information behaviour and health literacy on COVID-19 vaccination of the general population in Germany-findings from the CoSiD study. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2022;65(12):1289-1298 [FREE Full text] [doi: 10.1007/s00103-022-03617-9] [Medline: 36329209]
16. Schäfer M, Stark B, Werner AM, Mülder LM, Reichel JL, Heller S, et al. Informiert = Geimpft? Das Informationsverhalten und die COVID-19-Impfentscheidung bei Studierenden. Publizistik 2023 Jan 26;68(1):37-68. [doi: 10.1007/s11616-023-00779-7]
17. Meng L, Masters NB, Lu PJ, Singleton JA, Kriss JL, Zhou T, et al. Cluster analysis of adults unvaccinated for COVID-19 based on behavioral and social factors, national immunization survey-adult COVID module, United States. Prev Med 2023;167:107415 [FREE Full text] [doi: 10.1016/j.ypmed.2022.107415] [Medline: 36596324]
18. Wulkotte E, Schmid-Küpke NK, Neufeind J, Wichmann O. COVID-19-Impfquotenmonitoring in Deutschland als Einwanderungsgesellschaft (COVIMO-Fokuserhebung): report. Rober Koch Institut. 2022. URL: https://www.rki.de/DE/Themen/Infektionskrankheiten/Impfen/Forschungsprojekte/abgeschlossene-Projekte/COVIMO/Downloads_Uebersetzungen/COVIMO-9-Report-Deutsch.html [accessed 2025-04-02]
19. Lippke S, Schüz B. In: Haring R, editor. Modelle gesundheitsbezogenen Handelns und Verhaltensänderung. Springer Berlin Heidelberg: Gesundheitswissenschaften; 2019.
20. forsa Politik- und Sozialforschung GmbH. Befragung von nicht geimpften Personen zu den Gründen für die fehlende Inanspruchnahme der Corona-Schutzimpfung. Ergebnisbericht. Bundesministeriums für Gesundheit 2021:1-40 [FREE Full text]

XSL•FO
RenderX

21. Respondi Mingle. URL: https://mingle.respondi.co.uk/ [accessed 2026-04-21]
22. Hoffmeyer-Zlotnik JH, Hanefeld U, Herter-Eschweiler R, Mohr S. Demographische Standards : eine gemeinsame Empfehlung des ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V., der arbeitsgemeinschaft sozialwissenschaftlicher institute e.V. (ASI) und des Statistischen Bundesamtes. Wiesbaden: Statistisches Bundesamt 2010;53(3):618-619. [doi: 10.1007/s11577-001-0100-8]
23. Lechert Y, Schroedter JH, Lüttinger P. Die Umsetzung der Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus-Zusatzerhebung 1971 und die Mikrozensen 1976-2004. Social Science Open Access Repository 2006:1 [FREE Full text]
24. Schenk L, Bau AM, Borde T, et al. Mindestindikatorensatz zur Erfassung des Migrationsstatus - Empfehlungen für die epidemiologische Praxis. Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung. 2020 Sep. URL: https://edoc.rki.de/handle/176904/1061 [accessed 2026-04-30]
25. Kajikhina K, Koschollek C, Sarma N, Bug M, Wengler A, Bozorgmehr K. Recommendations for collecting and analysing migration-related determinants in public health research. J Health Monit 2023;8(1):52-72 [FREE Full text] [doi: 10.25646/11144] [Medline: 37064417]
26. Fachkommission der Bundesregierung zu den Rahmenbedingungen der Integrationsfähigkeit. Gemeinsam die Einwanderungsgesellschaft gestalten. Bericht der Fachkommission der Bundesregierung zu den Rahmenbedingungen der Integrationsfähigkeit. 2020. URL: https://www.fachkommission-integrationsfaehigkeit.de/resource/blob/1786706/1880170/917bc43f62136ed26ecef8125a4c9cdf/bericht-de-artikel-data.pdf?download=1
27. Bundesinstitut für Bau-, Stadt- und Raumforschung. Laufende Stadtbeobachtung? Raumbeobachtung. Stadt- und Gemeindetypen in Deutschland. 2020. URL: https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/gemeinden/StadtGemeindetyp/StadtGemeindetyp.html
28. Schaeffer D, Vogt D, Berens E, Hurrelmann K. Gesundheitskompetenz der Bevölkerung in Deutschland: Ergebnisbericht: Universität Bielefeld. Fakultät für Gesundheitswissenschaften: Universität Bielefeld; 2016.
29. BMG. URL: http://zusammengegencorona.de/ [accessed 2026-04-20]
30. Jerusalem M, Schwarzer R. SWE. Skala zur Allgemeinen Selbstwirksamkeitserwartung: [Verfahrensdokumentation, Autorenbeschreibung und Fragebogen]. Trier: Open Test Archive; 2003.
31. Vaccination Readiness. URL: https://www.vaccination-readiness.com/ [accessed 2026-04-21]
32. Link E. Gesundheitskommunikation mittels Gesundheitsportalen und Online-Communitys. In: Rossmann C, Hastall MR, editors. Handbuch der Gesundheitskommunikation: Kommunikationswissenschaftliche Perspektiven [Book in German]. Wiesbaden: Springer Fachmedien Wiesbaden; 2019:159-170.
33. Döring N. Die Bedeutung von Videoplattformen für die Gesundheitskommunikation. In: Rossmann C, Hastall MR, editors. Handbuch der Gesundheitskommunikation: Kommunikationswissenschaftliche Perspektiven [Book in German]. Wiesbaden: Springer Fachmedien Wiesbaden; 2019:171-185.

Abbreviations

OR: odds ratio

Edited by C Baur; submitted 11.Aug.2025; peer-reviewed by F Yeh; comments to author 29.Dec.2025; revised version received 13.Mar.2026; accepted 13.Mar.2026; published 29.May.2026.

Please cite as:
Liebner L, Babitsch B, Schmidt L
Online Information Behavior Regarding COVID-19 Vaccination and Its Association With Vaccination Behavior Based on Cluster Analysis of User Groups: Cross-Sectional Study
JMIR Infodemiology 2026;6:e82221
URL: https://infodemiology.jmir.org/2026/1/e82221
doi:10.2196/82221
PMID:

©Lea Liebner, Birgit Babitsch, Lisa Schmidt. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 29.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Weight Loss Medication Discourse: Mixed Methods Analysis of US-Based Facebook Posts

Elizabeth Dennard[1*], MPH; Katrina Makres[1*], MPH; Amrutha Alibilli[1*]; Vidur Jain[1*]; Anne Wang[1*]; Roger Abim-Karmon[2*]; Shaniece Criss[2*], ScD; Yulin Hswen[1*], ScD; Quynh C Nguyen[3*], PhD; Alexandria Ratzki-Leewing[4,5*], PhD; Rozalina G McCoy[4,5,6*], MD; Thu T Nguyen[1,5], ScD

[1]Deparment of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, 4254 Stadium Drive, College Park, MD, United States

[2]Furman University, Greenville, SC, United States

[3]National Institutes of Health, Bethesda, MD, United States

[4]University of Maryland School of Medicine, Baltimore, MD, United States

[5]University of Maryland Institute for Health Computing, North Bethesda, MD, United States

[6]Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, United States

[*]these authors contributed equally

Corresponding Author:
Elizabeth Dennard, MPH
Deparment of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, 4254 Stadium Drive, College Park, MD, United States

Abstract

Background: Despite the documented clinical efficacy of weight loss medications, few large-scale mixed methods studies have captured the experiences of individuals taking these medications.

Objective: This study aims to examine dominant themes and public narratives about the perceived benefits, challenges, identity-based experiences, and the broader sociocultural framing of glucagon-like peptide-1 receptor agonist use on Facebook.

Methods: We used CrowdTangle to collect 2500 US-based Facebook posts from January 1, 2022, to May 31, 2024. Bidirectional encoder representations from transformers topic modeling was used to identify latent topics, followed by a qualitative thematic analysis.

Results: An analysis of 2500 Facebook posts revealed distinct thematic patterns across five unique subcategories: (1) healthy weight management, (2) weight loss medications, (3) discussion specific to semaglutide, (4) discussions specific to Ozempic, and (5) public figures and media. Thirteen overarching themes emerged through this thematic analysis, with the most common themes including weight management programs (n=702), neutral content related to weight loss medication use (n=466), harms of weight loss medication use (n=373), benefits of weight loss medication use (n=329), and access to these medications (n=188).

Conclusions: Social media provides critical insights into individual experiences with pharmacotherapy and weight management. Enhanced public education may optimize the safe use of glucagon-like peptide-1 receptor agonists for weight loss.

(JMIR Infodemiology 2026;6:e89732) doi:10.2196/89732

KEYWORDS

weight loss; obesity; digital health; internet; research

Introduction

In the United States, more than one-third of adults are currently affected by obesity [1], a figure projected to approach one-half by 2030 [2]. Obesity is a major driver of type 2 diabetes, cardiovascular disease, and other leading causes of death in the United States [3]. Glucagon-like peptide-1 receptor agonists (GLP-1 RAs) have emerged as a highly effective pharmacologic approach to treat obesity and type 2 diabetes. These agents mimic the incretin hormone glucagon-like peptide-1 (GLP-1) to stimulate insulin secretion, delay gastric emptying, and suppress appetite. Multiple large-scale randomized trials, including STEP, SUSTAIN, and SURMOUNT [4,5], have demonstrated clinically meaningful weight loss and cardiometabolic benefits in individuals with obesity [6,7]. In response, regulatory approvals have expanded, and demand has surged; recent estimates suggest that 1 in 8 US adults has used a GLP-1 RA, with 40% citing weight loss as their primary goal [8].

XSL•FO
RenderX

Even though more than 45% of US adults meet eligibility criteria for GLP-1 RA use, uptake remains uneven due to limitations posed by high costs, insurance restrictions, and the still evolving evidence on long-term outcomes [9]. Discontinuation is common and often followed by weight regain, especially in the absence of sustained lifestyle support [10]. Therapeutic plateaus, unrealistic expectations, and adverse events may further undermine adherence. Frequently reported adverse events include gastrointestinal symptoms—nausea, vomiting, abdominal pain, constipation, and/or diarrhea—often experienced within the first month of GLP-1 RA use and during dose escalation [11]. Other less frequently assessed adverse events of GLP-1 RA use among people with overweight or obesity include gastroesophageal reflux disease, gastroenteritis, pancreatitis, and gallbladder-related outcomes (cholelithiasis and cholecystitis) [12]. Overall, current findings support the widespread use of GLP-1 RAs to treat obesity, diabetes, and other cardiovascular-kidney-metabolic disorders, while also highlighting frequently reported adverse events that require further investigation to inform patient counseling and adverse effect mitigation strategies.

Moreover, the cultural framing of these medications has shifted rapidly, with celebrity endorsements and esthetic motivations fueling widespread media attention and off-label use [13,14]. Deeper insight into real-world lived experiences and expectations of GLP-1 RA users is essential for informing equitable, person-centered treatment and education strategies. While clinical trials and guideline developments have centered on efficacy and safety, much less research has been conducted on how these medications are understood outside formal health care settings. Recently, social media platforms have become rich venues for user-driven health narratives—providing diverse perspectives that remain largely unexplored in the GLP-1 RA literature.

Previous studies have leveraged social media data to examine discourse surrounding pharmaceutical advertising and user-reported behavioral experiences [15,16]. However, findings from these studies focused on advertising and specific user experiences, leaving broader patterns of GLP-1 RA discourse unexplored. Analyzing this discourse at scale requires tools capable of preserving narrative nuance while uncovering broader thematic patterns. This study leverages bidirectional encoder representations from transformers topic (BERTopic), a transformer-based topic modeling approach, paired with qualitative content analysis to examine public narratives about GLP-1 RA use on Facebook. By applying this mixed method strategy to a large corpus of user-generated posts, we aimed to elucidate dominant themes related to perceived benefits and harms, access challenges, identity-based experiences, and the broader sociocultural framing of these medications.

Methods

Data Source

Over time, Facebook has evolved into one of the most globally used platforms, attracting over 3 billion annual users [17]. Approximately 71% of US adults report using the platform with disproportionately higher usage among older adults and women compared to other platforms [18]. This pattern aligns with GLP-1 use among women who are more likely to be currently taking a GLP-1 RA [19]. Facebook posts were collected using CrowdTangle, a publicly available tool that enables systematic exploration of content across social media platforms [20]. CrowdTangle provides data from public accounts (7 million Facebook pages, groups, and verified profiles) and allows users to refine results using parameters such as keywords, timeframes, and geolocation [20]. Currently available tools for researchers include Meta Content Library and Content Library API [21]. For this study, data were collected using keyword-based filters, yielding information such as post content, image URLs, captions, dates of posts, and engagement metrics.

Data Preparation

Data from English-language posts originating from US-based users were collected between January 1, 2022, and May 31, 2024. GLP-1 RA use has surged since 2022. Tirzepatide was approved by the US Food and Drug Administration (FDA) for diabetes management in May 2022 and was quickly used for off-label obesity management before receiving FDA approval in June 2023 [22-24]. To examine the rise in GLP-1 RA discussions online, we used data collected through CrowdTangle from January 2022 to May 2024; CrowdTangle was discontinued in August 2024 [21]. The keyword list (Table 1) was developed based on a combination of prior literature and clinical expertise. A total of 64,202 posts were retrieved. Text preprocessing included the removal of symbols, hyperlinks, HTML tags, and other special characters. Key terms were identified using regular expression methods.

XSL•FO
RenderX

Table . List of keywords used for capturing posts from Meta.

ID	Keywords
1	Weight loss drugs
2	Ozempic
3	Weight loss medication
4	Weight loss treatments
5	Weight management
6	Mounjaro
7	Tirzepatide
8	Weight loss medications
9	Miracle drug
10	Miracle weight loss
11	Miracle jab
12	Ozempy
13	Semaglutide
14	Wegovy
15	GLP-1[a]
16	GLP-1 agonist
17	GLP-1s
18	Weight management drugs
19	Weight management meds
20	Weight management medications
21	Skinny pen
22	Skinny jab
23	Skinny jots
24	Weight loss shot
25	Miracle diet drug
26	Liraglutide
27	Saxenda
28	Victoza

[a]GLP-1: glucagon-like peptide-1.

Quantitative Analysis

To identify key topics across Facebook posts, we applied BERTopic, a modern topic modeling approach that uses advanced language models to group similar content based on meaning [25]. Unlike traditional techniques such as latent Dirichlet allocation, which rely on word frequency patterns, BERTopic captures deeper contextual relationships between words, allowing for more meaningful and nuanced topic discovery [26]. The process began by converting each post into a numerical representation of its meaning using a pretrained language model. These representations were then simplified using uniform manifold approximation and projection (UMAP), which helps organize complex data. Next, a clustering algorithm known as hierarchical density-based spatial clustering of applications with noise grouped the posts into clusters based on similarity [27,28]. To interpret each topic, we used a modified version of term frequency–inverse document frequency (TF-IDF) called class-based TF-IDF, which highlights the most distinctive words within each group [25]. This technique improves the clarity of each topic and sets BERTopic apart from other topic modeling tools.

BERTopic was implemented in Python (Python Software Foundation) using the BERTopic library, with default settings for dimensionality reduction and clustering [25]. Several exploratory runs were conducted by adjusting parameters to evaluate the stability and interpretability of the results. Four parameters in total were adjusted: UMAP dimensionality reduction parameters (n_neighbors, n_components) and BERTopic parameters (nr_topics, min_topic_size). In the final model, we used n_neighbors=13, n_components=9, nr_topics=5, and min_topic_size=100. The final model was selected based on semantic coherence and the clarity of topic keywords.

XSL•FO
RenderX

Following deduplication, the dataset consisted of 50,013 unique Facebook posts. These topics provided an initial thematic structure that helped organize the large Facebook dataset and informed the subsequent qualitative content analysis.

Qualitative Content Analysis

A thematic analysis was conducted on a random sample of 500 Facebook posts for each of the 5 topics identified using BERTopic, for a total of 2500 posts. Thematic analysis is a widely used qualitative method that allows for the analysis of free-response data using an established codebook [29]. First, the authors implemented an inductive approach by reviewing 20 posts from each subcategory and discussing emerging themes to establish an understanding of the dataset through familiarization [29]. Then, the authors constructed themes through pattern formation and identification, assessed thematic diversity, and tested the codebook's operational abilities. The final codebook comprised 13 overarching themes, each containing unique subthemes (Table 2). Once the codebook was finalized, 2 authors independently conducted postscreening of the 2500 posts. Conflicts were resolved through discussion between the two reviewers until consensus was reached based on detailed justifications. The finalized themes and subthemes were then applied to guide a comprehensive thematic analysis of the sample, offering insight into the range of topics discussed in public discourse surrounding GLP-1 RA medications on Facebook.

XSL•FO
RenderX

Table . Codebook for thematic analysis.

Codebook ID	Themes	Subthemes
1	Natural and whole foods	• Multiple foods: health outcome • Multiple foods: nonhealth outcome • Protein: health outcome • Protein: nonhealth outcome • Fiber: health outcome • Fiber: nonhealth outcome • Whole grains: health outcome • Whole grains: nonhealth outcome • Vegetables: health outcome • Vegetables: nonhealth outcome • Oils: health outcome • Oils: nonhealth outcome • Other: health outcome • Other: nonhealth outcome
2	Physical activity	• Health outcome • Nonhealth outcome
3	Vegetarian or vegan	• Health outcome • Nonhealth outcome
4	Supplements	• Health outcome • Nonhealth outcome
5	Weight loss medications—promotion	• Weight loss • Health benefits • Both weight loss and health benefits • Celebrities
	Weight loss medications—neutral	• Education • Updates on regulations
6	Weight loss medications—harms	• Harming health • Exploiting insecurities • Too much weight loss • Temporary results of weight loss medications • Celebrities
7	Speculation about weight loss medication use	• N/A[a]
8	Preference for unmedicated weight loss	• N/A
9	Access to weight loss medications	• Availability of medication • Cost • Insurance • Lawsuits and investigations
10	Weight management programs	• Selling or promoting weight management programs • Advice
12	Hormone therapy	• N/A
13	Research studies	• Weight loss • Health benefits • Both weight loss and health benefits
14	Business expansion (products capitalizing on the rise in use of weight loss medications)	• N/A
99	Not relevant	• N/A

XSL•FO
RenderX

[a]N/A: not applicable.

Ethical Considerations

This study did not involve human participants and was conducted in accordance with ethical standards determined by the University of Maryland College Park Institutional Review Board (2072551 - 1). Given the nature of the study, examination of publicly available social media posts did not involve direct interaction or communication with human subjects. Additionally, posts were anonymized to uphold user privacy.

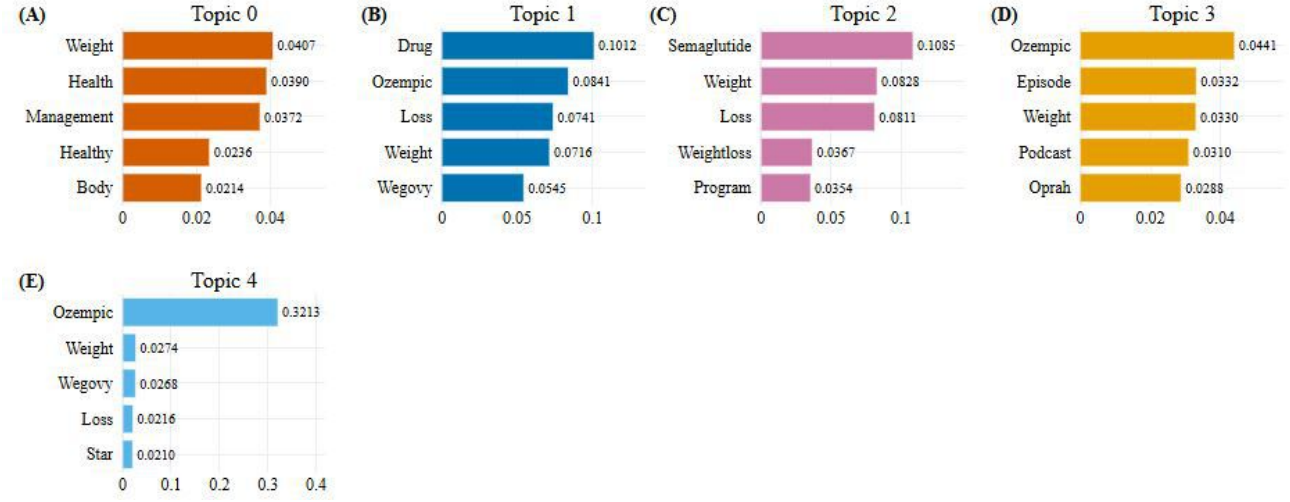
Results

Overview

BERTopic was used to analyze 50,013 posts referencing GLP-RAs and initially explored several configurations. A target topic size of 6 with a minimum topic size of 100 posts was chosen to prevent overly granular clusters. One topic was automatically flagged as a low-density outlier, yielding a total of five topics. The 5-topic model was then selected based on the clarity of cluster separation visualized in the intertopic distance map (Multimedia Appendix 1), which is a 2D UMAP projection of topic embeddings representing relative semantic distances between topics. There was no overlap between the 5 topics, and comparable distances were demonstrated, supporting the validity of the model.

Topics were categorized using the top keywords identified using class-based TF-IDF scores (Figure 1). BERTopic modeling yielded 5 distinct topics: healthy weight management, weight loss medications, semaglutide, public figures, and Ozempic. These broad topics serve as a foundation for understanding the dominant narratives surrounding weight loss drugs on social media. They provide a structured lens to interpret public discourse, identify recurring patterns, and uncover underlying concerns and motivations.

Figure 1. Topic word scores. This figure illustrates the top 5 keywords for each topic (A-E) based on class-based term frequency–inverse document frequency scores. These terms represent the most representative words within each topic.



Topic 0 (healthy weight management) emerged as a prominent theme characterized by keywords such as “weight,” “health,” and “management.” Posts in this category focused on lifestyle-based approaches to combating obesity. Topic 1 (weight loss medications) captured broader discussions about pharmaceutical interventions, featuring user experiences, medication comparisons, and commentary on side effects. Topic 2, characterized by the keyword “semaglutide,” reflected in-depth discourse on the clinical use and effectiveness of this specific medication. Topic 3 (public figures) centered on the role of celebrities and influencers in shaping perceptions of weight loss treatments. Finally, topic 4 (Ozempic) focused on its off-label use for weight loss, concerns around access and affordability, and growing public interest.

Then, a content analysis was conducted on a random sample of 2500 Facebook posts. As depicted in Table 2, 13 distinct themes emerged during the qualitative analysis, each comprising various subthemes. The most popular themes across the sample included weight management programs, neutral weight loss medication posts, harms of weight loss medications, promotion of weight loss medications, and access to weight loss medications. Table 3 provides a detailed overview of each theme and subtheme, accompanied by illustrative quotations to highlight specific examples of Facebook user statements. A total of 89 posts were excluded from the final analysis because they lacked substantive references to weight management or the population of interest.

XSL•FO
RenderX

Table . Examples of weight management posts by theme and subtheme.

Themes and subthemes	Post
Natural and whole foods and supplements	
Multiple foods	"When it comes to maintaining a healthy and balanced diet, it's important to remember the significance of incorporating a variety of nutrient-rich foods. One excellent addition to your daily intake is nuts and berries. These small yet powerful ingredients are packed with essential vitamins, minerals, and antioxidants that offer numerous health benefits...Incorporating nuts and berries into your daily diet can enhance your overall nutritional intake and contribute to a healthier lifestyle. So, remember to include them in your meals, snacks, or even as toppings for yogurt, cereal or however you like!"
Protein	"Top fat loss nutrition tip I wish I knew earlier: get 30 - 40 grams protein at every meal Why? Glad you asked top 5 reasons to eat high protein foods. Muscle growth: protein is essential for building and repairing muscles, which is crucial for overall strength and fitness. Weight management: high-protein diets can help you feel full for longer periods, reducing overall calorie intake and supporting weight loss or maintenance. Metabolism boost: protein has a higher thermic effect compared to fats or carbs, meaning your body burns more calories digesting it, potentially aiding in weight management..."
Fiber	"Trying to lose weight? Fiber is your friend! High-fiber foods can help to promote a feeling of fullness and satiety by delaying digestion and adding bulk to meals. For instance, starches like chickpeas (found in hummus) may help with weight management. Try it with some pita slices and fresh veggies. It's delicious!"
Whole grains	"Is 'oat-zempic' a substitute for Ozempic? Experts say this trend does not mimic the way these drugs used for weight loss work and could have negative health impacts."
Vegetables	"Eating fruits and vegetables offers a myriad of health benefits. Packed with essential nutrients, fiber, and vitamins, they support overall well-being. Studies link them to a lower risk of chronic diseases, improved weight management, and better digestive health. Share your favorite fruits or veggies!"
Other	"Hydrating doesn't have to be boring! Spice up your water with refreshing mint, orange, and cucumber for the perfect summer drink that will keep you healthy and energized!"
Supplements	"Spirulina is a nutrient-dense blue-green algae that offers several health benefits. Here are some key advantages of incorporating spirulina into your diet: Rich in Nutrients Spirulina is a powerhouse of nutrients, including protein, vitamins (B-complex vitamins, especially B12), minerals (iron, calcium, magnesium), and essential fatty acids. High Protein Content It is an excellent source of plant-based protein, making it a valuable addition to vegetarian and vegan diets. Antioxidant Properties Spirulina contains antioxidants like phycocyanin, which helps protect cells from damage caused by free radicals. Weight Management The protein and nutrient content in spirulina may contribute to a feeling of fullness, potentially aiding in weight management. Blood Sugar Control Preliminary research indicates that spirulina may have a positive effect on blood sugar levels, making it potentially beneficial for individuals with diabetes or those at risk..."
Physical activity	
Health outcome	"Elevate Your Health with Cardio! Burn calories, boost mood, and strengthen your heart. From weight management to increased endurance, these workouts have got your back (and heart) covered!"

XSL•FO
RenderX

Themes and subthemes	Post
No health outcome	"Beta-hydroxybutyrate, or BHB, is a key ingredient in SiselTHIN. BHB's are naturally occurring salts found in the body during ketosis. Getting into ketosis can be difficult to achieve through diet alone. Exogenous (provided externally) BHB's can provide the critical fuel needed during a keto diet and will help you get into ketosis faster and stay there so you can achieve your weight management goals. SiselTHIN is part of our January promo pack. Have you gotten yours yet? Get your packs and save while the deal is still available."
Diet (vegan or vegetarian)	
Health outcome	"Did you know plant-based diets can contribute to weight management and overall well-being? This Lent, embark on a journey of self-discovery with BombayFoodJunkies. Discover a world of flavorful and healthy vegan options on our menu! Ready to try plant-based food for 40 days? Book a table now and enjoy yummy, guilt-free meals!"
Weight loss medications—promotion	
Weight loss	"Unlock the secret to feeling full and satisfied! Discover the hormone that signals to your brain that your stomach is full, leading to decreased hunger and increased feelings of satisfaction. This hormone, called GLP-1, can even help your body use more energy from stored fat."
Health benefits	"Ozempic protects kidneys, boosts survival in diabetes patients with chronic kidney disease."
Both	"Drugs like Wegovy and Ozempic have demonstrated success in treating obesity and diabetes. Can they treat depression too?"
Celebrities	"Whoopi Goldberg revealed she used weight loss drug Mounjaro to help her shed some pounds after weighing nearly 300 pounds in late 2021."
Weight loss medications—neutral	
Education	"Here's what psychiatrist Allison Young, MD, says patients should be asking their doctors if they think new prescription weight loss drugs might help with weight gain due from antidepressants or other drugs prescribed for mental illness."
Regulatory	"Based on the new, unprecedented results, Eli Lilly announced plans to seek fast-track approval from the U.S. Food and Drug Administration to advertise and sell the drug for weight management."
Weight loss medications—harms	
Harming health	"Ozempic weight loss has patients needing to get rid of sagging skin. The primary reason for this new challenge with sagging skin is the loss of skeletal muscle."
Exploiting insecurities	"In the first few months of 2023, it seems like society has taken some serious steps backward when it comes to our relationship with our bodies. There are new headlines about Ozempic and weight loss nearly every day, and many of us are still reeling from the American Academy of Pediatrics' January recommendation that kids as young as 12 be prescribed weight loss drugs (and that those as young as 13 be prescribed surgery)... In our latest article, we talked with experts about what parents can say and do to help their kids build resilience and push back against the idea that we should all be striving to be thinner. Give it a read for actionable advice on protecting your kiddo from today's rampant fatphobia"
Too much weight loss	"A great supplement to consider if you are experiencing rapid weight loss due to certain medications. Ask one of the Rocky Mountain Pharmacystaff. They'll fill you in on the details."
Temporary results	"Sun Kim cautions against using popular weight-loss drugs as a quick fix: "We have to think of obesity as a chronic disease," she explains. "Someone with high blood pressure doesn't expect to stop taking blood pressure medications once normal blood pressure has been achieved. If you stop taking semaglutide, you generally regain the weight back."
Celebrities	"Lizzo fans criticize her for taking Ozempic to lose weight."
Access to weight loss medications	

XSL•FO
RenderX

Themes and subthemes	Post
Availability	"For months, multiple diabetes drugs like Ozempic have been in high demand with short supply."
Cost	"Wegovy's high price and the huge increase in people taking it companies reconsidering when and how to reimburse use of such treatments to prevent a steep spike in health insurance costs."
Insurance	"America's screwed-up healthcare system is preventing millions from receiving a new "life-saving" weight-loss drugs like Wegovy and Ozempic."
Lawsuits and investigations	"Federal prosecutors accuse 36-year-old Isis Navarro Reyes of using the social media platform to illegally sell bacteria-contaminated prescription medications, including Ozempic, that she had obtained from Central and South America."
Weight management programs	
Selling or promoting programs	"If you currently use Saxenda or Ozempic contact us for more information or to apply! If you participated in a Saxenda or Ozempic study with us in the last 6 months, you will not be able to participate in this session. Thank you for understanding."
Advice	"Most of us want to eat better. But how should you start? Here's a helpful article that offers some simple ways to develop a healthier diet without depriving yourself and explains how working with a nutritionist can help support your weight loss or weight management goals."
Research studies or statistics	
Weight loss	"In this trial, adolescents with obesity assigned to weekly subcutaneous semaglutide plus lifestyle intervention had a greater reduction in BMI than those who received lifestyle intervention alone."
Health benefits	"Novo Nordisk's Ozempic slows diabetic kidney disease progression in trial"
Both	"A late-stage trial of Eli Lilly's weight loss drug Zepbound showed promising results in treating sleep apnea symptoms in obese adults, the company announced Wednesday, the latest findings suggesting popular weight loss drugs have the potential to treat conditions other than obesity or type 2 diabetes."
Business expansion	
None	"Rapid weight loss can be for your body confidence, but it can lead to some unexpected changes in your facial appearance. We're here to help!"
Hormone therapy	
None	"Hormone harmony: boost estrogen levels and balance hormones PCOS Support: Aid is reducing symptoms commonly related with polycystic ovary syndrome such as irregular menstrual cycle Weight Management: Chromium has been shown to regulate and support blood sugar levels and in turn help in fat loss Skin & Hair Strength: L-Glutathione is an antioxidant that can help reduce wrinkles and improve complexion with added biotin for hair and nail health (Click here to order"
Speculation	
None	"Curious about Ashley Graham's stunning weight loss journey? Wondering if it was through a dedicated diet plan, the buzzed-about Ozempic, or perhaps weight loss surgery? Dive into the details and discover the truth behind her transformation. Learn more:"
Preference for unmedicated weight loss	

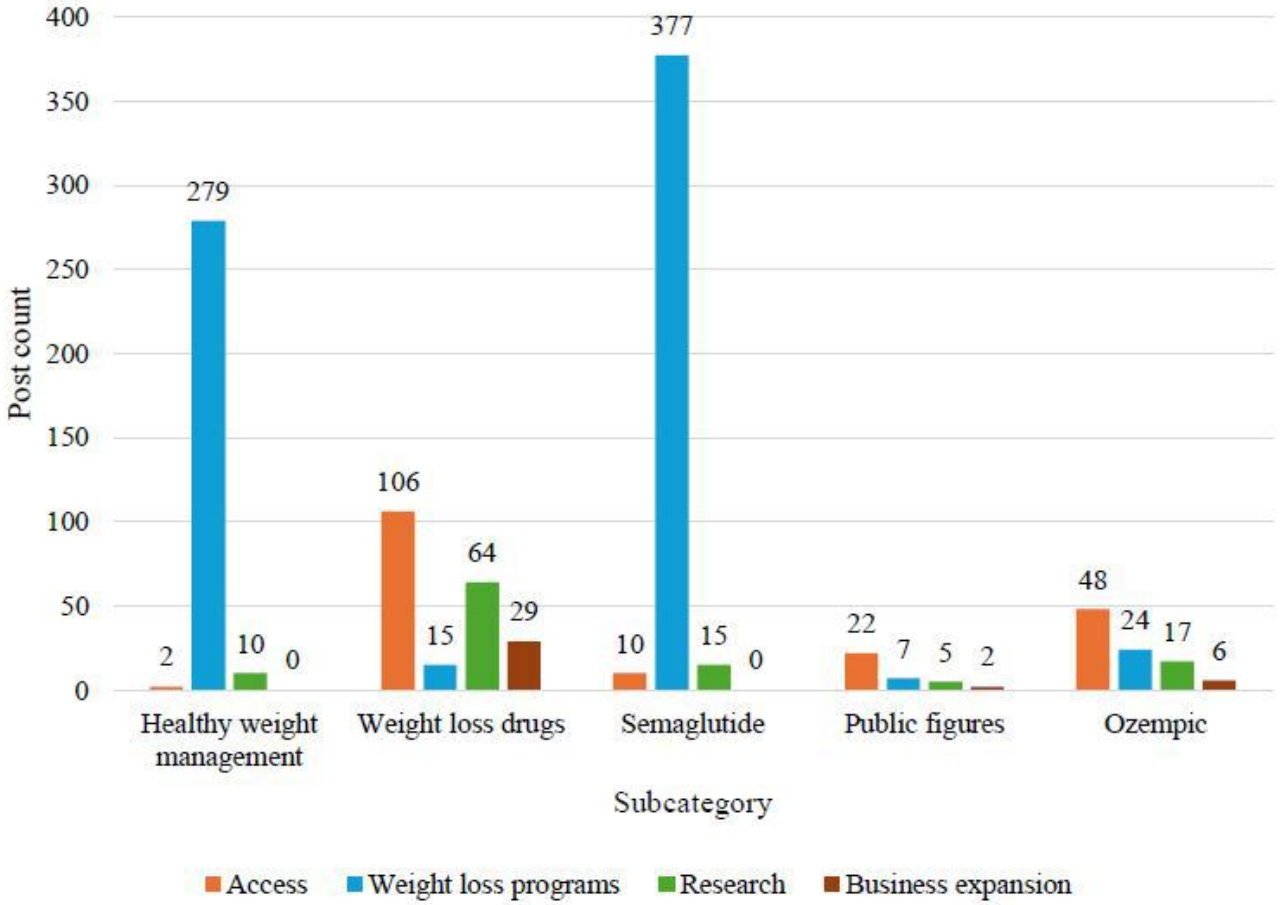
XSL•FO
RenderX

Themes and subthemes	Post
None	"You don't need to manage your weight. Your body is already doing this for you. Anyone selling weight management as a strategy for health promotion is selling a product that will result in an eventual decline of physical or mental health or both. I appreciate that you might want to attempt to manage your body. I understand that this is a product sold, repeatedly. In different packages and as various plans. Please save your money. The body cant discern weight management strategies from famine. No matter how fancy the package, it is not fancy business. And it is only about business. ($$$) Not yours...What are you doing this weekend? *my page is educational and does not represent nutrition therapy or medical care."

Quantitative Content Analysis of the 5 Topics

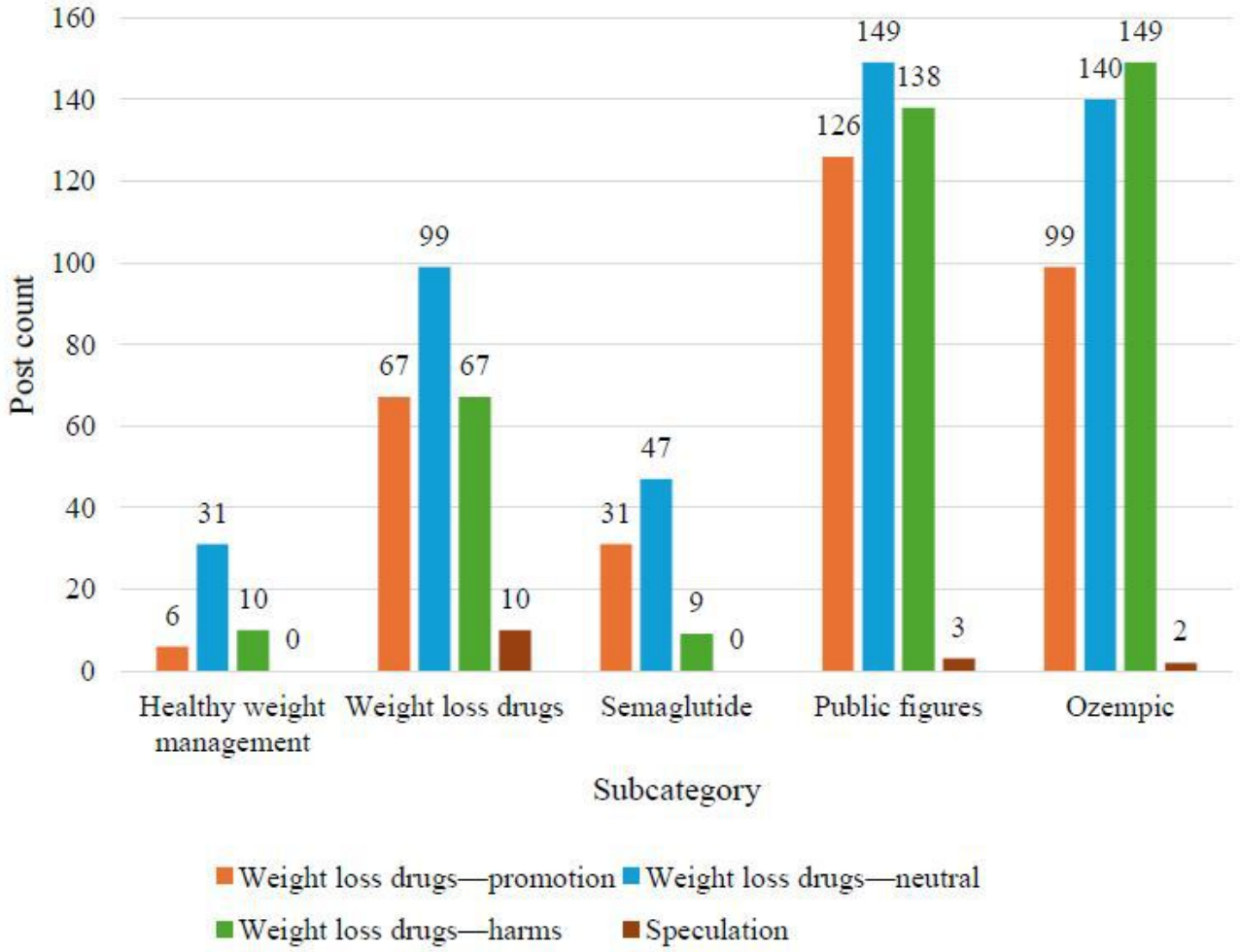
The five BERTopic-derived topics included (1) healthy weight management, (2) weight loss medications, (3) semaglutide, (4) public figures and media, and (5) Ozempic. Across the 5 topics, several overarching themes emerged, such as natural and whole foods and supplements, physical activity, vegan or vegetarian diets, attitudes toward weight loss medications, speculation about medication use, preference for nonpharmacologic options, access to weight loss medications, weight loss programs, commercial expansion, hormone therapy, and research studies. Additionally, post counts for each theme are described across the 5 BERTopic-derived topics (Figures 2-4). The 5 most prevalent themes were as follows: "weight management programs," "weight loss medications—neutral," "weight loss medications—harms," "weight loss medications—promotion," and "access to weight loss medications." "Weight management programs" referred primarily to program promotion, supplement sales, and company advertisements. "Weight loss medications—neutral" included informational content and regulatory updates. "Weight loss medications—harms" referred to negative aspects such as adverse events and exploitative messaging. "Weight loss medications—promotion" focused on perceived benefits, and "access to weight loss medications" included issues related to cost, insurance, and availability. Illustrative examples of posts for each theme and subtheme are presented in Table 3.

Figure 2. Weight management posts by subcategory and theme (group 1: access, weight loss programs, research, and business expansion). This bar graph illustrates the total number of weight management posts across subcategories (x-axis) and themes (represented by the colors of the bars). The x-axis categorizes posts by theme (access, weight loss programs, research, and business expansion) and subcategory (healthy weight management, weight loss medications, semaglutide, public figures, and Ozempic), while the y-axis displays the corresponding post counts.



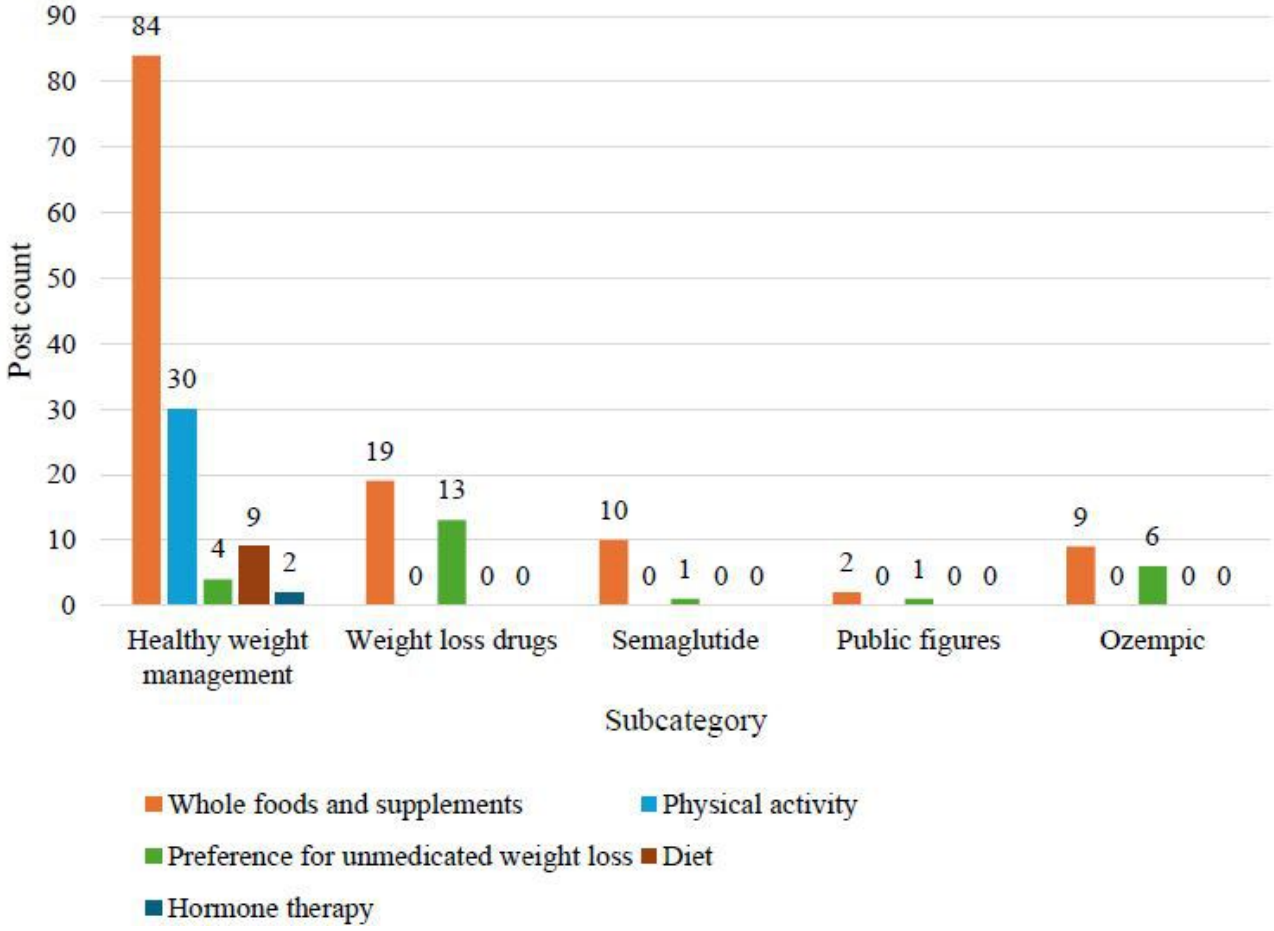
XSL•FO
RenderX

Figure 3. Weight management posts by subcategory and theme (group 2: weight loss medications—promotion, weight loss medications—neutral, weight loss medications—harms, and speculation). This bar graph illustrates the total number of weight management posts across subcategories (x-axis) and themes (represented by the colors of the bars). The x-axis categorizes posts by theme (weight loss medication promotion, neutral, harms, and speculation of weight loss medication use) and subcategory, while the y-axis displays the corresponding post counts.



XSL•FO
RenderX

Figure 4. Weight management posts by subcategory and theme (group 3: natural and whole foods and supplements, physical activity, preference for unmedicated weight loss, diet, and hormone therapy). This bar graph illustrates the total number of weight management posts across subcategories (x-axis) and themes (represented by the colors of the bars). The x-axis further categorizes posts by theme (natural and whole foods and supplements, physical activity, diet, preference for unmedicated weight loss, and hormone therapy), while the y-axis provides the corresponding post counts.



Healthy Weight Management Subcategory

The most common themes in this group were "weight management programs," including 279 (55.8%) posts; "natural and whole foods and supplements," including 84 (17%) posts; and "weight loss medications—neutral," including 31 (6%) posts. Most "weight management programs" posts (n=245) promoted programs, while fewer (n=34) offered health advice. Most posts referenced health outcomes, such as diabetes, obesity, and cardiovascular disease. "Weight loss medications—neutral" posts included information, expert opinions, or federal guidance on weight loss medications. This category also included notable references to physical activity (n=30).

Weight Loss Medications Subcategory

In this subcategory, the most common themes included "access to weight loss medications," comprising 106 (21.2%) posts; "weight loss medications—neutral," comprising 99 (20%) posts; "weight loss medications—promotion," comprising 67 (13%) posts; and "weight loss medications—harms," comprising 67 (13.4%) posts. Access posts referenced availability (n=46), cost (n=15), insurance coverage (n=26), and lawsuits or investigations (n=19). Neutral posts included educational content, with fewer posts mapping to the "updates on regulations" subtheme. Promotional posts highlighted weight loss (n=30), health benefits (n=9), both weight loss and health benefits (n=18), and celebrity posts (n=10), allowing for further discussion regarding the weight loss benefits of these medications. With identical theme counts, posts focused on harms raised concerns about health risks, exploiting insecurities, excessive weight loss, temporary results, and celebrity posts.

Semaglutide Subcategory

The distribution of posts was dominated by "weight management programs," including 377 (75.4%) posts, followed by "weight loss medications—neutral," including 47 (9%) posts and "promotion," including 31 (6%) posts. Most posts related to this topic promoted paid or supplement-based weight management programs (n=373), consistent with the commercial focus observed across the full dataset. Neutral posts covered educational content and webinar information (n=45) or regulatory updates (n=2), while promotional posts centered around weight loss success (n=23) and weight loss combined with health benefits (n=7).

Public Figures and Media Subcategory

This subcategory involved weight loss medication posts, with "neutral" accounting for 149 (29.8%) posts, "harms" accounting for 138 (27.6%) posts, and "promotion" comprising 126 (25.2%) posts. "Neutral" posts included educational facts or webinar information (n=148) and FDA regulation updates (n=1).

XSL•FO
RenderX

Harm-related posts frequently referenced celebrities (n=114). Some discussed how weight loss medications harm one's health (n=18), while others noted the exploitation of insecurities (n=6). This category also included posts in which celebrities were critical of weight loss medications or public complaints regarding celebrity weight loss medication usage. Promotional posts highlighted celebrities (n=115), including reports of positive results. Additional subthemes from positive posts included weight loss (n=6), health benefits with weight loss (n=3), and general health benefits (n=2). There was an almost even divide of posts about celebrities in the harm and promotion themes regarding weight loss medications. The public figures and media subcategory had the highest number of posts surrounding celebrities, with a total of 229 posts.

Ozempic Subcategory

The most common themes were "weight loss medications—harms," including 149 (29.8%) posts; "neutral," including 140 (28%) posts; and "promotion," including 99 (20%) posts. Harms of weight loss medications included adverse events (n=51), exploiting insecurities (n=5), excessive weight loss (n=1), temporary results (n=2), and celebrity posts (n=90). The discussion of the negative impacts of Ozempic on Facebook was more common than that of its positive effects. The majority of neutral weight loss medication posts were linked to education (n=134), with only a few mapping to the regulatory subtheme (n=6). For positive posts linked to weight loss medications, celebrity posts were identified most frequently, while posts highlighting weight loss, health benefits, or both appeared less frequently. Unlike other subcategories, Ozempic included the most posts linked to the harms of weight loss medications (n=149).

Themes Across the Data Sample: Descriptions and Example Quotes

Weight Management Programs

Posts under this theme frequently promoted weight management programs, advertised weight loss medications, and shared general advice. These posts often included persuasive sales messages:

> *Transform your journey, sculpt your dreams!! Introducing the ultimate duo: Emsculpt and semaglutide/Tirzepatide For EXTREME fast major and amazing results!! Elevate your results and redefine your contours and step into a confidence that's uniquely yours. Contact us/text us to learn more.*

Conversely, other posts from established programs were more neutral and often informative, offering general advice on nutrition, exercise, and supplements while supporting consumer weight management journeys: "Nutrition tip: choose whole foods over processed ones for a nutrient-rich diet that supports your weight management goals."

Weight Loss Medications—Neutral

These posts shared educational information about GLP-1 RAs. From expert webinars to FDA updates, the neutral posts advanced factual knowledge about weight loss medications: "FDA approves weight-loss drug Wegovy to reduce heart-disease risks."

Weight Loss Medications—Harms

Various posts highlighted the negative physical and psychological consequences of using weight loss medications. Common physical adverse health outcomes included gastrointestinal distress:

> *Ozempic and Wegovy can cause gastrointestinal issues such as constipation and nausea. Can increasing your fiber intake help?*

In addition, long-term weight loss efficacy was questioned by users discussing excessive or temporary weight loss: "Although Ozempic and Wegovy have been found to help persons with obesity lose weight, they are not necessarily a foolproof approach for long-term weight loss and weight management." Common psychological harms included exploiting consumer insecurities by targeting body image, social comparison, and conformity. Additionally, increased media coverage and celebrity endorsements of weight loss medications may heighten the pressure on consumers to conform to Western beauty standards and perpetuate fatphobia.

Weight Loss Medications—Promotion

Users commonly discussed weight loss and other health benefits associated with GLP-1 RA use. Various posts highlighted the duality of both weight loss and improved holistic health outcomes:

> *23lbs down and starting month 3 this week!! Compounded Semaglutide I feel INCREDIBLE! My autoimmune issues are at an all time low! My skin is clearer. And I had to buy NEW clothes!!!*

Access to Weight Loss Medications

The final most prevalent theme included access to weight loss medications, which encompassed GLP-1 RA availability, affordability, lawsuits, and investigations. Many users were concerned about GLP-1 RA shortages: "Can't get my Ozempic and am out. Anyone know which pharmacy has the 2 in stock?" Furthermore, the increasing costs presented another accessibility barrier: "Zepbound has a much lower sticker price than its rival, Novo Nordisk's Wegovy." Additionally, insurance coverage was described across numerous posts: "Some lawmakers say the United States cannot afford to keep a decades-old law that prohibits Medicare from paying for new weight loss drugs." Finally, lawsuits and investigations often discussed regulatory pricing and other guidelines that may impact public access to the medications: "Bernie Sanders called a vote to subpoena the chief of Novo Nordisk's U.S. division over pricing of Ozempic and Wegovy."

Overall, social media discussions about weight management revealed a complex landscape with a myriad of themes and a wide spectrum of public sentiment. Social media users shared their weight loss successes and perceived health benefits, as well as adverse physical and psychological events they experienced. The data also revealed exposure to marketing, including a variety of promotional programs and advertisements. Public figures and celebrities were commonly featured in

XSL•FO
RenderX

messaging related to endorsements and critical discourse on weight loss medications. Finally, posts that discussed medication access underscored significant concerns about GLP-1 RA shortages, high costs, limited insurance coverage, and ongoing lawsuits that could impact weight loss medication pricing and availability.

Discussion

Principal Findings

Mixed methods analysis of Facebook posts discussing GLP-1 RA medications revealed 13 distinct thematic patterns across 5 unique topics. The most prevalent themes included weight management programs, weight loss medications, and access to weight loss medications. Healthy weight management posts primarily featured weight loss programs, natural and whole foods, supplements, and physical activity. In the weight loss medications subcategory, access and educational posts were most frequently identified. Semaglutide-related posts were dominated by program promotions, while public figures and media primarily focused on weight loss medications, with a high percentage of celebrity influence in both harm and promotional contexts. Ozempic posts most frequently discussed harms and education, with fewer posts promoting the use of weight loss medications. Finally, online discourse was heavily shaped by commercial interests, personal experiences, and celebrity influence, highlighting a need for evidence-based messaging around weight loss interventions.

Our findings highlight novel themes across identified topics while also confirming current sentiments regarding weight loss medications among social media users. Facebook users often shared personal experiences, motivations for GLP-1 RA use, and research supporting the use of weight loss medications. These sentiments align with existing studies that describe the benefits of GLP-1 RA use for sustainable weight loss [30]. In addition to physical health benefits, studies have documented the mental health benefits of weight loss medications. A recent systematic review reported associations between GLP-1 RAs and reductions in depressive symptoms, suicidal ideation, alcohol use, substance use, and binge eating behaviors [31]. However, some adverse psychiatric outcomes have been documented in pharmacovigilance analyses, with varying effects across different types of weight loss medications [31].

Beyond clinical outcomes, prior research highlights the stigma surrounding GLP-1 agonists, rooted in cultural ideals favoring “disciplined” weight loss—an image embedded in fatphobic ideologies [13]. High costs, limited availability, and lack of insurance coverage exacerbate class-based fatphobia, disproportionately affecting lower-income individuals with fewer options for pharmacological treatment [13]. Scholars emphasize that sustainable cultural change requires initiatives beyond medical interventions, including public health campaigns and educational resources that foster body diversity, empathy, and an understanding of the complexity of obesity as a multifaceted health condition [13].

The widespread promotion of GLP-1 RAs has been heavily influenced by celebrity culture where the pressure to maintain one’s physical appearance and weight is critical. Weight management culture in the United States is entangled with long-standing trends of fatphobia, body image pressures, and evolving medical approaches to obesity, as evidenced by increased usage of GLP-1 RAs [13]. Celebrities’ early adoption of GLP-1 RAs amplifies the desirability of these medications. High-profile celebrities like Oprah Winfrey have acknowledged their use of weight loss medications, while others are widely suspected of doing the same [32]. The growing visibility of celebrity use of weight loss medications can further influence public attitudes toward obesity, fatphobia, and body image [13].

Strengths and Limitations

By analyzing 2500 randomly selected Facebook posts, this study offers insights into public perceptions and information-sharing practices surrounding weight loss medications. By analyzing diverse subcategories, we captured a wide range of content, including educational, commercial, and celebrity influences. This study addresses existing data gaps, such as social media discourse and the cultural framing of GLP-1 RA medications, which are relevant to public health communication, patient education, and policy development. Additionally, this study employed BERTopic, a modeling technique that captures contextual and semantic relationships in text and is well suited to informal social media discourse. Compared to traditional models, BERTopic enhances topic coherence, dynamic reduction, and outlier detection. After removing duplicates and outliers, we randomly sampled 2500 posts for detailed analysis, ensuring a diverse dataset spanning educational content, personal experiences, and commercial promotion.

However, several limitations merit attention. First, the analysis was limited to Facebook and may not reflect discussions occurring on other social media platforms. Additionally, Crowdtangle data did not provide information on the type of accounts that originated from the included Facebook posts. Second, while the keyword list was informed by prior literature and author expertise, it was not exhaustive, potentially missing relevant online discussions. Third, we lacked demographic and engagement metrics (eg, likes and shares), which could have offered additional insights for interpreting influence. Finally, the analysis was restricted to English-language posts, limiting generalizability to non-English-speaking communities.

Conclusions

This study underscores the complexity of social media discourse surrounding GLP-1 RAs and the intersection of commercial, clinical, and cultural narratives. As weight loss medications gain increasing popularity across diverse populations, it is imperative to ensure that their use is not only effective but also safe. This mixed methods study highlights the depth and complexity of social media discourse surrounding GLP-1 RAs as an interweaving narrative of commercial, clinical, and cultural perspectives. Our analysis reveals that GLP-1 RA use is shaped by individual and social contexts, including personal values, lived experiences, expectations, and stigma. Future studies should build upon this descriptive and exploratory study to examine long-term patterns of GLP-1 RA use, benefits, and side effects across diverse populations. Additionally, future studies can consider social media platform diversity and how

XSL•FO
RenderX

social media users navigate sources of evidence-based information to counteract the spread of misinformation online. Research is also needed to support understanding of user experiences with non-FDA-approved products (such as compounded drugs) or medications obtained outside of traditional health care settings (such as directly from the manufacturer), as data on adverse effects would not be fully captured in electronic health records or claims. Our analysis lays important groundwork for future work focused on understanding GLP-1RAs through a more complete, person-centered lens—a critical component of evidence-based medicine.

Acknowledgments

Generative artificial intelligence tools were not used at any stage in the preparation of this manuscript.

Funding

Research reported in this publication was supported by the National Institute on Minority Health and Health Disparities (R01MD015716 [TTN]). This research was supported (in part) by the Intramural Research Program of the National Institutes of Health (NIH) (grant number ZIA NR000043; QCN Nguyen). Drs. McCoy, Ratzki-Leewing, and Nguyen are investigators at the University of Maryland-Institute for Health Computing, which is supported by funding from Montgomery County, Maryland, and The University of Maryland Strategic Partnership: MPowering the State, a formal collaboration between the University of Maryland, College Park, and the University of Maryland, Baltimore The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services.

Disclaimer

The contributions of the National Institutes of Health author(s) are considered works of the US Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the National Institutes of Health or the US Department of Health and Human Services.

Data Availability

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

Authors' Contributions

Substantial contributions to the conception, design, data acquisition, analysis, and interpretation were completed by KM, ED, AA, VJ, AW, RA-K, SC, YH, QCN, AR-L, RGM, and TTN. All authors contributed significantly to manuscript development, drafted the work, reviewed it critically, and submitted final approval of the manuscript for publication. Finally, all authors agreed to be accountable for all aspects of the work, ensuring that questions related to the integrity of any part of the work are appropriately investigated.

Conflicts of Interest

TTN is a member of the *JMIR Infodemiology* Editorial Board. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Intertopic distance map of topic embeddings generated by BERTopic with circles topic proportional to size.
[JPG File, 14 KB - infodemiology_v6i1e89732_app1.jpg]

References

1. Adult obesity facts. US Centers for Disease Control and Prevention. 2025. URL: https://www.cdc.gov/obesity/adult-obesity-facts/index.html [accessed 2026-04-22]
2. GBD 2021 US Obesity Forecasting Collaborators. National-level and state-level prevalence of overweight and obesity among children, adolescents, and adults in the USA, 1990–2021, and forecasts up to 2050. Lancet 2024 Dec;404(10469):2278-2298. [doi: 10.1016/S0140-6736(24)01548-4] [Medline: 39551059]
3. Type 2 diabetes. US Centers for Disease Control and Prevention (CDC). 2024. URL: https://www.cdc.gov/diabetes/about/about-type-2-diabetes.html [accessed 2026-04-22]
4. Aroda VR, Ahmann A, Cariou B, et al. Comparative efficacy, safety, and cardiovascular outcomes with once-weekly subcutaneous semaglutide in the treatment of type 2 diabetes: insights from the SUSTAIN 1–7 trials. Diabetes Metab 2019 Oct;45(5):409-418. [doi: 10.1016/j.diabet.2018.12.001] [Medline: 30615985]
5. Aronne LJ, Horn DB, le Roux CW, et al. Tirzepatide as compared with semaglutide for the treatment of obesity. N Engl J Med 2025 Jul 3;393(1):26-36. [doi: 10.1056/NEJMoa2416394] [Medline: 40353578]

XSL•FO
RenderX

6. Alfaris N, Waldrop S, Johnson V, Boaventura B, Kendrick K, Stanford FC. GLP-1 single, dual, and triple receptor agonists for treating type 2 diabetes and obesity: a narrative review. EClinicalMedicine 2024 Sep;75:102782. [doi: 10.1016/j.eclinm.2024.102782] [Medline: 39281096]
7. American Diabetes Association Professional Practice Committee. 9. Pharmacologic approaches to glycemic treatment: standards of care in diabetes-2025. Diabetes Care 2025 Jan 1;48(Supplement_1):S181-S206. [doi: 10.2337/dc25-S009] [Medline: 39651989]
8. Montero A, Sparks G, Presiado M, Hamel L. KFF health tracking poll may 2024: the public's use and views of GLP-1 drugs. KFF. 2024. URL: https://www.kff.org/health-costs/kff-health-tracking-poll-may-2024-the-publics-use-and-views-of-glp-1-drugs/ [accessed 2026-04-22]
9. Pandey A, Ye Y, Wells CR, Singer BH, Galvani AP. Estimating the lives that could be saved by expanded access to weight-loss drugs. Proc Natl Acad Sci U S A 2024 Oct 22;121(43):e2412872121. [doi: 10.1073/pnas.2412872121] [Medline: 39405358]
10. West S, Scragg J, Aveyard P, et al. Weight regain after cessation of medication for weight management: systematic review and meta-analysis. BMJ 2026 Jan 7;392:e085304. [doi: 10.1136/bmj-2025-085304] [Medline: 41500720]
11. Liu L, Chen J, Wang L, Chen C, Chen L. Association between different GLP-1 receptor agonists and gastrointestinal adverse reactions: a real-world disproportionality study based on FDA adverse event reporting system database. Front Endocrinol (Lausanne) 2022;13:1043789. [doi: 10.3389/fendo.2022.1043789] [Medline: 36568085]
12. Ismaiel A, Scarlata GGM, Boitos I, et al. Gastrointestinal adverse events associated with GLP-1 RA in non-diabetic patients with overweight or obesity: a systematic review and network meta-analysis. Int J Obes 2025 Oct;49(10):1946-1957. [doi: 10.1038/s41366-025-01859-6]
13. Heitmann BL. The impact of novel medications for obesity on weight stigma and societal attitudes: a narrative review. Curr Obes Rep 2025 Feb 5;14(1):18. [doi: 10.1007/s13679-025-00611-5] [Medline: 39907856]
14. Fong S, Carollo A, Lazuras L, Corazza O, Esposito G. Ozempic (glucagon-like peptide 1 receptor agonist) in social media posts: unveiling user perspectives through Reddit topic modeling. Emerging Trends in Drugs, Addictions, and Health 2024 Dec;4:100157. [doi: 10.1016/j.etdah.2024.100157]
15. Rad J, Melendez-Torres GJ. Critical discourse analysis of social media advertisements for GLP-1 receptor agonist weight loss drugs: implications for public perceptions and health communication. BMC Public Health 2025 Sep 1;25(1):2996. [doi: 10.1186/s12889-025-24197-8] [Medline: 40890660]
16. Arillotta D, Floresta G, Papanti Pelletier GD, et al. Exploring the potential impact of GLP-1 receptor agonists on substance use, compulsive behavior, and libido: insights from social media using a mixed-methods approach. Brain Sci 2024 Jun 20;14(6):617. [doi: 10.3390/brainsci14060617] [Medline: 38928616]
17. Facebook—statistics & facts. Statista. 2024. URL: https://www.statista.com/topics/751/facebook/ [accessed 2026-04-22]
18. Americans' social media use 2025. : Pew Research Center; 2025 URL: https://www.pewresearch.org/wp-content/uploads/sites/20/2025/11/PI_2025.11.20_Social-Media-Use_REPORT.pdf [accessed 2026-04-22]
19. Poll: 1 in 8 adults say they are currently taking a GLP-1 drug for weight loss, diabetes or another condition, even as half say the drugs are difficult to afford. KFF. 2025. URL: https://www.kff.org/public-opinion/poll-1-in-8-adults-say-they-are-currently-taking-a-glp-1-drug-for-weight-loss-diabetes-or-another-condition-even-as-half-say-the-drugs-are-difficult-to-afford/ [accessed 2026-04-22]
20. Shiffman N, Silverman B. CrowdTangle opens public application for academics. Meta. 2020. URL: https://research.facebook.com/blog/2020/7/crowdtangle-opens-public-application-for-academics/ [accessed 2020-04-22]
21. CrowdTangle. Meta Transparency Center. 2024. URL: https://transparency.meta.com/researchtools/other-data-catalogue/crowdtangle/ [accessed 2026-04-22]
22. Dhirani D, Shahid A, Mumtaz H. A new kind of diabetes medication approved by the FDA: Is there hope for obesity? Int J Surg 2023 Feb 1;109(2):81-82. [doi: 10.1097/JS9.0000000000000044] [Medline: 36799809]
23. Farzam K, Patel P. Tirzepatide. In: StatPearls: StatPearls Publishing; 2024. URL: https://www.ncbi.nlm.nih.gov/books/NBK585056 [accessed 2026-04-22]
24. FDA approves new medication for chronic weight management. US Food and Drug Administration (FDA). 2023. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-new-medication-chronic-weight-management [accessed 2026-04-22]
25. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv. Preprint posted online on Mar 11, 2022. [doi: 10.48550/arXiv.2203.05794]
26. Egger R, Yu J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. Front Sociol 2022;7:886498. [doi: 10.3389/fsoc.2022.886498] [Medline: 35602001]
27. McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection. J Open Source Softw 2018;3(29):861. [doi: 10.21105/joss.00861]
28. Campello R, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. Presented at: 17th Pacific-Asia Conference, Advances in Knowledge Discovery and Data Mining, PAKDD; Apr 14-17, 2013. [doi: 10.1007/978-3-642-37456-2_14]

XSL•FO
RenderX

29. Terry G, Hayfield N, Clarke V, Braun V. Thematic analysis. In: Willig C, Stainton-Rogers W, editors. The SAGE Handbook of Qualitative Research in Psychology: SAGE Publications Ltd; 2017:17-36. [doi: 10.4135/9781526405555.n2]
30. Wittert G, Caterson I, Finer N. The clinical effectiveness of weight loss drugs. Obes Res Clin Pract 2007 Jan;1(1):1-78. [doi: 10.1016/j.orcp.2006.12.001] [Medline: 24351426]
31. Tempia Valenta S, Nicastri A, Perazza F, et al. The impact of GLP-1 receptor agonists (GLP-1 RAs) on mental health: a systematic review. Curr Treat Options Psych 2024;11(4):310-357. [doi: 10.1007/s40501-024-00331-y]
32. Callahan C. 32 celebrities who've opened up about taking ozempic or weight-loss drugs. TODAY. 2025. URL: https://www.today.com/health/celebrities-on-ozempic-rcna129740 [accessed 2026-04-22]

Abbreviations

BERTopic: bidirectional encoder representations from transformers topic
FDA: Food and Drug Administration
GLP-1: glucagon-like peptide-1
GLP-1 RA: glucagon-like peptide-1 receptor agonist
TF-IDF: term frequency–inverse document frequency
UMAP: uniform manifold approximation and projection

Edited by T Mackey; submitted 16.Dec.2025; peer-reviewed by J Rucklidge, J Hill; revised version received 14.Mar.2026; accepted 28.Mar.2026; published 07.May.2026.

Please cite as:
Dennard E, Makres K, Alibilli A, Jain V, Wang A, Abim-Karmon R, Criss S, Hswen Y, Nguyen QC, Ratzki-Leewing A, McCoy RG, Nguyen TT
Exploring Weight Loss Medication Discourse: Mixed Methods Analysis of US-Based Facebook Posts
JMIR Infodemiology 2026;6:e89732
URL: https://infodemiology.jmir.org/2026/1/e89732
doi:10.2196/89732

© Elizabeth Dennard, Katrina Makres, Amrutha Alibilli, Vidur Jain, Anne Wang, Roger Abim-Karmon, Shaniece Criss, Yulin Hswen, Quynh C Nguyen, Alexandria Ratzki-Leewing, Rozalina G McCoy, Thu T Nguyen. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 7.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Online Information-Seeking Behavior Regarding Vitreous Floaters in the United States: Google Trends Analysis

Thanaphat Seeboonruang, BA; Qais A Dihan, MD; Jamie L Surgent-Nahay, MS; Muhammad Z Chauhan, MS, MD; Ahmed F Shakarchi, MPH, MD; Ahmed B Sallam, MD, PhD

Department of Ophthalmology, University of Arkansas for Medical Sciences, 4301 W Markham St, #523, Little Rock, AR, United States

Corresponding Author:
Ahmed B Sallam, MD, PhD
Department of Ophthalmology, University of Arkansas for Medical Sciences, 4301 W Markham St, #523, Little Rock, AR, United States

Abstract

Background: Vitreous floaters are a prevalent phenomenon, with some patients experiencing persistent symptoms that substantially degrade their vision. Public interest and information-seeking behavior regarding vitreous surgery are challenging to quantify using claims alone.

Objective: The aim of this study is to evaluate temporal and geographic patterns of public interest in vitreous floaters and related interventions in the United States and to contextualize this interest relative to retina specialist availability.

Methods: In this retrospective observational study, we queried Google Trends for search terms originating in the United States from 2008 to 2025. We extracted the relative search volume (RSV) using a composite of terms combined with the Google Trends "+" operator. State-level counts of vitreoretinal surgeons from the American Society of Retina Specialists (ASRS) database were used to calculate surgeon density per 100,000 population. We calculated a search-to-surgeon index (SSI) as the ratio of state-level RSV to retina surgeon density to serve as an exploratory metric of public interest relative to specialist availability. We evaluated associations using the Mann-Kendall trend test, Sen slope, and Spearman correlation.

Results: From 2008 to 2025, nonsurgical RSV increased by 138%, demonstrating a significant temporal increasing trend (τ=0.600; P<.001). States with a higher proportion of residents aged 40 years or older demonstrated higher RSV (ρ=0.51; P<.01). The median SSI was 123 (IQR 100-127; range 48-311). Surgical RSV increased by 133% (τ=0.443; P<.001). The median surgical SSI was 133 (IQR 107-164; range 34-341).

Conclusions: This analysis demonstrates substantial geographic variability in public interest in vitreous floaters across the United States. Integrating Google Trends data with surgeon distribution provides an exploratory framework for evaluating relative interest in vitreoretinal care, though it does not directly measure clinical demand.

(JMIR Infodemiology 2026;6:e94821) doi:10.2196/94821

KEYWORDS

public interest; vitreous floaters; Google Trends; vitrectomy; demand index; digital ophthalmology

Introduction

Vitreous floaters are a prevalent phenomenon, with large community surveys suggesting that two-thirds of individuals notice floaters, and one-third perceive them as a visual impairment [1]. Most floaters are benign, but some patients may experience persistent symptoms that substantially degrade their vision and quality of life [2]. The true prevalence of symptomatic floaters requiring clinical intervention remains uncertain, but recent studies suggest a high burden in the population [3,4]. The increase in both aging populations and the incidence of myopia in the United States might cause an increase in public interest in floater-related surgical intervention. Public interest and information-seeking behavior regarding vitreous surgery for symptomatic floaters are challenging to quantify using claims or surgical volume alone, as these metrics reflect use rather than public awareness. Google Trends offers an option for quantifying real-time public interest. Accordingly, we used Google Trends to characterize temporal and regional variation in US public search interest for vitreous floaters and floater-directed interventions [5].

Methods

Ethical Considerations

This study did not use patient data but rather publicly available data on search engine trends in conjunction with publicly available US Census data. As such, no direct patient contact or intervention was involved, and individual informed consent was not required. Therefore, our study was exempted by the

XSL•FO
RenderX

University of Arkansas for Medical Sciences Institutional Review Board. Our study was conducted in accordance with the Declaration of Helsinki.

Google Trends Query Strategy

We queried Google Trends for searches originating in the United States between 2008 and 2025. Google Trends reports relative search volume (RSV) on a normalized scale from 0 to 100, with 100 representing peak popularity for a given search term for the selected time and geographical location. The searches were performed using a composite of search terms, which we grouped into 2 categories. We use the Google Trends "+" operator to combine exact syntax and capture public interest. The nonsurgical category included "floaters," "eye floaters," "cobwebs in vision," "visual floaters," and "floating specks." The surgical category included "vitreous floaters," "vitrectomy," "symptomatic vitreous opacities," "floaterectomy," and "YAG vitreolysis." Although "floaterectomy" is a colloquial and nonstandard medical term, it was intentionally included to capture niche public search variations that patients may use during online information seeking; the term is used to describe combined phacoemulsification and vitrectomy procedures for floaters [6]. The search terms were selected based on consensus among fellowship-trained vitreoretinal surgeons to capture both colloquial, symptom-based layperson queries and formal, treatment-based clinical inquiries. At the time of data collection (January 10, 2026), Google Trends allowed a maximum of 5 syntax inputs per query.

These terms reflect established clinical precedent, supported by previous studies evaluating vitreous floaters and prior evidence-based reviews that specifically use keywords such as "floaters," "vitreous opacities," "floaterectomy," and "vitrectomy" in their search strategies [7,8].

Retina Surgeon Density

We obtained state-level counts of vitreoretinal surgeons from the American Society of Retina Specialists (ASRS) member database [9]. We calculated retina surgeon density for each state by dividing the number of ASRS members by the 2025 US Census Bureau state population estimates and multiplying by 100,000 [10,11].

Search-to-Surgeon Index

We calculated a search-to-surgeon index (SSI) to contextualize public interest relative to specialist availability [12,13]. We defined SSI as the ratio of state-level RSV to retina surgeon density (SSI=state-level RSV/retina surgeon density). Because RSV is a normalized, unitless scale rather than an absolute search count, the SSI is used strictly as an exploratory relative metric to rank geographic disparities rather than to quantify absolute clinical volume or health care demand.

Population Age Distribution

Because Google Trends does not provide demographic data on users' age, we independently obtained state-level population age distributions from the 2025 US Census Bureau. We used the proportion of residents aged 40 years or older as a proxy for populations at higher risk for symptomatic vitreous floaters. We evaluated the correlation between age distribution, RSV, and SSI using the Spearman correlation test.

Statistical Analysis

We evaluated associations using the Mann-Kendall trend test with the Sen slope estimator and the Spearman correlation test performed using Microsoft Excel.

Results

State-Level Retina Specialist Availability and Population Age Structure

We identified 1992 US-based ASRS members. Mean retina surgeon density was 0.58 (SD 0.2; range 0.17-2.09) per 100,000 residents. Delaware (2.09), Maryland (1.04), and Vermont (0.93) had the highest densities, whereas Wyoming (0.17), Alaska (0.27), and New Hampshire (0.28) had the lowest. Census data showed the highest proportions of residents aged 40 years or older in Maine, Montana, New Hampshire, Pennsylvania, Vermont, and West Virginia, while the lowest proportions were in Alaska, Texas, Utah, Georgia, and North Dakota.

Nonsurgical Floaters Search Trends

From 2008 to 2025, RSV for nonsurgical floater terms increased by 138%. Temporal trend analysis demonstrated a strong positive monotonic association between time and RSV values (Kendall τ=0.600; P<.001), with an approximate median increase of 0.172 RSV units per month (95% CI 0.153 - 0.190). The minimum RSV of 42 occurred in November 2008 (Figure 1). There was a moderate positive correlation between the proportion of residents aged 40 years or older and RSV (ρ=0.45; P<.01). In 2025, the median RSV across states was 68 (IQR 61-73; range 48-100), with the highest values in Delaware (100), Alabama (77), and Tennessee (77) and the lowest in Utah (48), Wyoming (53), Montana (55), and Oregon (56) (Figure 2).

XSL•FO
RenderX

Figure 1. Trends in public interest in nonsurgical vitreous floaters over time. Time series of RSV for composite nonsurgical floater–related search terms from January 2008 to December 2025. RSV: relative search volume.

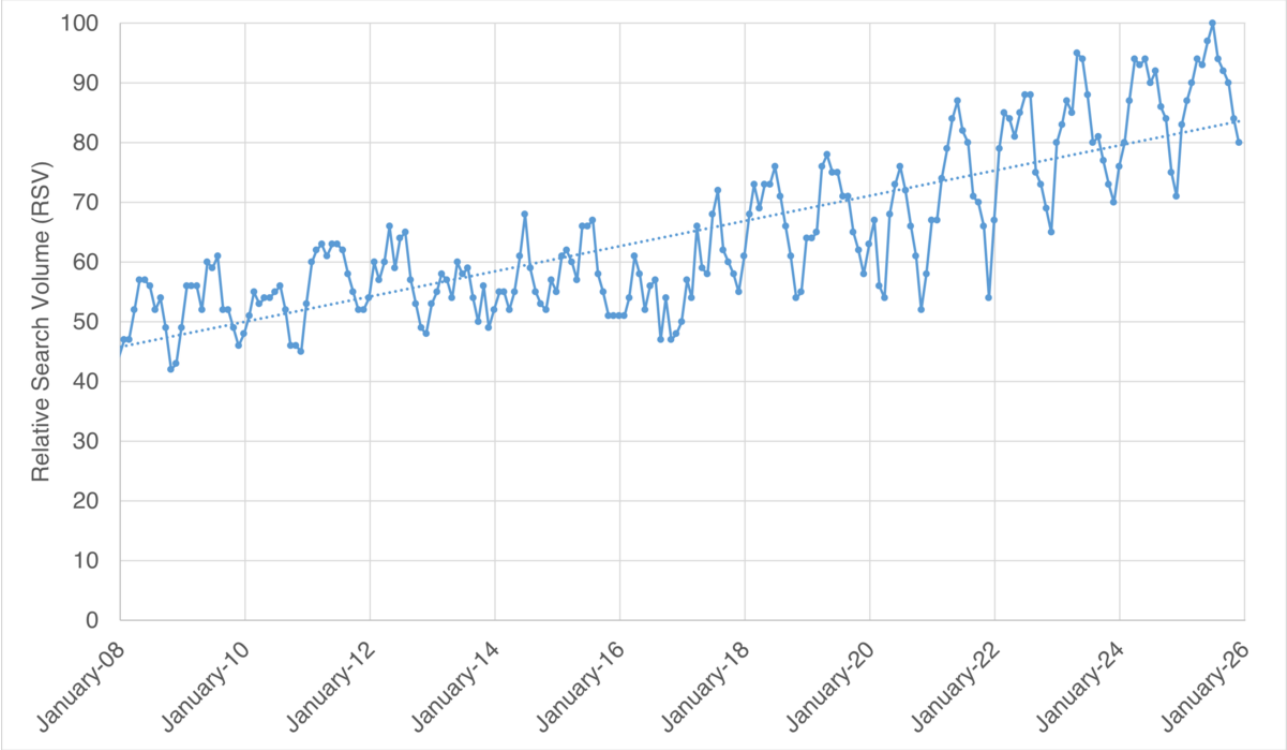
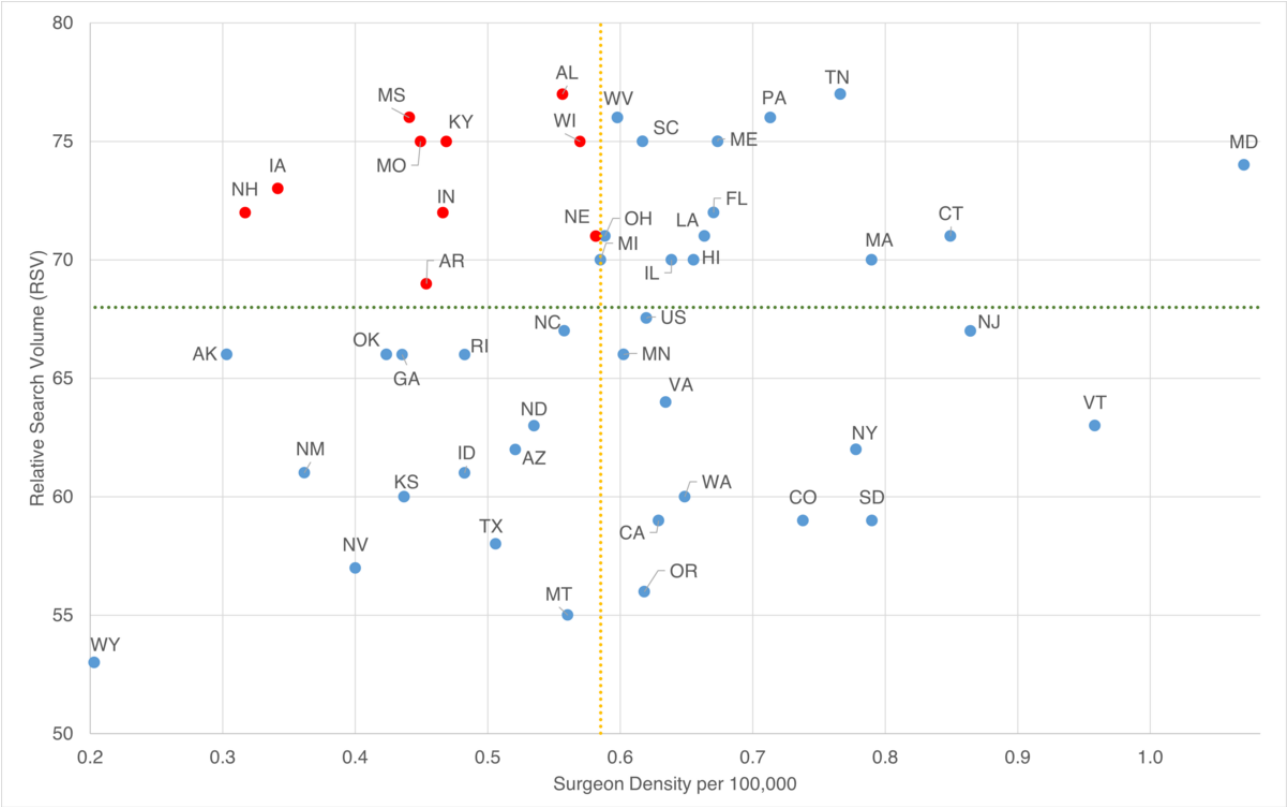


Figure 2. State-level relationship between nonsurgical public interest and surgeon density. Relationship between relative search density (RSV) for composite nonsurgical–related search terms and surgeon density per 100,000 people in 2025. Green dash: median RSV. Orange dash: median surgeon density. Red dot: state with RSV above the median and surgeon density below the median. Delaware (RSV 47.8; surgeon density 2.09) was excluded from the figure to enhance interpretability.

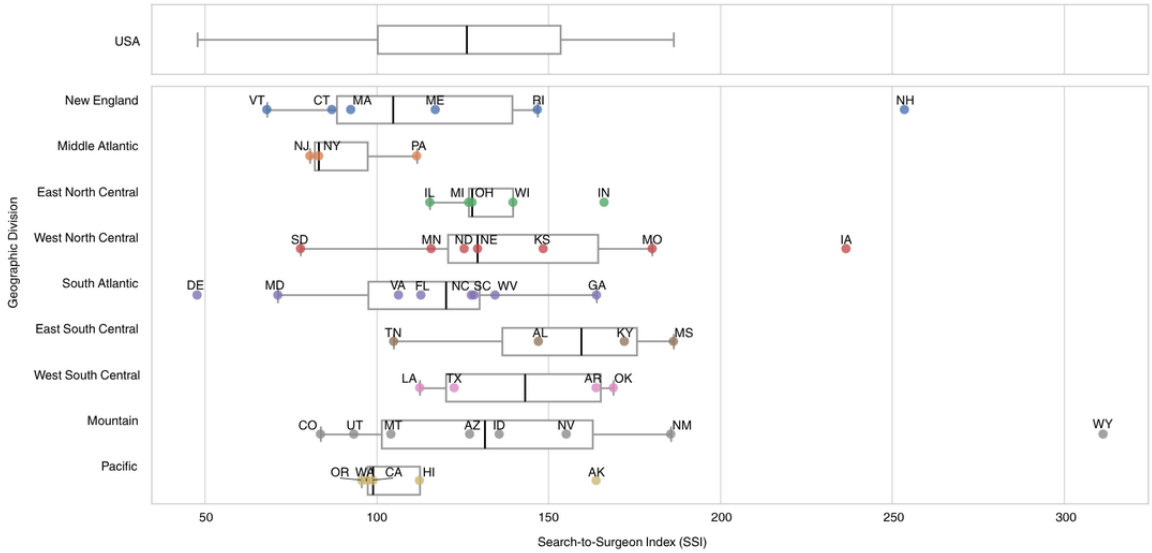


XSL•FO
RenderX

Nonsurgical SSI

The median nonsurgical SSI across states was 123 (IQR 100-127; range 48-311). Wyoming (311), New Hampshire (254), and Alaska (244) had the highest SSI values, whereas Delaware (48), Vermont (68), and Maryland (71) had the lowest (Figure 3). There was no significant correlation between population age distribution and SSI (ρ=−0.08; *P*>.10).

Figure 3. Geographic variation in the search-to-surgeon index (SSI) for nonsurgical floaters across the US. SSI for nonsurgical floaters across the US by Census Bureau geographic division. Overall US SSI is represented by the topmost box plot.



Surgical Floater Search Trends

From 2008 to 2025, RSV increased by 133%. Temporal trend analysis demonstrated a moderate positive monotonic association between time and RSV values (Kendall τ=0.443; *P*<.001) with an approximate median increase of 0.101 RSV units per month (95% CI 0.084 - 0.118) (Figure 4). In 2025, the median RSV across states was 70 (IQR 62-79; range 50-100). West Virginia (100), South Dakota (100), Montana (95), and Nebraska (95) demonstrated the highest RSV values, while Alaska (50), Georgia (58), California (58), Wyoming (58), and Mississippi (58) demonstrated the lowest (Figure 5). The proportion of residents aged 40 years and older showed a low but statistically significant positive correlation with surgical RSV (ρ=0.39; *P*<.01).

XSL•FO
RenderX

Figure 4. Trends in public interest in surgical vitreous floaters over time. Time series of RSV for composite surgical floater–related search terms from January 2008 to December 2025. RSV: relative search volume.

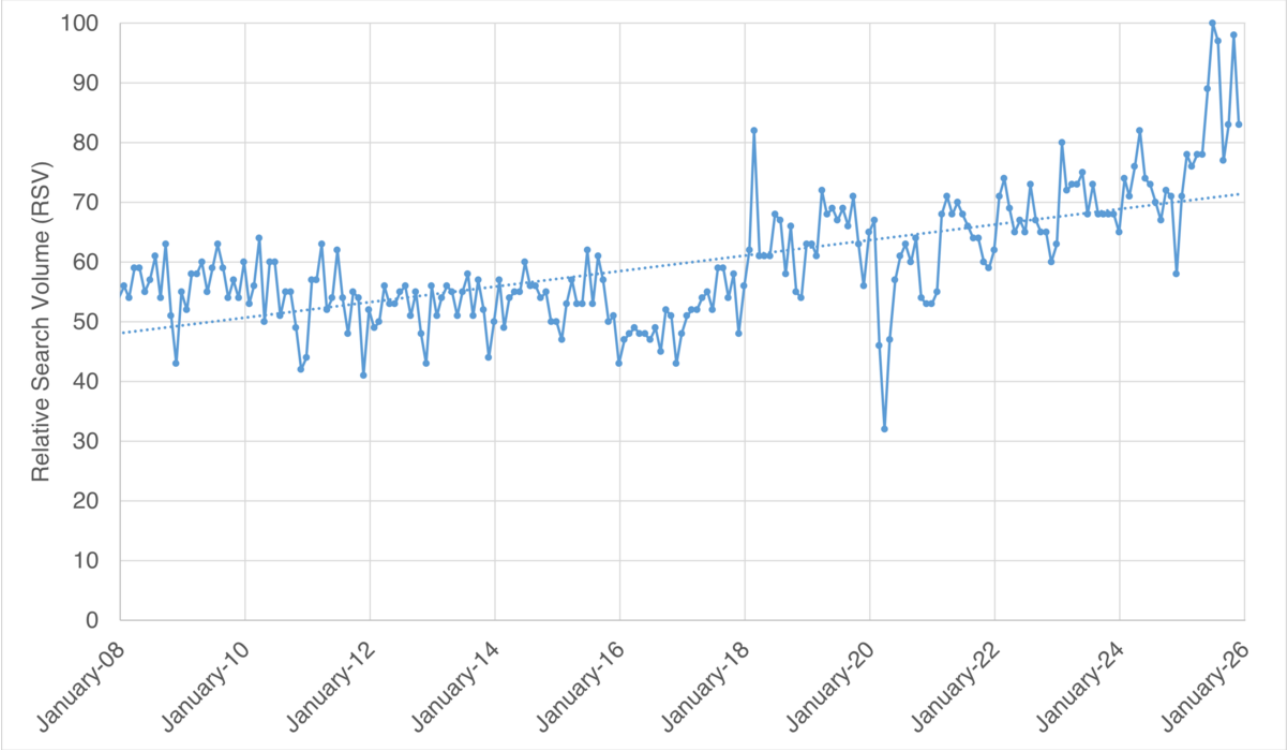
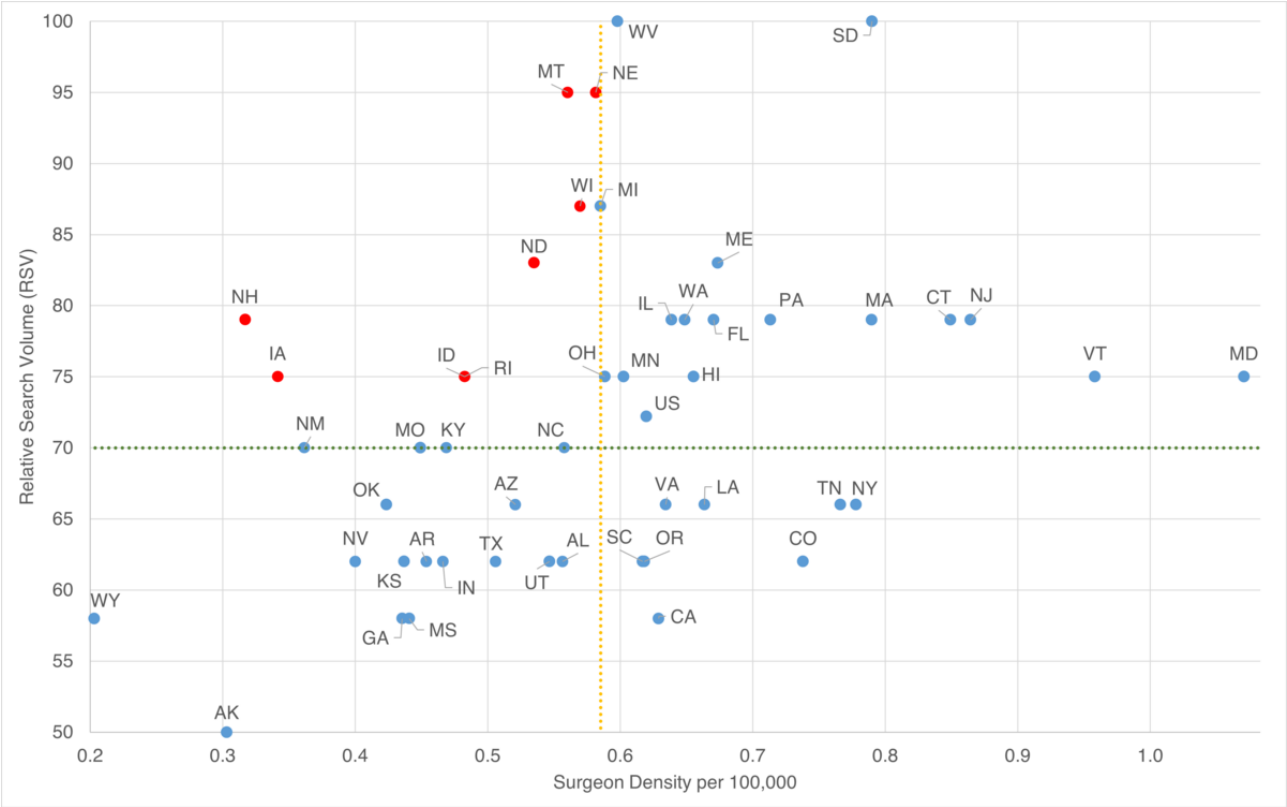


Figure 5. State-level relationship between surgical public interest and surgeon density. Relationship between relative search volume (RSV) for composite surgical–related search terms and surgeon density per 100,000 people in 2025. Green dash: median RSV. Orange dash: median surgeon density. Red dot: state with RSV above the median and surgeon density below the median. Delaware (RSV 33.5; surgeon density 2.09) was excluded from the figure to enhance interpretability.

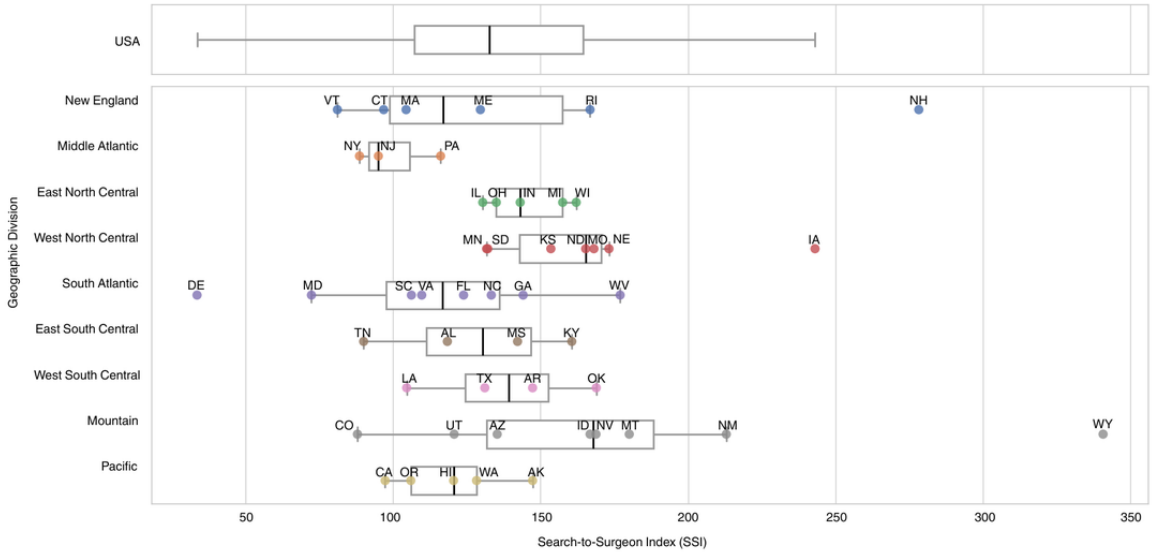


XSL•FO
RenderX

Surgical SSI

The median surgical SSI across states was 133 (IQR 107-164; range 34-341), with the highest values in Wyoming (341), New Hampshire (278), and Iowa (243) and the lowest in Delaware (34), Maryland (72), and Vermont (81) (Figure 6). Population age distribution did not significantly correlate with surgical SSI (ρ=−0.02; P>.10).

Figure 6. Geographic variation in the search-to-surgeon index (SSI) for surgical floaters across the US. SSI for surgical floaters across the US by Census Bureau geographic division. Overall US SSI is represented by the topmost box plot.



Discussion

Principal Findings

Public search interest in vitreous floaters and related interventions increased substantially across the US over the study period. These findings suggest symptomatic floaters are an increasingly important topic for online health information seeking.

Notably, our data demonstrated a significant upward inflection in RSV around March 2018, which coincided with the publication of a randomized clinical trial demonstrating improvement in floater symptoms after yttrium-aluminum-garnet (YAG) laser vitreolysis, and subsequent media coverage [14,15]. However, this correlation does not prove causality; multiple factors, including broader media coverage, social media activity, and changes in search engine algorithms, could have simultaneously influenced general search behavior. Similarly, a temporary decline in search growth in early 2020 coincided with the COVID-19 pandemic. While this mirrors reported reductions in elective ophthalmic procedures, it remains a speculative association without direct data on changes in individual health search patterns [16].

When public interest was interpreted relative to local specialist availability via the SSI, regional variations emerged. For instance, Delaware's high surgeon density and SSI reflect its smaller population and proximity to major metropolitan referral centers, which act as cross-state catchments. Due to its extreme positive skew in surgeon density (2.09 per 100,000), Delaware was excluded from select visual analyses to preserve the interpretability of the scatterplots for the remaining states. Conversely, states with high SSI values and low surgeon densities such as Wyoming and Alaska may be indicative of regions where public interest outpaces local capacity, raising the possibility of regional disparities in subspecialty access [17,18].

Limitations

Our study has several limitations. First, due to the inherent nature of internet search data, state-level RSV is an indirect indicator that cannot be used to infer individual-level care-seeking behavior, true clinical demand, or disease prevalence. Online search behavior is influenced by multiple confounding factors—such as educational campaigns, baseline internet access, and digital health literacy—which were not empirically adjusted for in this study. Furthermore, failure to adjust for state-level confounding variables, such as socioeconomic status, health insurance rates, and local referral patterns, increases the possibility of spurious relationships.

Second, our search strategy was constrained by Google Trends, which restricts exact-match term combinations (using the "+" operator) to a maximum of 5 inputs per query. While our composite categories used the 5 most representative terms determined by expert consensus, this platform limitation inherently omits less common and niche search variations, potentially underestimating the true totality of public interest. In the absence of published data on the exact usage rates of these search terms, our query strategy relied on expert consensus and historical literature precedent, which may not capture all relevant patient searches.

Third, using ASRS membership as a proxy for the total retinal workforce likely underestimates the true density of treating

XSL•FO
RenderX

providers, as many patients with floater symptoms are initially evaluated by general ophthalmologists and optometrists. ASRS density is a crude proxy that does not account for an individual surgeon's clinical capacity, surgical volume, or interstate referral dynamics.

Finally, without external validation against actual patient attendance data, referral rates, or surgical volume, the operational relevance of the SSI to workforce policy decisions remains strictly exploratory.

Conclusions

Public search interest in vitreous floaters and related interventions has increased substantially across the United States from 2008 to 2025. Integrating Google Trends data with surgeon distribution provides an exploratory framework for evaluating relative public interest in vitreoretinal care. Future studies incorporating more granular workforce and referral data are required to validate whether these search patterns translate into measurable gaps in clinical access.

Acknowledgments

The authors attest that no generative AI tools were used in the generation of the scientific content, data analysis, or initial drafting of this manuscript.

Funding

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

Conflicts of Interest

None declared.

References

1. Webb BF, Webb JR, Schroeder MC, North CS. Prevalence of vitreous floaters in a community sample of smartphone users. Int J Ophthalmol 2013;6(3):402-405. [doi: 10.3980/j.issn.2222-3959.2013.03.27] [Medline: 23826541]
2. Albrecht M, Auffarth GU, Friedrich M, Kessler LJ, Khoramnia R. Vision degrading myodesopsia from vitreous floaters in the young: an important aspect of myopia. Surv Ophthalmol 2025;70(2):265-282. [doi: 10.1016/j.survophthal.2024.10.001] [Medline: 39424075]
3. Lippek J, Rynko L, Framme C, Kermani O, Johannsmeier S, Ripken T. Investigating symptomatic vitreous opacities: an online survey and field of view reconstruction. Klin Monbl Augenheilkd 2025 Oct;242(10):991-1000. [doi: 10.1055/a-2676-7455] [Medline: 40763795]
4. Woudstra-de Jong JE, Manning-Charalampidou SS, Vingerling JHR, Gerbrandy SJF, Pesudovs K, Busschbach JJ. The impact of vitreous floaters on quality of life: a qualitative study. J Patient Rep Outcomes 2025 Aug 18;9(1):102. [doi: 10.1186/s41687-025-00934-w] [Medline: 40824500]
5. Cheong JH, Nayer ZH, Trief DF. Demand for cataract surgery in the United States: Google Trends analysis. J Cataract Refract Surg 2025 Dec 1;51(12):1148-1150. [doi: 10.1097/j.jcrs.0000000000001736] [Medline: 40643603]
6. Mossa F, Delaney YM, Rosen PH, Rahman R. Floaterectomy: combined phacoemulsification and deep anterior vitrectomy. J Cataract Refract Surg 2002 Apr;28(4):589-592. [doi: 10.1016/s0886-3350(01)01104-x] [Medline: 11955895]
7. Delaney YM, Oyinloye A, Benjamin L. Nd:YAG vitreolysis and pars plana vitrectomy: surgical treatment for vitreous floaters. Eye (Lond) 2002 Jan;16(1):21-26. [doi: 10.1038/sj.eye.6700026] [Medline: 11913884]
8. Ivanova T, Jalil A, Antoniou Y, Bishop PN, Vallejo-Garcia JL, Patton N. Vitrectomy for primary symptomatic vitreous opacities: an evidence-based review. Eye (Lond) 2016 May;30(5):645-655. [doi: 10.1038/eye.2016.30] [Medline: 26939559]
9. Find a retina specialist. American Society of Retina Specialists. URL: https://www.asrs.org/find-a-specialist [accessed 2026-03-02]
10. Feng PW, Ahluwalia A, Feng H, Adelman RA. National trends in the United States eye care workforce from 1995 to 2017. Am J Ophthalmol 2020 Oct;218:128-135. [doi: 10.1016/j.ajo.2020.05.018] [Medline: 32445703]
11. State population by characteristics: 2020-2025. US Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-detail.html [accessed 2026-09-02]
12. Lieu AC, Chuter BG, Radgoudarzi N, et al. Geographic patterns of ocular oncologist supply and patient demand for uveal melanoma treatment in the United States: a supply and demand analysis. Clin Ophthalmol 2024;18:2487-2502. [doi: 10.2147/OPTH.S472064] [Medline: 39246555]
13. Doepker CP, Pakhchanian H, Raiker R, Lakhani DA, Hogg JP. Google Trends data of radiologists who accept Medicare: a potential tool for predicting state demand. Curr Probl Diagn Radiol 2022;51(1):46-50. [doi: 10.1067/j.cpradiol.2021.03.004] [Medline: 33814217]
14. Shah CP, Heier JS. YAG laser vitreolysis vs sham YAG vitreolysis for symptomatic vitreous floaters: a randomized clinical trial. JAMA Ophthalmol 2017 Sep 1;135(9):918-923. [doi: 10.1001/jamaophthalmol.2017.2388] [Medline: 28727887]
15. Stuart A, Sebag J, Shah CP, Gaurav SK. To treat—or not to treat—vitreous floaters. EyeNet 2018 [FREE Full text]

XSL•FO
RenderX

16. Azad AD, Mishra K, Lee EB, et al. Impact of early COVID-19 pandemic on common ophthalmic procedures volumes: a US claims-based analysis. Ophthalmic Epidemiol 2022 Dec;29(6):604-612. [doi: 10.1080/09286586.2021.2015394] [Medline: 34935591]
17. Pepper CM, Sandefer RH, Gray MJ. Recruiting and retaining physicians in very rural areas. J Rural Health 2010;26(2):196-200. [doi: 10.1111/j.1748-0361.2010.00282.x] [Medline: 20447007]
18. Ahmed A, Ali M, Dun C, Cai CX, Makary MA, Woreta FA. Geographic distribution of US ophthalmic surgical subspecialists. JAMA Ophthalmol 2025 Feb 1;143(2):117-124. [doi: 10.1001/jamaophthalmol.2024.5605] [Medline: 39745719]

Abbreviations

ASRS: American Society of Retina Specialists
RSV: relative search volume
SSI: search-to-surgeon index
YAG: yttrium-aluminum-garnet

Edited by T Mackey; submitted 06.Mar.2026; peer-reviewed by A Jamali, H Tanabe, K Negiloni; revised version received 12.Jul.2026; accepted 04.Aug.2026; published 13.Oct.2026.

Please cite as:
Seeboonruang T, Dihan QA, Surgent-Nahay JL, Chauhan MZ, Shakarchi AF, Sallam AB
Online Information-Seeking Behavior Regarding Vitreous Floaters in the United States: Google Trends Analysis
JMIR Infodemiology 2026;6:e94821
URL: https://infodemiology.jmir.org/2026/1/e94821
doi:10.2196/94821

© Thanaphat Seeboonruang, Qais A Dihan, Jamie L Surgent-Nahay, Muhammad Z Chauhan, Ahmed F Shakarchi, Ahmed B Sallam. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 13.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Data for Linguistic Minority Group Research in Canada: Proof-of-Concept Centralized Health Care Metadata Repository Development and Usability Study

Vincent Martin-Schreiber[1*], BScN, MEng; Cayden Peixoto[2*], HBSc, MSc; Ricardo Batista[3], MSc, MD, PhD; Christopher Belanger[4], HBSc, MBA, PhD; Peter Tanuseputro[5], HBSc, MSc, MD; Amy T Hsu[6], PhD; Lise M Bjerre[2,7,8], MDCM, PhD

[1]Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada

[2]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[3]Akausivik Inuit Family Health Team, Ottawa, ON, Canada

[4]Telfer School of Management, University of Ottawa, Ottawa, ON, Canada

[5]Department of Family Medicine and Primary Care, University of Hong Kong, Hong Kong, China (Hong Kong)

[6]Bruyère Health Research Institute, Ottawa, ON, Canada

[7]Institut du Savoir Montfort, Ottawa, ON, Canada

[8]Department of Family Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Lise M Bjerre, MDCM, PhD
Department of Family Medicine
Faculty of Medicine
University of Ottawa
600 Peter Morand Crescent
Suite 201
Ottawa, ON, K1G 5Z3
Canada
Phone: 1 613 562 5800 ext 2982
Email: lbjerre@uottawa.ca

Abstract

Background: Language barriers between Canadian patients and health care providers are associated with poorer health outcomes, including decreased patient safety and quality of care, misdiagnosis and longer treatment initiation times, and increased mortality. However, research exploring language as a social determinant of health is limited, as Canadian health data are scattered across many jurisdictions, each with its own policies and procedures. This fragmentation makes it difficult for researchers to identify, locate, and use existing data. This paper presents the results of a pilot study that attempts to address this gap by creating a metadata repository (MDR) to act as a central source of information about what data are available at which data holdings across Canada.

Objective: This project aimed to (1) create a proof-of-concept MDR for Canadian health data at the variable level; (2) identify and label language-related variables existing within the MDR data; and (3) develop an interactive, public-facing web application to let users browse and search the MDR.

Methods: Metadata were collected from 5 Canadian health data sources, including 4 provincial data holdings and 1 national survey, and pooled to create a data repository. Then, we performed bottom-up labeling of language-related variables within the pooled metadata by first using a search string algorithm across all variable labels, names, and definitions and then consensus screening these variables using a derived, standardized definition of language or linguistic variables. Using the *Shiny* web framework in R, we then developed an openly accessible web application to allow users to search the proof-of-concept MDR.

Results: A total of 850,343 variables were collected and included in the repository, with most coming from Ontario (n=712,037, 83.7%) and Manitoba (n=97,051, 11.4%) provincial data holdings. Among all variables in the repository, 213,696 (25.1%) were confirmed to be language related.

XSL•FO
RenderX

Conclusions: Developing a national MDR would be a transformative opportunity for Canadian researchers to leverage the full scope of Canadian health administrative data. Although a top-down approach with consistent engagement of and collaboration between provincial data holdings and federal data agencies is ideal to develop a national MDR, this study demonstrates the feasibility of a bottom-up approach in contributing to this overarching goal.

(JMIR Infodemiology 2026;6:e77242) doi:10.2196/77242

KEYWORDS

metadata; metadata repository; variables; language; linguistic

Introduction

Background

Canada's publicly funded health care system generates a vast amount of data covering factors as wide ranging as pharmacy or prescription records, laboratory results, and health care services [1,2]. These data hold immense potential for health care research and for health policy and planning. However, because health care is administered differently across each of Canada's provinces and territories, the data are scattered across a large number of agencies and institutions, each with its own data policies and procedures [3]. This makes it difficult for researchers to access Canada's provincial health data and also creates a more fundamental problem—it is often difficult for researchers to even discover what types of data are available, where they are held, and how to access them. This fragmentation has contributed to significant differences in the availability and accessibility of administrative and other health data across provinces, posing a major challenge for interprovincial or pan-Canadian health care research. This "data fragmentation" can create particular problems for health care research related to patient and health care provider language abilities.

Language as a social determinant of health is an important and emerging topic in health research [4,5], and language barriers between Canadian patients and health care providers are associated with misdiagnosis and longer treatment initiation times [6]; negative experiences for patients [7,8] and physicians [9,10]; and, in hospital settings, decreased patient safety and quality of care [11,12] as well as increased mortality [13]. This issue is of specific concern in Canada, an officially bilingual country in which 76.1% of the population are native English speakers, 22% are native French speakers, and 18% are bilingual [14]. Although French speakers and English speakers can be found across the country, most French speakers live in the provinces of Quebec and New Brunswick. Despite the importance of language-related health research, the data fragmentation described previously makes it difficult and time consuming to even discover what language-related data are available, let alone access and analyze them. This paper presents the results of a pilot study that attempts to bridge this gap by creating a "metadata repository" (MDR) to serve as a central source of information about which data are available at which locations across Canada.

Metadata can be defined as "data about data" [15], and for this project, we sought to create a repository of variable-level metadata. In this context, variable-level metadata include information such as the institution holding the variable, the larger collection or "library" to which it belongs, and a plaintext description. To help illustrate the utility of MDR in light of Canada's bilingual health care context, we put a special focus on identifying language-related variables. In addition to our final metadata dataset, we also created an interactive public-facing web application to let users browse and search the repository.

We discuss the current state of health data and metadata management in Canada and outline the principles and scope guiding our pilot project subsequently.

Current Initiatives in Canada

There are currently 2 main health metadata initiatives in Canada: the Health Data Research Network (HDRN) Canada's Data Access Support Hub (DASH) and the Strategy for Patient-Oriented Research (SPOR) Canadian Data Platform (CDP). The HDRN is a pan-Canadian network of health data–holding organizations, and it established DASH [16] to guide researchers and streamline access to data held by its members. However, DASH only helps researchers access data housed at member organizations of HDRN Canada, and its services are not free to use.

The SPOR CDP, announced in 2019 by Canada's Ministry of Health, is intended to function as a single portal for researchers to request access to administrative, clinical, and social data from sources across the country [17]. To achieve this goal, the SPOR CDP aims to harmonize and validate definitions for key analytic variables (eg, chronic diseases) while expanding the sources, types, and linkages of data available to researchers (eg, social data). Standardizing data definitions allows information exchanged between data holdings to be equally understood by all parties, a concept known as semantic interoperability [18]. Semantic interoperability is especially important as it allows researchers to combine datasets. Canada is known to lag in health data interoperability [19-21], and the development of metadata standards, a set of guidelines that establish a common way of structuring and understanding data [15], would be very helpful. However, the CDP platform was originally announced as a 7-year initiative and is still ongoing as of 2026.

In addition to larger metadata projects, some data-holding organizations also have public-facing websites that allow users to search their metadata. For example, the Institute for Clinical Evaluative Sciences (ICES) in Ontario provides a publicly accessible data dictionary of their metadata that is searchable at the variable level [22]. Although such resources can be helpful, they lead to the problem of data fragmentation described previously, as researchers must visit each organization's website and consolidate results themselves.

XSL•FO
RenderX

Although there are clear use cases for larger projects such as DASH or the CDP and smaller, institution-level metadata websites, they do not offer a free-to-use and up-to-date repository of health metadata from across Canada. The goal of this study is to take the initial steps toward bridging this gap.

A National MDR: Pilot Project Principles

In this study, we were guided by 2 sets of principles: the principles of findability, accessibility, interoperability, and reusability (FAIR) data stewardship [23] and a bottom-up principle of researcher-driven development.

The FAIR principles were developed by Wilkinson et al [23] to address the challenges in managing large amounts of data. The FAIR principles stipulate that both data and metadata should be findable, accessible, interoperable, and reusable by researchers [23,24]. Clearly, the fragmented landscape of Canadian administrative health data does not adhere to FAIR principles in this sense, which creates what we view as unnecessary delays and roadblocks to potentially life-saving research.

We also postulate that there is a useful role for researchers to play in creating a pragmatic and useful national MDR within the current Canadian health data landscape. Given the size and complexity of administrative health databases and the dappled policy environment governing data access across Canada, creating an MDR through the top-down approach at the organizational level would take a large degree of coordination, political will, and resources to harmonize data selection, definition, collection, and sharing procedures across all provincial and territorial health data holdings. Although an MDR built through top-down standardization would be ideal, there is no guarantee that one will be available in Canada soon.

According to the Public Health Agency of Canada, federal, provincial, and territorial governments are currently working to improve the sharing of public health information [25]. However, a data-sharing agreement between these governments is not expected until the end of 2026, with bilateral agreements to follow and then a lengthy process of harmonizing definitions and processes across the data holdings. In the meantime, a simpler solution built by and for researchers has the potential to provide value now.

Methods

Data Sources and Data Collection

To ensure that our proof-of-concept MDR is robust and inclusive, we aimed to include metadata from a variety of national and provincial administrative health data sources. Administrators and data custodians at national and provincial data holdings (Table 1) were contacted via email between January 2023 and September 2023 to request access to the metadata from all held administrative health datasets, ideally in a raw data format such as CSV. Among the data custodians contacted, metadata were provided by or accessible from the ICES [26], the Manitoba Centre for Health Policy (MCHP) [27], the Institut de la statistique du Québec [28], and the New Brunswick Institute for Research, Data and Training [29]. We also obtained and included metadata from the Canadian Longitudinal Study on Aging [30].

Metadata files were obtained from MCHP and the New Brunswick Institute for Research, Data and Training. For the other 4 data holdings, data scraping [31] was performed by a member of the research team (VM-S) to extract metadata from publicly available online sources and data dictionaries. Detailed explanations of how data were collected from each data holding are provided in Multimedia Appendix 1. Once the metadata from the 5 included data holdings were pooled into a single CSV file, the metadata were organized according to commonly reported data elements across sources, including data holding, dataset name, dates available, variable label, variable name, and variable definition, as reported by the respective data source.

Table 1. Data dictionary availability from administrative health data custodians by province.

Province	Data custodian	Publicly accessible data dictionary or catalog
Alberta	Alberta Health	Yes[a]
BC[b]	Population Data BC	Yes
Manitoba	Manitoba Centre for Health Policy	Yes
New Brunswick	New Brunswick Institute for Research, Data and Training	Yes
Newfoundland and Labrador	Newfoundland and Labrador Centre for Health Information	Yes
Nova Scotia	Health Data Nova Scotia	Yes[a]
Ontario	Institute for Clinical Evaluative Sciences	Yes
PEI[c]	Health PEI	No
Quebec	Régie de l’assurance maladie du Québec	No
Saskatchewan	eHealth Saskatchewan	No

[a]Available only upon request.

[b]BC: British Columbia.

[c]PEI: Prince Edward Island.

XSL•FO
RenderX

Data Labeling

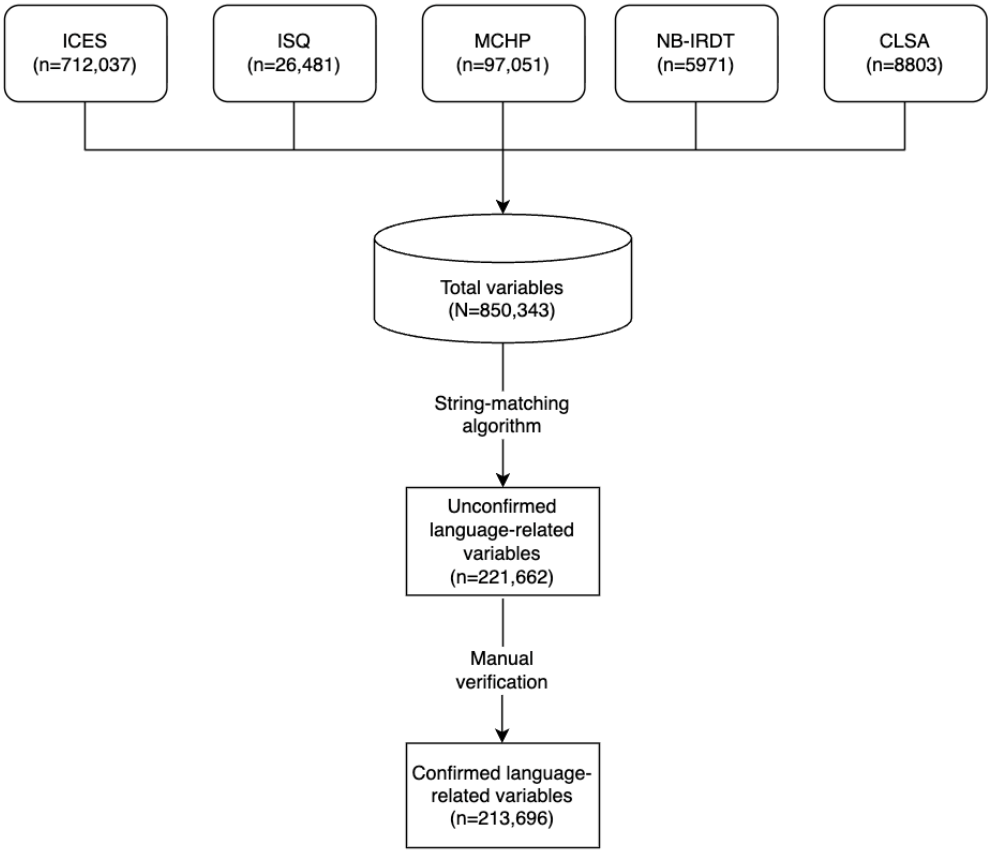
Data labeling (or tagging) is the common process of assigning one or more descriptive tags or labels to a dataset [32], which can make it easier to search and filter results while enabling other uses of the data (eg, machine learning) [33]. To provide an example of searchability in our proof-of-concept MDR, we identified potential language-related variables. We used a naive string-searching algorithm, which works by checking for the occurrence of a pattern (or string) at every possible position in the text [34]. Given Canada's status as a bilingual English-speaking and French-speaking country, we identified potentially linguistic variables as those matching any of the following text strings: "french," "english," "lang," "spoken," "speak," "ling," and "franc."

From here, 2 members of the research team (VM-S and CP) independently reviewed 9.9% (84,068/850,343) of the overall variable names and definitions in the dataset (these were taken from the list of potential language-related variables; Figure 1) to agree on the criteria to define what a language variable is, placing higher value on the most common definitions. The 2 researchers then met to reach a consensus on the standardized definition for tagging language-related variables within the proof-of-concept MDR: "any variable that directly or indirectly provides information regarding the linguistic characteristics of an individual, a health professional, or an organization." This definition aimed to be extremely broad to be able to accommodate any form of research method, including Bayesian statistical approaches.

Both researchers then independently screened all previously identified variables, including variable names and definitions, to identify all language-related variables according to our standardized definition. Screening results were then compared to ensure consensus in the labeled variables between the 2 researchers. Any conflicts in the identification of language-related variables or in the application of the standardized tagging definition were resolved via conversation with a third member of the research team (LMB).

To quantitatively assess the reliability of this screening process, we calculated interrater reliability using Cohen κ [35], which measures the consistency with which both researchers independently applied the standardized definition, accounting for agreement that would be expected by chance. This approach is a standard practice in systematic reviews and content analysis methodologies where operational definitions are developed through iterative refinement and consensus-building discussion [36]. To reduce potential bias, we calculated Cohen κ on 137,594 of the 221,662 (62.0%) variables screened after the standardized definition was established (Figure 1), excluding the 84,068 (37.9%) variables from the initial screening phase used for definition development.

Figure 1. Flowchart for identification of language-related variables from health care data holdings. CLSA: Canadian Longitudinal Study on Aging; ICES: Institute for Clinical Evaluative Sciences; ISQ: Institut de la statistique du Québec; MCHP: Manitoba Centre for Health Policy; NB-IRDT: New Brunswick Institute for Research, Data and Training.



XSL•FO
RenderX

Ethical Considerations

All data used in this study were limited to publicly available metadata, which posed no privacy risk or potential for harm. No personal health information, patient data, or confidential research data were accessed. The collected metadata consisted solely of variable names, descriptions, dataset structures, and data availability information—content that data custodians have chosen to make publicly available to facilitate research discovery and data access applications. For these reasons, approval from a research ethics board was neither required nor sought.

Results

Overview

Across the 5 included data sources, metadata from a total of 850,343 variables were collected and included in our repository. The number of metadata variables collected from each data holding is presented in Table 2. Among the data holdings, the ICES (712,037/850,343, 83.7%) and MCHP (n=97,051, 11.4%) data holdings contained the most variables.

Among the initial 850,343 variables in our repository, 221,662 (26.1%) potential or unconfirmed language-related variables were identified by using a search string algorithm across variable labels, names, and definitions. Consensus screening of these variables using a derived, standardized definition of language or linguistic variables identified 213,696 (25.1%) confirmed language-related variables in our repository (Figure 1).

Interrater reliability for the independent screening process was assessed using observed percent agreement and Cohen κ. The 2 researchers initially agreed on 96.3% (132,538/137,594) of these postdefinition variables, with 5056 (3.7%) disagreements that were subsequently resolved through consensus-building discussion. The calculated Cohen κ was 0.621 (95% CI 0.611-0.632), indicating substantial agreement between the 2 researchers.

Table 2. Number of metadata variables included in the proof-of-concept metadata repository by data holding (N=850,343).

Data holding	Variables, n (%)
Institute for Clinical Evaluative Sciences	712,037 (83.7)
Institut de la statistique du Québec	26,481 (3.1)
Manitoba Centre for Health Policy	97,051 (11.4)
New Brunswick Institute for Research, Data and Training	5971 (0.7)
Canadian Longitudinal Study on Aging	8803 (1.0)

Creating a Usable Proof-of-Concept Web Interface

To facilitate exploration and use of the MDR, we developed a prototype web application that allows users to browse and search the MDR over the internet [37]. Although data management frameworks exist, such as the DataHub Project [38] and the Comprehensive Knowledge Archive Network (CKAN) [39], we developed our application in R (version 4.3.1; R Foundation for Statistical Computing) using *Shiny* [40]. *Shiny* is an open-source R package that makes it easy to build interactive web applications directly using R, a programming language widely used for statistical computing and graphics. *Shiny* was chosen due to its relative simplicity compared to other web development frameworks and the research team's familiarity with the R programming language. The user interface of the *Shiny* app was designed with user-friendliness and functionality in mind, and it allows users to search by keyword, filter by data properties (eg, data holding and linguistic properties), and browse through paginated results. The *Shiny* application was built into a Docker image and hosted on a public platform-as-a-service provider.

Discussion

Principal Findings

Canada's health data are scattered across many organizations and jurisdictions, each with its own policies and procedures, making it difficult for researchers to identify, locate, and use existing data [23,41]. To address this gap, we developed a proof-of-concept MDR containing metadata for more than 850,000 variables from 5 different Canadian data holdings and performed bottom-up labeling of 213,696 (25.1%) of the 850,343 language-related variables within the repository to help researchers easily identify language-related data within the vast landscape of Canadian health data. We also developed an openly accessible web application to allow users to search for the MDR [37].

Building a Bottom-Up MDR: Lessons Learned

Our pilot project demonstrated the feasibility of a bottom-up approach to building an MDR for Canadian health data, but we learned several important lessons that we summarize here. First, complex, manual effort was required to collect (or "scrape") data that are publicly available on the internet. Web scraping is very fast when it works, but each data source needs a bespoke approach. Some websites are straightforward to scrape (eg, those that use backend application programming interfaces) that can be queried directly), but others use an architecture that is not well suited to automatic data collection (eg, those that require repeated form submissions or client-side JavaScript). In addition, the scraping logic is custom-built to each website's design at that moment in time, and if repositories update their websites, the scraping code will need to be updated as well.

Second, there is a need for robust internal data management practices when developing an MDR. We initially prioritized simplicity as well as data portability and transparency; therefore, we stored our data in plaintext CSV files. However, as we collected more data, we were surprised by the size of our final dataset, at a little more than 1 GB. Although this is small

XSL•FO
RenderX

compared to many geospatial or genetic datasets, files of this size are unwieldy to work with, since common office software, such as Microsoft Excel, may not be able to load all variables and can be slow and difficult to transfer to others. For any similar projects, we suggest that a simple data-storage format, such as CSV files, is appropriate for initial feasibility studies, but the project should move quickly to a more sophisticated centralized data-storage solution (eg, a database or a large-file storage solution with version tracking) once feasibility has been established.

Finally, we learned that *Shiny* has several limitations that make it ill-suited for public-facing web applications with datasets this large. *Shiny* creates a new R session for each user and loads the entire dataset into server memory. For a typical dataset measured in KBs or MBs, the overhead is negligible; however, since our data are approximately 1 GB, our application runs out of memory and crashes with more than a few concurrent users. So, although *Shiny* was indispensable to us for rapid prototyping on a local computer, for production deployments, we suggest a different framework in which the data are stored in a single database and queried as needed, as opposed to the server loading a new in-memory copy of the dataset for each user. The user interface could be written using any web development framework (eg, Phoenix and React) and the open-source database software such as PostgreSQL, which is commonly used in large commercial and government projects, would be capable of handling queries on a million-row dataset with millisecond-level response times [42]. Direct access to the application programming interface could also be added, but implementation details of this potential future project are outside the scope of this paper.

Limitations

Although our proof of concept provides a working example of a bottom-up labeled MDR, our methodology is not without limitations. For our initial screening of language variables, we used a search string algorithm to first identify potential language-related variables within all datasets in the proof-of-concept MDR. This search string may not have been exhaustive and could have missed potential language-related variables within the included datasets. Moreover, to best assess the accuracy of our algorithm, it would need to be tested against manually screened datasets as a gold standard for our definition. Although this process was considered too time consuming for the scope of this proof-of-concept project, given the size of the datasets used, it would allow us to evaluate measures such as sensitivity and specificity.

Finally, because variables in our repository were web scraped from various data sources, our repository reflects what variables were available at the point in time of data scraping and would require a repeat of the scraping, screening, and labeling process to update the repository as it is. In addition, there may have been additional metadata in the data holdings that were not made publicly available and therefore were not scrapable, meaning our proof-of-concept MDR may not be exhaustive of variables from the included data holdings. Nonetheless, without top-down policies and procedures in place to allow for easy data collection and labeling processes across Canada, our language-variable data labeling provides a working example of how bottom-up data labeling can be performed by researchers.

Future Directions

Although currently in a beta version, we have plans to expand the MDR to include variables from additional Canadian administrative health data holdings, such as Population Data British Columbia [43], and data from Statistics Canada [44]. Moreover, additional variable tagging can be performed to identify sociodemographic variable types within all included datasets for research purposes, such as sex, gender, race, ethnicity, income, and immigration status. Regarding language-related variables specifically, subtagging can be performed for more specific variable definitions [45], including knowledge of official languages (French or English), variables indicating first language or mother tongue, or variables related to patient–health care provider language concordance.

We also intend to develop a new MDR web application that overcomes the limitations of our *Shiny* app by using a backend database, so that the entire dataset does not need to be loaded into server memory repeatedly for each user.

Finally, we believe that creating a *top-down* national MDR is a worthy goal that should be pursued in tandem with *bottom-up* efforts such as ours. However, such a project would face a number of governance, legal, ethical, and administrative barriers and require a high degree of alignment across diverse organizations so as not to create numerous delays in the collection and integration of data from provincial and organizational data custodians [46,47]. In other words, an ideal top-down MDR will need intense collaboration between many organizations, and although this is beyond our power as individual researchers, we hope Canada's data custodians will rise to the challenge.

Conclusions

This paper addresses the need for a national MDR of administrative and other health data in Canada, underscoring how an MDR can address issues caused by data fragmentation and increase the FAIRness of health care data across the country. However, complex challenges hinder the development of a top-down health data MDR in Canada. We developed a proof-of-concept MDR of administrative health data from 5 different data sources and performed bottom-up labeling of language-related variables within the repository to help researchers easily identify language data in the vast landscape of Canadian health data. This MDR is publicly available online as a searchable data dictionary [37].

Our proof-of-concept MDR illustrates the methodological limitations of a bottom-up approach, which can be complementary and synergistic with but cannot replace top-down approaches to the development of such a repository. Engagement of and collaboration between provincial data holdings and federal data agencies are critical to ensuring a pan-Canadian MDR is comprehensive and can be kept up to date. A national MDR would make it simple and straightforward for Canadian researchers to leverage the full scope of Canadian health data, and open opportunities for new studies as researchers discover datasets previously unknown to them. We

XSL•FO
RenderX

believe that this could be transformative, and we hope this pilot project demonstrates the feasibility of a bottom-up approach in contributing toward this overarching goal.

Acknowledgments

The authors acknowledge the contributions of the following steering committee members and other contributors who, while not fulfilling all authorship criteria, have generously provided their expertise and contributed to the success of this project: Marie-Hélène Chomienne, associate professor and research chair, Chaire de recherche en francophonie internationale et santé de l'immigrant ou du réfugié d'Afrique francophone subsaharienne, University of Ottawa; Jan Warnke, data model analyst, l'Hôpital Jeffery Hale; Claire Kendall, senior scientist, Bruyère Health Research Institute; Cynthia Kendell, assistant professor and research implementation scientist, Department of Medicine, Dalhousie University; and Louise Bouchard, co-director, Réseau Francophonie Observatoire de Recherche Collaborative en Études sur la Santé et les services en contexte minoritaire.

Funding

This work was funded through a Canadian Institutes of Health Research Catalyst Grant: Official Language Minority Communities in Health Research (grant 472426; principal investigator LMB).

Data Availability

The data for this study are available for browsing [37], and the full dataset is available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection methods and compliance by source.
[DOCX File , 22 KB - infodemiology_v6i1e77242_app1.docx]

References

1. Cadarette SM, Wong L. An introduction to health care administrative data. Can J Hosp Pharm 2015 Jun 25;68(3):232-237 [FREE Full text] [doi: 10.4212/cjhp.v68i3.1457] [Medline: 26157185]
2. Lucyk K, Lu M, Sajobi T, Quan H. Administrative health data in Canada: lessons from history. BMC Med Inform Decis Mak 2015 Aug 19;15(1):69 [FREE Full text] [doi: 10.1186/s12911-015-0196-9] [Medline: 26286712]
3. Kendell C, Levy AR, Porter G, Gibson E, Urquhart R. Factors affecting access to administrative health data for research in Canada: a study protocol. Int J Popul Data Sci 2021 Sep 23;6(1):1653 [FREE Full text] [doi: 10.23889/ijpds.v6i1.1653] [Medline: 34632104]
4. Mansoor Y, Wong T, Comeau JL. Language: the ignored determinant of health. Paediatr Child Health 2024 Sep 12;29(3):168-170. [doi: 10.1093/pch/pxad066] [Medline: 38827371]
5. Batista R, Reaume M, Roberts R, Seale E, Rhodes E, Sucha E, et al. Prevalence and patterns of multimorbidity among linguistic groups of patients receiving home care in Ontario: a retrospective cohort study. BMC Geriatr 2023 Nov 09;23(1):725 [FREE Full text] [doi: 10.1186/s12877-023-04267-5] [Medline: 37946126]
6. de Moissac D, Bowen S. Impact of language barriers on quality of care and patient safety for official language minority francophones in Canada. J Patient Exp 2019 Mar 18;6(1):24-32 [FREE Full text] [doi: 10.1177/2374373518769008] [Medline: 31236448]
7. Jutras C, Gauthier AP, Timony PE, Côté D, Kpazaï G. Expérience de francophones en Ontario chez leur médecin de famille: concordance et discordance linguistique. Divers Res Health J 2020 Mar;3:12-33 [FREE Full text] [doi: 10.28984/drhj.v3i0.310]
8. Bowen S. The impact of language barriers on patient safety and quality of care. Société Santé en français. 2015 Aug. URL: https://www.santefrancais.ca/wp-content/uploads/2018/11/SSF-Bowen-S.-Language-Barriers-Study-1.pdf [accessed 2025-12-15]
9. Timony PE, Gauthier AP, Serresse S, Goodale N, Prpic J. Barriers to offering French language physician services in rural and northern Ontario. Rural Remote Health 2016;16(2):3805 [FREE Full text] [Medline: 27316568]
10. Gauthier AP, Timony PE, Serresse S, Goodale N, Prpic J. Strategies for improved French-language health services: perspectives of family physicians in northeastern Ontario. Can Fam Physician 2015 Aug;61(8):e382-e390 [FREE Full text] [Medline: 26505060]

XSL•FO
RenderX

11. Reaume M, Batista R, Talarico R, Guerin E, Rhodes E, Carson S, et al. In-hospital patient harm across linguistic groups: a retrospective cohort study of home care recipients. J Patient Saf 2022 Jan 01;18(1):e196-e204. [doi: 10.1097/PTS.0000000000000726] [Medline: 32433437]
12. Reaume M, Batista R, Talarico R, Rhodes E, Guerin E, Carson S, et al. The impact of hospital language on the rate of in-hospital harm. A retrospective cohort study of home care recipients in Ontario, Canada. BMC Health Serv Res 2020 Apr 21;20(1):340 [FREE Full text] [doi: 10.1186/s12913-020-05213-6] [Medline: 32316965]
13. Seale E, Reaume M, Batista R, Eddeen AB, Roberts R, Rhodes E, et al. Patient-physician language concordance and quality and safety outcomes among frail home care recipients admitted to hospital in Ontario, Canada. CMAJ 2022 Jul 11;194(26):E899-E908 [FREE Full text] [doi: 10.1503/cmaj.212155] [Medline: 35817434]
14. Statistics on official languages in Canada. Government of Canada. URL: https://www.canada.ca/en/canadian-heritage/services/official-languages-bilingualism/publications/statistics.html [accessed 2025-12-15]
15. Ulrich H, Kock-Schoppenhauer AK, Deppenwiese N, Gött R, Kern J, Lablans M, et al. Understanding the nature of metadata: systematic review. J Med Internet Res 2022 Jan 11;24(1):e25440 [FREE Full text] [doi: 10.2196/25440] [Medline: 35014967]
16. About DASH. Health Data Research Network Canada. URL: https://www.hdrn.ca/dash/about-dash/ [accessed 2025-12-15]
17. Health minister announces launch of SPOR Canadian Data Platform. Health Data Research Network Canada. URL: https://www.hdrn.ca/en/news/formal-launch-of-the-spor-canadian-data-platform/ [accessed 2025-12-15]
18. de Mello BH, Rigo SJ, da Costa CA, da Rosa Righi R, Donida B, Bez MR, et al. Semantic interoperability in health records standards: a systematic literature review. Health Technol (Berl) 2022;12(2):255-272 [FREE Full text] [doi: 10.1007/s12553-022-00639-w] [Medline: 35103230]
19. Read KB, Gibson G, Leahey A, Peterson L, Rutle S, Shi J, et al. Identifying metadata commonalities across restricted health data sources: a mixed methods study exploring how to improve the discovery of and access to restricted datasets. J Escience Librariansh 2024;13(2):e907 [FREE Full text] [doi: 10.7191/jeslib.907]
20. Expert Advisory Group. Pan-Canadian health data strategy: toward a world-class health data system. Public Health Agency of Canada. 2022. URL: https://tinyurl.com/ycfb2r8z [accessed 2026-01-23]
21. Affleck E, Sutherland E, Lindeman C, Golonka R, Price T, Murphy T, et al. Human factor health data interoperability. Healthc Pap 2024 Jan 31;21(4):47-55. [doi: 10.12927/hcpap.2024.27272] [Medline: 38482657]
22. Data dictionary. Institute for Clinical Evaluative Sciences (ICES). URL: https://datadictionary.ices.on.ca/Applications/datadictionary/Default.aspx [accessed 2025-12-15]
23. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
24. Boeckhout M, Zielhuis GA, Bredenoord AL. The FAIR guiding principles for data stewardship: fair enough? Eur J Hum Genet 2018 Jul 17;26(7):931-936 [FREE Full text] [doi: 10.1038/s41431-018-0160-0] [Medline: 29777206]
25. Modernizing public health information sharing. Government of Canada. URL: https://www.canada.ca/en/public-health/services/data/modernizing-information-sharing.html [accessed 2025-12-15]
26. Institute for Clinical Evaluative Sciences homepage. Institute for Clinical Evaluative Sciences (ICES). URL: https://www.ices.on.ca/ [accessed 2025-12-15]
27. Manitoba Centre for Health Policy. University of Manitoba. URL: https://umanitoba.ca/manitoba-centre-for-health-policy/ [accessed 2025-12-15]
28. Institut de la statistique du Québec homepage. Institut de la statistique du Québec. URL: https://statistique.quebec.ca/en [accessed 2025-12-15]
29. DataNB. University of New Brunswick. URL: https://www.unb.ca/nbirdt/ [accessed 2025-12-15]
30. CLSA homepage. Canadian Longitudinal Study of Aging. URL: https://www.clsa-elcv.ca/ [accessed 2025-12-15]
31. Web scraping. Statistics Canada. URL: https://www.statcan.gc.ca/en/our-data/where/web-scraping [accessed 2025-12-15]
32. Mahalle PN, Shinde GR, Ingle YS, Wasatkar NN. Data Centric Artificial Intelligence: A Beginner's Guide. Cham, Switzerland: Springer; 2023.
33. Paun S, Artstein R, Poesio M. Statistical Methods for Annotation Analysis. Cham, Switzerland: Springer; 2022.
34. Abdeen RA. An algorithm for string searching. Int J Comput Appl 2019 Oct 17;177(10):17-22 [FREE Full text] [doi: 10.5120/ijca2019919484]
35. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
36. Belur J, Tompson L, Thornton A, Simon M. Interrater reliability in systematic review methodology: exploring variation in coder decision-making. Sociol Methods Res 2018 Sep 24;50(2):837-865. [doi: 10.1177/0049124118799372]
37. CLOSM - data catalogue (beta version). Health Data Dictionary. 2024. URL: https://healthdatadictionary.ca/ [accessed 2025-12-15]
38. Datahub-project. GitHub. URL: https://github.com/datahub-project/datahub [accessed 2025-12-15]
39. CKAN. GitHub. URL: https://github.com/ckan [accessed 2025-12-15]
40. Rstudio/shiny. GitHub. URL: https://github.com/rstudio/shiny [accessed 2025-12-15]

XSL•FO
RenderX

41. Neumann J. FAIR data infrastructure. Adv Biochem Eng Biotechnol 2022;182:195-207. [doi: 10.1007/10_2021_193] [Medline: 35091812]
42. Salunke SV, Ouda A. A performance benchmark for the PostgreSQL and MySQL databases. Future Internet 2024 Oct 19;16(10):382. [doi: 10.3390/fi16100382]
43. Population Data BC homepage. Population Data BC. URL: https://www.popdata.bc.ca/ [accessed 2025-12-15]
44. Data. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/en/type/data [accessed 2025-12-15]
45. Batista R, Hsu AT, Bouchard L, Reaume M, Rhodes E, Sucha E, et al. Ascertaining the Francophone population in Ontario: validating the language variable in health data. BMC Med Res Methodol 2024 Apr 27;24(1):98 [FREE Full text] [doi: 10.1186/s12874-024-02220-7] [Medline: 38678174]
46. Katz A, Enns J, Wong ST, Williamson T, Singer A, McGrail K, et al. Challenges associated with cross-jurisdictional analyses using administrative health data and primary care electronic medical records in Canada. Int J Popul Data Sci 2018 Oct 05;3(3):437 [FREE Full text] [doi: 10.23889/ijpds.v3i3.437] [Medline: 34095523]
47. Become an ICES Scientist. Institute for Clinical Evaluative Sciences (ICES). URL: https://www.ices.on.ca/join-our-research-community/become-an-ices-scientist/ [accessed 2025-12-15]

Abbreviations

CDP: Canadian Data Platform
CKAN: Comprehensive Knowledge Archive Network
DASH: Data Access Support Hub
FAIR: findability, accessibility, interoperability, and reusability
HDRN: Health Data Research Network
ICES: Institute for Clinical Evaluative Sciences
MCHP: Manitoba Centre for Health Policy
MDR: metadata repository
SPOR: Strategy for Patient-Oriented Research

Edited by I Brooks; submitted 09.May.2025; peer-reviewed by H Spechbach, B Cahill, M Al Zoubi; comments to author 29.Oct.2025; revised version received 19.Dec.2025; accepted 14.Jan.2026; published 09.Feb.2026.

Please cite as:
Martin-Schreiber V, Peixoto C, Batista R, Belanger C, Tanuseputro P, Hsu AT, Bjerre LM
Health Data for Linguistic Minority Group Research in Canada: Proof-of-Concept Centralized Health Care Metadata Repository Development and Usability Study
JMIR Infodemiology 2026;6:e77242
URL: https://infodemiology.jmir.org/2026/1/e77242
doi:10.2196/77242
PMID:41543876

©Vincent Martin-Schreiber, Cayden Peixoto, Ricardo Batista, Christopher Belanger, Peter Tanuseputro, Amy T Hsu, Lise M Bjerre. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 09.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Audience-Specific Health Communication: Mixed Methods Evaluation of the Maria Ciência AI-Assisted Knowledge Translation Tool

Mariana Araújo-Pereira[1], PhD; Klauss Villalva-Serra Jr[1], MD; Gustavo Pires-Ramos[2], BSc; Beatriz Sousa-Peres[2], MD; Joanã Nascimento Conceição-Oliveira[1], BSc; Sarah Dourado Maiche[1], MD; Rebeca Rebouças da Cunha Silva[3], BSc; Bruno de Bezerril Andrade[1], MD, PhD

[1]Laboratório de Pesquisa Clínica e Translacional, Instituto Gonçalo Moniz, R Waldemar Falcão 121, Salvador, Brazil

[2]Universidade Federal da Bahia, Salvador, Brazil

[3]Curso de Mestrado de Tecnologias em Saúde, Escola Bahiana de Medicina e Saúde Pública, Salvador, Brazil

Corresponding Author:
Bruno de Bezerril Andrade, MD, PhD
Laboratório de Pesquisa Clínica e Translacional, Instituto Gonçalo Moniz, R Waldemar Falcão 121, Salvador, Brazil

Abstract

Background: Scientific misinformation remains a major barrier to effective health communication. Bridging the gap between academic research and public understanding requires tools that simplify scientific language and adapt content to diverse audiences.

Objective: This study presents Maria Ciência (LPCT-IGM), a specialized GPT-based assistant for science communication. The tool supports researchers in translating peer-reviewed scientific findings through simple prompts into accessible, ethically appropriate materials tailored for children, the general public, health professionals, and policymakers.

Methods: The tool was configured using prompt engineering techniques and guided by curated reference materials on inclusive and nonstigmatizing scientific language. Materials derived from 47 public health papers resulted in 188 outputs, which were assessed by 121 evaluators using 4 criteria: clarity, level of detail, language suitability, and content quality. In addition, outputs generated by Maria Ciência were compared with those produced by a base large language model and with human-written science communication materials. Readability and linguistic accessibility were assessed using multiple established metrics.

Results: Worldwide, mean scores were high: clarity (4.90), language suitability (4.78), content quality (4.72), and level of detail (4.56), on a 5-point scale. Materials for children and the general public consistently achieved the highest ratings across all criteria. A targeted comparison with the base large language model demonstrated superior performance of Maria Ciência in contextual stability. Readability analyses indicated that Maria Ciência's outputs were significantly more accessible than human-written texts, while maintaining high legibility classifications.

Conclusions: Maria Ciência demonstrates the potential of artificial intelligence–assisted tools to enhance knowledge translation and counter scientific misinformation by producing scalable, audience-specific content that balances accessibility and informational integrity.

(JMIR Infodemiology 2026;6:e78843) doi:10.2196/78843

KEYWORDS

science communication; custom GPT; public health; scientific literacy; education

Introduction

Scientific misinformation is one of the most pressing challenges of our time, with direct consequences for public trust, the implementation of health policies, and the protection of population health [1]. The COVID-19 pandemic sharply illustrated how inaccurate or distorted information can compromise collective responses, with measurable impacts on morbidity and mortality [2]. The term "infodemic," adopted by the World Health Organization, refers to the overwhelming volume of both accurate and misleading content that undermines access to trustworthy guidance. This phenomenon is amplified by digital platforms and social media, where misinformation spreads faster than corrective content. Efforts to manage infodemics have become a policy priority for global and national institutions, particularly in response to phenomena such as vaccine hesitancy, which has been linked to the resurgence of diseases such as measles.

Addressing misinformation requires more than reactive fact-checking; it demands the proactive translation of scientific knowledge into accessible, contextually relevant communication. Effective communication must also be timely,

XSL•FO
RenderX

audience-centered, and grounded in the social and cultural contexts of the target populations. However, the scientific community itself often struggles to engage effectively with nonspecialist audiences. The persistence of a publication-centered academic culture, combined with time constraints and a lack of training, limits researchers' ability to participate in outreach activities [3]. While efforts to integrate science communication into academic curricula are increasing, there remains a need for structural support and practical tools to facilitate engagement beyond scholarly environments.

Recent advances in artificial intelligence (AI) have expanded its application beyond clinical decision support into health communication and knowledge translation. Capability- and function-oriented reviews of AI in health care have highlighted a growing shift toward systems designed to support information delivery, user engagement, and audience adaptation, rather than exclusively diagnostic or predictive tasks. These perspectives emphasize that the value of AI in health increasingly depends on how effectively systems align outputs with user needs, contexts, and levels of expertise. In parallel, research on explainable AI has underscored transparency, interpretability, and usability as central requirements for trust, acceptance, and ethical deployment, particularly in health-related applications [4]. Although explainable AI has been extensively discussed in the context of clinical decision support systems, its core principles are equally relevant for health communication tools that aim to produce clear, audience-appropriate, and trustworthy content.

This paper presents Maria Ciência: an AI-assisted platform designed to translate scientific content into tailored, accessible formats for diverse audiences. Developed using a custom GPT model, the tool supports researchers in generating science communication materials adapted for children, adults with low literacy, health professionals, and decision-makers. Importantly, Maria Ciência is not intended to function as a clinical decision support system, nor to generate diagnostic or therapeutic recommendations. Instead, it operates as an audience-facing knowledge translation tool, focused on mediating peer-reviewed scientific evidence into formats appropriate for different levels of literacy, expertise, and social context.

The platform was created with the goal of enhancing the reach and impact of health-related scientific information, particularly in contexts where misinformation can influence individual behaviors and public health outcomes. The approach integrates AI with ethical oversight, thematic supervision, and practical communication strategies, aligning with emerging efforts to design responsible, user-centered AI systems for health communication. In contrast to generic chatbot applications, Maria Ciência is grounded in bioethical principles, equity-driven design, and a commitment to cultural and educational inclusivity. By enabling the same scientific input to be transformed into multiple outputs, the platform offers a scalable and adaptable response to the challenges of misinformation and scientific inaccessibility in public health and beyond.

Methods

Study Design

This study presents the development and evaluation of Maria Ciência, an AI-powered assistant designed to translate peer-reviewed research into audience-specific communication products. The tool was developed with the objective of promoting inclusive, accurate, and culturally sensitive dissemination of scientific knowledge, particularly in the field of public health. The methodological approach combined AI-supported content generation, audience-specific language adaptation, and empirical evaluation based on stakeholder feedback.

Creation and Structure of the Tool

Maria Ciência was developed on the ChatGPT Plus (OpenAI) platform, a commercial version of ChatGPT (OpenAI) that allows for the creation of customized GPTs (also known as AI assistants) through detailed configuration of instructions, role definitions, and operational parameters. The primary function of Maria Ciência is to enable researchers to input scientific materials, such as peer-reviewed papers, and select the intended target audience. Based on this input, the assistant generates adapted communication materials suitable for various reader profiles, including children, the general population, health professionals, and policy decision-makers. These outputs are designed to be practical for use in educational, clinical, and public outreach contexts. ChatGPT is an AI-generated content developed by OpenAI that uses a transformer decoder-only architecture. The GPT model used in this approach was ChatGPT 4.5, released in February 2025. This version stands out as one of the most current versions alongside the ChatGPT o4-mini version, achieving in accuracy tests a 62.5% correctness rate, the highest value in software quality tests among the versions, and the lowest hallucination rate at 37.1%.

Prompt Configuration and Operational Guidelines

The Maria Ciência assistant was configured through prompt engineering techniques combined with a carefully curated set of reference documents that established linguistic, ethical, and stylistic parameters. This configuration aimed to guide the assistant in generating respectful, inclusive, and audience-appropriate outputs. Training the AI assistant with carefully selected documentation greatly improves the accuracy and relevance of its responses [5,6]. In this study, we used reference materials on scientific dissemination, inclusive languages, and health dictionaries to fine-tune the model. Of note, the tool was initially tested in Brazilian Portuguese.

The assistant was assigned the explicit role of a Specialized Science Communicator, with advanced knowledge in public health, immunology, infectious and chronic diseases, and health communication. Its core objective was to translate complex scientific knowledge into formats that could be readily understood by diverse audiences. The interaction protocol included three primary steps: (1) understanding user needs: prompting users to specify the target audience (children, general public, health students, or health managers), (2) continuous engagement: encouraging deeper interaction through follow-up

XSL•FO
RenderX

questions, and (3) final content generation: producing the final adapted text aligned with the audience's profile.

Content adaptation guidelines were explicitly defined for each audience segment: (1) children: playful and narrative-driven writing inspired by Writing for Young Minds, incorporating contextual summaries and storytelling techniques; (2) general public: simplification of scientific concepts with actionable health information, drawing from accessible Brazilian sources such as Superinteressante and Profissão Biotec; (3) health students: simplified explanations while maintaining technical terminology, with emphasis on key health concepts; and (4) health professionals and managers: structured summaries including population characteristics, methodological overviews, and actionable public health recommendations (typically 5 suggested policy improvements).

Additionally, for social media content, the assistant adapted outputs for platforms such as Instagram (Meta) and LinkedIn, ensuring appropriate tone, hashtags, and visual alignment for each context.

Training Materials

Among the sources used were the "Na ponta das línguas: pequeno glossário para apoiar o enfrentamento do estigma e da discriminação," carried out with the support of the Brazilian Ministry of Health, which offers terminology to reduce stigma in health communication [7], as well as the "UNAIDS Terminology Guidelines" [8], which provide recommendations for respectful and accurate language related to HIV and global health. The "Global TB Dictionary" was used to ensure technical accuracy in tuberculosis-related content, and also the "UNICEF Terminology Dictionary" for the protection of children regarding matters of a sexual nature [9]. To support outputs for younger audiences, the tool's language was informed by resources such as "Writing for Young Minds" and adapted scientific texts such as those published in Frontiers for Young Minds [10]. Additional materials, including guides to science communication [11,12] and examples of accessible writing from Brazilian science outreach initiatives such as "Profissão Biotec" [13], were used to calibrate tone, clarity, and structure across all outputs.

The training process emphasized the use of nonstigmatizing and inclusive language, minimization of unnecessary technical jargon, clear structuring of information according to literacy level, and contextual sensitivity to cultural and ethical dimensions of health communication. An additional focus was placed on ensuring equitable representation and fairness in responses across diverse population groups, particularly those historically marginalized in public health communication. The outputs generated by Maria Ciência included short narratives and analogies for children, simplified papers for the general population, technical summaries for health professionals, strategic briefs for health managers, and communication materials designed for social media. Additionally, Maria Ciência includes, in its presentation on the question bar, preconfigured prompts for users with guiding questions for using the chatbot, such as "translate this article for the general public," "how to explain this concept to children," or "create an accessible summary for children." Finally, the assistant was programmed to provide appropriate attribution for all generated materials, including the original scientific source, first author, journal, and year of publication.

Generation of Outputs for Evaluation

Following the configuration of the assistant, we selected 47 peer-reviewed papers (on public health, infectious diseases, and epidemiology) from our institution to serve as the training material. This approach provided the opportunity to invite the original authors of these papers to participate as evaluators in the assessment team. To minimize potential evaluation bias, the chatbot outputs for each paper were generated by external collaborators who were not part of the research group and who did not have a scientific background. Using only the preconfigured guiding questions in the chatbot interface, these collaborators generated 4 outputs per paper, including "for children," "for health managers," "for social media," and "for the general public," resulting in a total of 188 outputs. All outputs were generated using a dedicated user account created exclusively for this purpose, to minimize potential bias from prior sessions or unrelated model interactions.

Public Evaluation of Chatbot Outputs

Following the generation of 188 outputs (4 per paper across 47 selected papers), a public evaluation process was conducted to assess the quality and appropriateness of the content produced by Maria Ciência. A standardized online evaluation form was developed for this purpose, structured to allow systematic feedback from diverse audiences. The evaluation process engaged 5 stakeholder groups: authors of the original papers, health professionals and students, social media specialists, members of the general public, and health managers. The evaluation form was disseminated through social media channels to reach a broad and heterogeneous audience. In addition, the original authors of the selected papers were invited to participate in the evaluation. This dual approach enabled the inclusion of both expert and lay perspectives in the assessment process. Participation was voluntary and anonymous.

Of note, evaluators were allowed to assess materials not exclusively targeted to their own audience profile to capture broader perceptions of clarity, tone, and appropriateness, reflecting real-world dissemination contexts in which health communication materials are often encountered by diverse audiences. In addition, children were not recruited as evaluators in this study, given that research involving minors requires parental permission and age-appropriate assent procedures. Given the exploratory nature of this evaluation and the absence of a child-centered protocol, the assessment of child-focused materials relied on adult evaluators.

Each participant was asked to assess selected chatbot outputs according to the following criteria: (1) Clarity: Is the text clear and appropriate for the intended audience? (2) Detail: Does the text provide sufficient and relevant information? (3) Language suitability: Is the language appropriate for the literacy level and context of the intended audience? (4) Content quality: Does the text maintain scientific accuracy and communicative effectiveness?

Participants assigned scores on a scale from 1 (poor) to 5 (excellent) for each criterion. Additionally, the form included

XSL•FO
RenderX

an open field for qualitative comments, enabling participants to provide contextual feedback on strengths, limitations, or suggestions for improvement. Qualitative feedback provided by anonymous evaluators was analyzed using thematic categorization. Comments were grouped into four predefined domains: (1) language (clarity, accessibility, and appropriateness of language use); (2) information (accuracy, level of detail, and appropriateness of content for the audience); (3) structure (organization, narrative flow, and format of the material); and (4) proposal (whether the material complied with the intended purpose and target audience). For each domain, comments were further classified as either criticism or praise (for language, information, and structure), or as complies with proposal or does not comply with proposal (for proposal). Each comment was reviewed independently and could be assigned to multiple categories when it addressed more than one thematic domain.

Accuracy Evaluation and Comparison With Base GPT

In addition to the public evaluation of the outputs, a focused accuracy evaluation was conducted to compare the performance of Maria Ciência with that of the base GPT-4.5 (OpenAI) model. This comparison aimed to assess whether the custom configuration of Maria Ciência, which incorporates training on inclusive, nonstigmatizing, and health-appropriate language, enhanced the model's ability to maintain contextual relevance and communicative precision in public health content generation. For this purpose, a selected set of questions previously answered by Maria Ciência was resubmitted to both Maria Ciência and the base GPT-4.5 (without the custom configuration) separately.

To minimize potential bias from session memory, priming, or model adaptation effects, the evaluations for Maria Ciência and base GPT-4.5 were conducted using separate user accounts. This ensured that the comparative responses were generated independently, reducing the risk of inadvertent learning from previous interactions within the same account. In addition, to assess response stability, the same questions were submitted multiple times in different conversational sequences, allowing us to evaluate whether the models maintained contextual coherence across repeated interactions.

Responses from both models were evaluated by a team composed of health care professionals and undergraduate students. The evaluation criteria included 4 key dimensions: whether the response accurately established the context of the question; whether contextual coherence was preserved throughout the conversation; whether there was any interruption or drift from the intended context; and, if such drift occurred, whether the model was able to recover and return to the appropriate context. Each dimension was scored on a qualitative scale from 1 (poor) to 5 (excellent). The comparative analysis of results aimed to determine whether Maria Ciência's configuration effectively enhanced contextual accuracy and stability, thus supporting its suitability for reliable and ethically appropriate use in public health communication. The following four dimensions were used to structure the assessment: (1) Establishment of a context: Does the answer fulfill the objective of the question according to the inserted context? (2) Continuity of conversation without specific context of the question: Do the following answers in the conversation with the chatbot lose the context in relation to the question? (3) Interruption of context: Do the answers interrupt or stop fulfilling the context of the question? (4) Return to context: Even after the interruption of the context or digression, is the chatbot able to return to the context of the question?

The tool operates in various languages, having been tested by the developers at Maria Ciência in Portuguese, English, Spanish, Italian, and French. Moreover, as it is a GPT assistant, the platform supports more than 50 languages [14]. Regardless of the language, the generated content follows the same principles of technical configuration, linguistic curation, and thematic supervision. This version preserves the commitment to accessibility, scientific accuracy, and ethical adequacy in knowledge translation, with a view toward application in international and multilingual contexts.

Comparative and Readability Analyses

We conducted additional analysis, including human-written science communication materials and outputs generated by a nonspecialized base GPT-4.5. First, we performed an active search for human-authored news papers and popular science texts related to the peer-reviewed papers previously included in the main evaluation. This search yielded human-written communication materials corresponding to 5 of the originally analyzed scientific papers, all published in open-access media outlets.

In addition, to include recent and independent examples, we searched Google News (on December 1, 2025) using the query "new study in infectious diseases." From the results, we selected the first five news reports that met the following criteria: (1) publication in non–university-affiliated media outlets, (2) reference to peer-reviewed scientific papers, and (3) availability of the original scientific papers under open-access conditions.

For these five additional papers, audience-oriented texts were generated using both Maria Ciência and the base GPT-4.5 model. The same simple prompt was applied to the base model across all cases ("transform this article into a text for the general population"), allowing a direct comparison between specialized and nonspecialized large language model outputs.

All materials (Maria Ciência outputs, base GPT-4.5 outputs, and human-written texts) were subjected to a standardized textual readability analysis. This analysis was applied to texts generated for all target audiences and encompassed both the original evaluation set and the additional comparative corpus. Readability metrics included: Flesch reading ease, Gulpease index, Flesch-Kincaid grade level, adapted Gunning fog index, automated readability index (ARI), Coleman-Liau index, letter-to-word ratio, syllable-to-word ratio, words-per-sentence ratio, and proportion of complex words. All readability analyses were conducted using the ALT (analysis of language and text) software, a tool specifically developed for Portuguese-language texts, as described by Moreno et al [15] (2023). This approach enabled a quantitative and language-appropriate comparison of textual complexity across human-written materials and AI-generated outputs, independently of subjective evaluator impressions.

XSL•FO
RenderX

Data Analysis

Quantitative data from the evaluation forms were analyzed using descriptive statistics. For each evaluation criterion, means and SDs were calculated. Frequencies and percentages were computed for the classification of evaluator identities and for the distribution of outputs across target audiences. All analyses were conducted using the structured database generated from the stakeholder evaluations. For comparative purposes, we compared each readability metric according to the text generator (Maria Ciência, base GPT-4.5, or human-written).

Ethical Considerations

This study was conducted as a public opinion survey evaluating AI-generated science communication materials. All data were collected through voluntary and anonymous participation, without the collection of identifiable or sensitive personal information, and without any form of intervention or risk to participants. In line with international ethical standards and in accordance with Brazilian Resolution CNS 510/2016, which exempts public opinion research involving nonidentified participants from requiring formal ethics committee approval, this study did not require submission to a research ethics committee.

Results

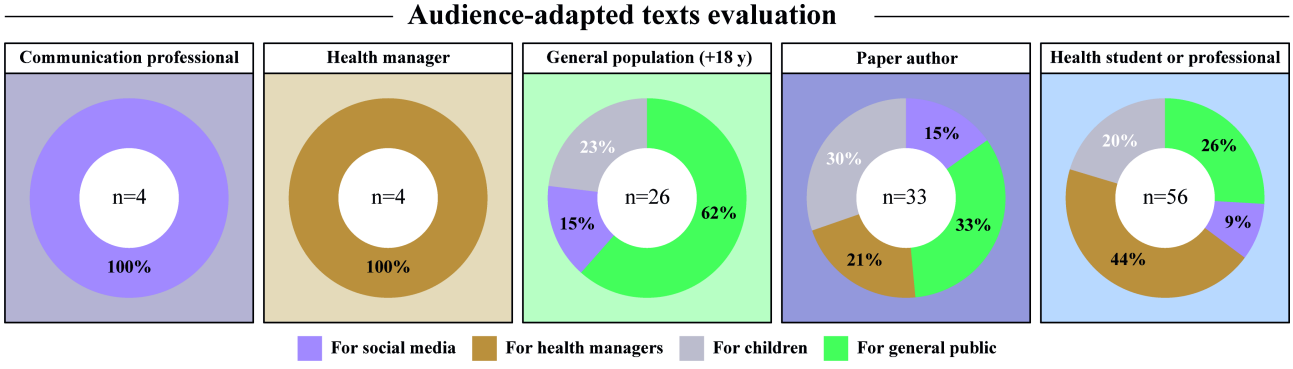
Overview

Before launching the public evaluation, we first analyzed the process of generating the 188 outputs used for assessment. The generation was conducted by external collaborators with no scientific background, using a dedicated user account created exclusively for this purpose. During this process, important differences were observed between the use of the free version and the paid version (ChatGPT Plus). In the free version, the model frequently (10%) exhibited technical limitations: it would often require questions to be reformulated and occasionally produce incomplete outputs. The average generation time per output in this version was approximately 15.2 (SD 2.3) seconds. After upgrading to the ChatGPT Plus version, performance improved in measurable operational terms. The model produced responses more rapidly and with greater consistency, showing a more direct communication style. In this version, the mean generation time decreased to 8.5 (SD 1.5) seconds, and the incidence of incomplete outputs was eliminated. Despite these differences, the free version remained capable of generating the requested outputs using the predefined prompts provided by Maria Ciência; however, it required longer processing time and occasional manual resubmission of prompts to ensure complete responses.

Evaluator Profile and Distribution of Reviewed Outputs

The evaluation of Maria Ciência involved 121 responses to the form, stratified across 5 distinct groups, each representing a key target audience of the tool. The distribution of respondents was as follows: health professionals and students (56/121, 44.6%), original authors of the scientific papers (33/121, 27.3%), members of the general population (26/121, 21.5%), communication specialists (4/121, 3.3%), and health managers (4/121, 3.3%; Figure 1). Each evaluator assessed outputs generated for one or more specific target audiences, including “for children” (n=27), “for health managers” (n=35), “for social media” (n=24), and “for the general public” (n=35). Among these groups, only the original paper authors had prior in-depth knowledge of the scientific content being communicated. Notably, Figure 1 illustrates the distribution of reviewed text types according to identity profile classification, revealing clear patterns in how different stakeholder groups engaged with the audience - adapted outputs. Communication professionals and health managers exclusively reviewed solely the texts targeted at their respective profiles (100%), while the other evaluator profiles diversified their reviewed texts. Given the small proportion of the evaluator sample, results related to these subgroups were reported descriptively and were interpreted as exploratory.

Figure 1. Donut charts summarizing the types of audience-adapted texts proportionally reviewed by each participant identity classification. Segment colors represent the percentage of texts tailored for different audiences, including (1) for social media (lavender), (2) for health managers (golden), (3) for children (gray), and (4) for the general public (green). Each donut reflects the distribution of text types reviewed by participants within a specific identity group.



Among the general population evaluators, 62% (16/26) assessed the “for the general public” texts, while 23% (6/26) provided feedback on “for children” versions, and 15% (4/26) on “for health managers.” Notably, none of the general population participants reviewed the “for social media” outputs. Health professionals and students provided a broader distribution of feedback, with 44% (24/56) evaluating “for health managers” outputs, 25% (14/56) “for the general public,” 20% (11/56) “for children,” and 9% (5/56) “for social media.” Finally, the paper authors demonstrated a balanced engagement across all 4

XSL•FO
RenderX

categories: 33% (11/33) reviewed “for the general public” outputs, 30% (10/33) “for children,” 21% (7/33) “for health managers,” and 15% (5/33) “for social media” (Figure 1). The distribution of respondents and their evaluation profiles across different target audiences is summarized in Table 1.

Table . Distribution of reviewer identities and rating frequencies by target audience. This table displays the frequency (n) and percentage (%) of reviewers’ self‐reported identity classifications and rating scores across 4 different domains, stratified by the number of reviews attributed to each target audience texts. *P* values were calculated using Fisher exact test (for nominal categorical counts) or Kruskal-Wallis tests (for ordinal rating distributions), comparing across the 4 target‐audience groups. A *P* value <.05 denotes a statistically significant difference among audiences.

	For children (n=27)	For health managers (n=35)	For social media (n=24)	For the general public (n=35)	*P* value
Variable					
Participant classification, n (%)					<.001
Paper author	10 (37)	7 (20)	11 (45.8)	5 (14.3)	
Communication professional	0 (0)	0 (0)	4 (16.7)	0 (0)	
General population (18+ years)	6 (22.2)	0 (0)	4 (16.7)	16 (45.7)	
Health manager	0 (0)	4 (11.4)	0 (0)	0 (0)	
Health student or professional	11 (40.7)	24 (68.6)	5 (20.8)	14 (40)	
Rating					
Clarity, n (%)					.21
1	0 (0)	0 (0)	0 (0)	0 (0)	
2	0 (0)	1 (2.9)	0 (0)	0 (0)	
3	0 (0)	2 (5.7)	0 (0)	0 (0)	
4	4 (14.8)	6 (17.1)	4 (16.7)	1 (2.9)	
5	23 (85.2)	26 (74.3)	20 (83.3)	34 (97.1)	
Detailing, n (%)					.28
1	0 (0)	0 (0)	0 (0)	0 (0)	
2	0 (0)	1 (2.9)	0 (0)	1 (2.9)	
3	2 (7.4)	8 (22.9)	4 (16.7)	1 (2.9)	
4	6 (22.2)	7 (20)	6 (25)	5 (14.3)	
5	19 (70.4)	19 (54.3)	14 (58.3)	28 (80)	
Language adequacy, n (%)					.20
1	0 (0)	0 (0)	0 (0)	0 (0)	
2	0 (0)	1 (2.9)	0 (0)	0 (0)	
3	3 (11.1)	1 (2.9)	2 (8.3)	1 (2.9)	
4	4 (14.8)	12 (34.3)	2 (8.3)	6 (17.1)	
5	20 (74.1)	21 (60)	20 (83.3)	28 (80)	
Content quality, n (%)					.11
1	0 (0)	0 (0)	0 (0)	0 (0)	
2	0 (0)	0 (0)	0 (0)	0 (0)	
3	0 (0)	4 (11.4)	1 (4.2)	0 (0)	
4	7 (25.9)	12 (34.3)	7 (29.2)	7 (20)	
5	20 (74.1)	19 (54.3)	16 (66.7)	28 (80)	

XSL•FO
RenderX

Public Evaluation of Audience-Adapted Outputs

Each output was independently evaluated by respondents from multiple stakeholder groups, using 4 evaluation criteria: clarity of the text, level of detail, suitability of language for the intended audience, and overall content quality. Participants rated each criterion on a 5-point scale, with 5 indicating strong agreement regarding the quality or appropriateness of the item assessed. The distribution of evaluator profiles and scoring frequencies is presented in Table 1.

Within this exploratory evaluation, the adapted texts were rated highly across all criteria (Table 1). Texts targeting the general public received strong evaluations. Among members of this audience, mean scores were 4.94 (SD 0.25) for clarity, 4.56 (SD 0.89) for detail, 4.62 (SD 0.62) for language suitability, and 4.75 (SD 0.45) for overall quality. Students and health professionals were equally enthusiastic, with scores of 5.00 (SD 0.00) for clarity, 4.86 (SD 0.36) for detailing, 4.93 (SD 0.27) for language, and 4.93 (SD 0.27) for quality. Of note, researchers who authored the original papers were somewhat more critical, assigning 5.00 (SD 0.00) for clarity, 4.80 (SD 0.45) for detail, 4.80 (SD 0.45) for language, and 4.60 (SD 0.55) for overall quality (Table 2).

Similarly, the child-focused outputs were well received. The general public assigned near-perfect ratings of 5.00 (SD 0.00) for clarity, 4.83 (SD 0.41) for detail, 4.83 (SD 0.41) for language suitability, and 4.83 (SD 0.41) for content quality. Students and health professionals also rated the child-focused content highly, with a means of 4.82 (SD 0.40) for clarity, 4.64 (SD 0.67) for detailing, 4.73 (SD 0.65) on language suitability, and 4.73 (SD 0.47) regarding overall quality. Again, the original authors were more reserved, assigning 4.80 (SD 0.42) for clarity, 4.50 (SD 0.71) for detail, 4.40 (SD 0.84) for language, and 4.70 (SD 0.48) for overall quality (Table 2).

Next, the version tailored for health managers was rated lower by managers themselves, who rated it 4.00 (SD 1.41) for clarity, 4.00 (SD 1.41) for detail, 4.00 (SD 0.82) for language suitability, and 4.25 (SD 0.96) for content quality. Students and health professionals provided comparably high scores, with 4.83 (SD 0.48) for clarity, 4.58 (SD 0.72) for detail, 4.62 (SD 0.71) for language suitability, and 4.67 (SD 0.56) for overall quality. Original paper authors again expressed greater variability in their evaluations, especially regarding level of detail, with a mean of 3.29 (SD 0.49), similarly lower ratings were reported for clarity, which was 4.29 (SD 0.76), 4.43 (SD 0.53) for language, and 3.71 (SD 0.49) for content quality (Table 2).

For evaluations aimed at social media texts, all evaluators rated these texts highly. General public reviewers rated this version at 5.00 (SD 0.00) for clarity, 4.75 (SD 0.50) for detail, 4.50 (SD 1.00) for language suitability, and 4.50 (SD 1.00) for overall quality. Students and health professionals similarly gave high scores, 5.00 (SD 0.00) for clarity, 4.40 (SD 0.89) for detail, 4.80 (SD 0.45) for language suitability, and 4.60 (SD 0.55) for content quality. Communication professionals assigned a perfect mean score of 5.00 (SD 0.00) across all 4 criteria. As observed for other target audiences, paper authors provided more conservative ratings, with scores of 4.64 (SD 0.50) for clarity, 4.09 (SD 0.83) for detail, 4.73 (SD 0.65) for language suitability, and 4.55 (SD 0.52) for overall quality (Table 2).

Taken together, within this exploratory evaluation, audience-adapted texts received high mean scores across all 4 assessed criteria, regardless of the target audience. Notably, texts adapted for children and for the general population received particularly high ratings across all dimensions, with most mean scores approaching the maximum value of 5. Health students or professionals, communication professionals, and the general adult population tended to assign higher scores overall, while health managers and paper authors demonstrated greater variability in their assessments, especially for detailing of texts targeting health managers or social media platforms (Figure 2). Among all groups, paper authors exhibited the greatest variability in their evaluations. These patterns indicate broad perceived acceptability of the materials in this sample, with subtle differences in perceived quality depending on the evaluator profile (Figure 3).

XSL•FO
RenderX

Figure 2. Mean evaluation scores of audience-adapted texts by criterion and evaluator population. Bubble plot presenting the mean scores for each evaluation criterion, clarity, detailing, language adequacy, and content quality, across four types of audience-adapted texts: (A) for children, (B) for the general population, (C) for health managers, and (D) for social media. Each colored bubble represents a different population of evaluators, as indicated in the legend on the right. The size of each bubble is proportional to the number of respondents from that population who rated the corresponding question for each text type.

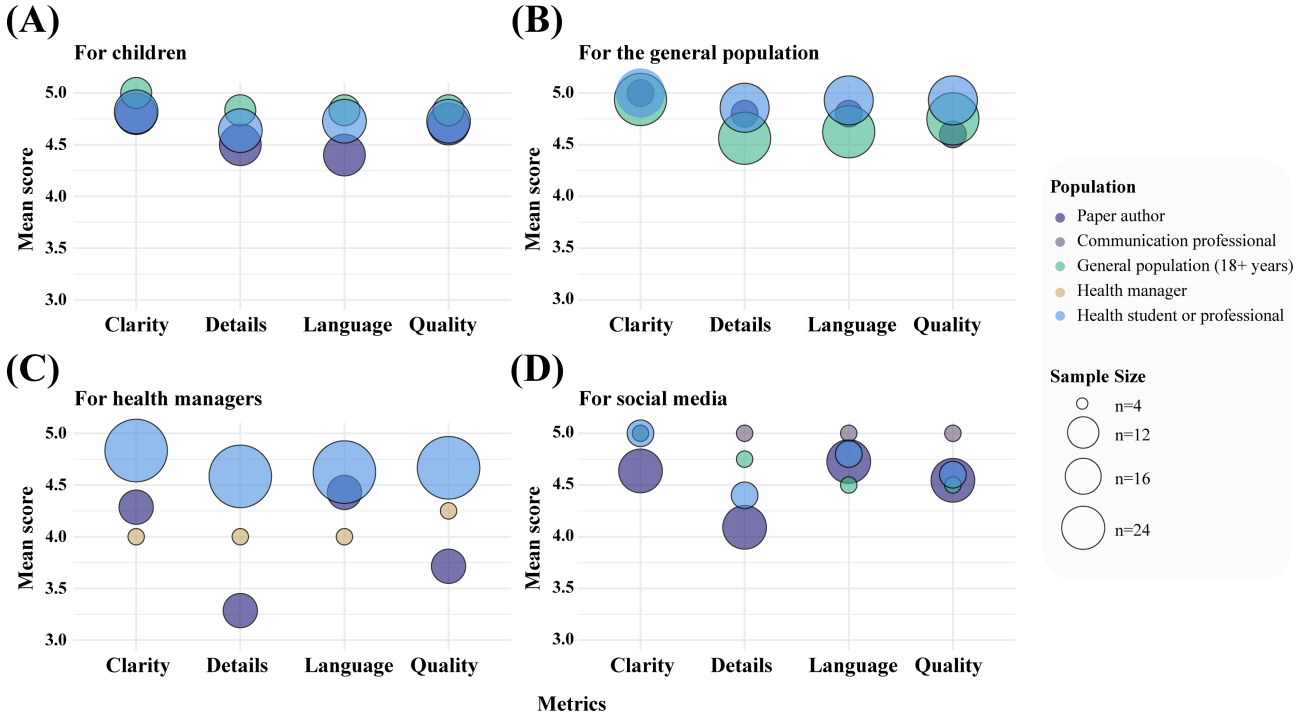
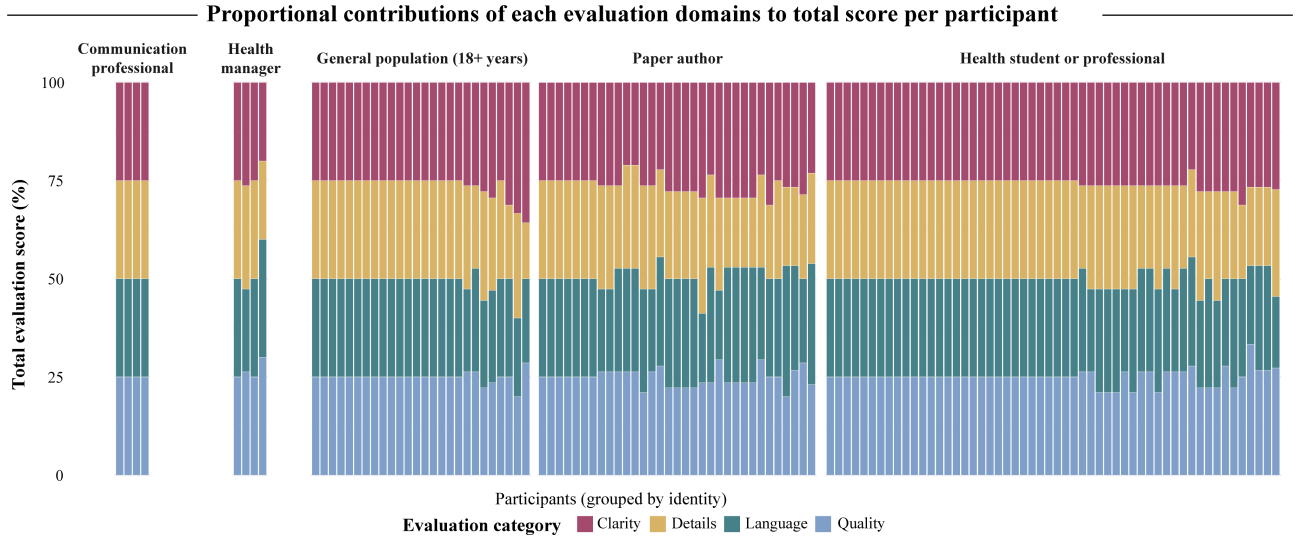


Figure 3. Proportional contributions of evaluation domains to each participant's total score, grouped by reviewer identity. A 100% stacked bar chart in which each vertical bar represents an individual participant's normalized total evaluation score (scaled to 100%), with participants organized along the x-axis and grouped by self-reported identity: paper author, communication professional, general population (18+ years), health manager, and health student or professional. Within each bar, colored segments depict the relative weight of each of the 4 evaluation categories, quality (light blue), language (teal), detailing (gold), and clarity (maroon), in that participant's overall total rating (sum of individual 1-5 domain scores). The y-axis indicates the percentage contribution of each category to the participant's total score.



XSL•FO
RenderX

Table . Average (SD) evaluation scores by audience and stakeholder group for each adapted output. This table presents the central tendency and dispersion (mean, SD) of four evaluation domains: (1) clarity, (2) detailing, (3) language suitability, and (4) content quality, across four target - audience contexts ("for children," "for health managers," "for social media," and "for the general public"). Within each context, scores are shown separately for each evaluator subgroup (authors, general public, students or health professionals, health managers, and communication professionals). All ratings were provided on a 5 - point scale ranging from 1 - 5.

Target audience and evaluator	Clarity	Detailing	Language suitability	Content quality
For the general public, mean (SD)				
Authors	5.00 (0.00)	4.80 (0.45)	4.80 (0.45)	4.60 (0.55)
General public	4.94 (0.25)	4.56 (0.89)	4.62 (0.62)	4.75 (0.45)
Students or health professionals	5.00 (0.00)	4.86 (0.36)	4.93 (0.27)	4.93 (0.27)
For children, mean (SD)				
Authors	4.80 (0.42)	4.50 (0.71)	4.40 (0.84)	4.70 (0.48)
General public	5.00 (0.00)	4.83 (0.41)	4.83 (0.41)	4.83 (0.41)
Students or health professionals	4.82 (0.40)	4.64 (0.67)	4.73 (0.65)	4.73 (0.47)
For health managers, mean (SD)				
Authors	4.29 (0.76)	3.29 (0.49)	4.43 (0.53)	3.71 (0.49)
Students or health professionals	4.83 (0.48)	4.58 (0.72)	4.62 (0.71)	4.67 (0.56)
Health managers	4.00 (1.41)	4.00 (1.41)	4.00 (0.82)	4.25 (0.96)
For social media, mean (SD)				
Authors	4.64 (0.50)	4.09 (0.83)	4.73 (0.65)	4.55 (0.52)
General public	5.00 (0.00)	4.75 (0.50)	4.50 (1.00)	4.50 (1.00)
Students or health professionals	5.00 (0.00)	4.40 (0.89)	4.80 (0.45)	4.60 (0.55)
Communication professionals	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)

In addition to quantitative ratings, anonymous evaluators were invited to provide open-ended comments on the chatbot-generated materials. A total of 68 comments were collected, distributed across the evaluated target audiences: "for children" (n=14), "for the general public" (n=27), "for health managers" (n=11), and "for social media" (n=16; Table 3). Across all categories, the overall tone of the comments was positive and constructive. For children's materials, praise regarding language was the most frequent (8/14, 57.1%), while 42.9% (6/14) of comments did not address language. Information praise (4/14, 28.6%) and information criticism (4/14, 28.6%) were also observed. Only 7.1% (1/14) of comments suggested the material did not fully comply with its intended proposal. For health manager materials, comments were more evenly distributed: information criticism (4/11, 36.3%), language praise (4/11, 36.3%), and proposal compliance (9/11, 81.8%) were most common, though 18.2% (2/11) indicated noncompliance with the proposal. For social media materials, the highest frequency of feedback related to information criticism (6/16, 37.5%) and structure criticism (4/16, 25%), reflecting the platform-specific communication challenges. A majority (13/16, 81.3%) of comments judged the materials as compliant with the intended proposal. For general public materials, language praise was dominant (18/27, 66.7%), with a smaller proportion of comments addressing information (only 29.6% provided any information-related feedback). Most comments (23/27, 85.2%) indicated that the outputs complied with the intended proposal. Across all targets, the high proportion of proposal compliance comments indicates strong overall alignment between the outputs and their target audiences. A full compilation of the anonymous comments is provided in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Distribution of types of comments by target audience. This table displays the frequency (n) and percentage (%). Comments were categorized into 4 domains: language, information, structure, and proposal. For each domain, comments were classified as criticism, praise, or not applicable (for language, information, and structure), or as complies with proposal or does not comply with proposal (for proposal). Comments could be assigned to multiple domains and categories.

Category	For children (n=14)	For health managers (n=11)	For social media (n=16)	For the general public (n=27)
Language, n (%)				
Criticism	0 (0)	1 (9.1)	2 (12.5)	2 (7.4)
Praise	8 (57.1)	4 (36.3)	5 (31.2)	18 (66.7)
Not applicable	6 (42.9)	6 (54.6)	9 (56.3)	7 (25.9)
Information, n (%)				
Criticism	4 (28.6)	4 (36.3)	6 (37.5)	5 (18.5)
Praise	4 (28.6)	2 (18.2)	1 (6.2)	3 (11.1)
Not applicable	6 (42.9)	5 (45.5)	9 (56.3)	19 (70.4)
Structure, n (%)				
Criticism	3 (21.4)	2 (18.2)	4 (25)	2 (7.4)
Praise	3 (21.4)	0 (0)	0 (0)	0 (0)
Not applicable	8 (57.1)	9 (81.8)	11 (68.8)	25 (92.6)
Proposal, n (%)				
Does not comply with the proposal	1 (7.1)	2 (18.2)	3 (18.7)	4 (14.8)
Complies with the proposal	13 (92.9)	9 (81.8)	13 (81.3)	23 (85.2)

Comparative Accuracy and Context Stability Evaluation

A detailed comparison between Maria Ciência and the base GPT-4.5 model was conducted across 4 conversational criteria: establishment of context, continuity of conversation, resilience to interruption, and return to context, stratified by target audience (Table S1 in Multimedia Appendix 1). For Maria Ciência, mean scores were high across all target audiences and criteria. For the general public, Maria Ciência achieved perfect continuity of conversation (5.00, SD 0.00) and high stability across all other dimensions (establishment of context: 4.75, SD 0.50; interruption of context: 4.75, SD 0.50; and return to context: 4.75, SD 0.50). In comparison, the base GPT-4.5 model exhibited lower mean scores across several criteria.

For children, Maria Ciência again showed higher mean scores, with perfect scores for interruption of context (5.00, SD 0.00) and strong scores across other criteria (establishment of context: 4.66, SD 0.58; continuity: 4.66, SD 0.58; and return to context: 4.66, SD 0.58). The base GPT-4.5 model showed substantially lower performance in this category, particularly in continuity (3.00, SD 0.82) and establishment of context (3.25, SD 0.50), indicating challenges in maintaining audience-appropriate conversation flow for younger users.

For social media outputs, scores remained high, with means of 4.50 (SD 0.58) for establishment of context and continuity of conversation, and 5.00 (SD 0.00) for criterion resilience to interruption. In contrast, the base GPT-4.5 model exhibited greater variability, particularly in outputs for children and health managers. For children, mean scores were notably lower in the establishment of context (3.25, SD 0.96), continuity of conversation (3.00, SD 0.82), and resilience to interruption (3.00, SD 0.82), highlighting difficulties in maintaining and recovering conversational context. For health managers, while context establishment remained high (4.75, SD 0.50), performance dropped in criteria resilience to interruption (3.75, SD 0.96) and continuity of conversation (4.25, SD 0.50). Across all targets, the base GPT-4.5 model showed more frequent context drift and reduced continuity compared to Maria Ciência (Table S1 in Multimedia Appendix 1).

Readability Evaluation

Readability metrics derived from the ALT framework revealed significant differences between texts generated by Maria Ciência and those generated by the base GPT-4.5 model. For materials designed for children, the overall grade level was higher for Maria Ciência outputs compared to base GPT-4.5 (MC: 8.23, SD 0.96 vs GPT: 7.56, SD 1.03; P=.002). While all texts from both generators were classified as having high legibility, differences emerged across several quantitative readability indices. Maria Ciência texts showed lower Flesch reading ease scores (MC: 67.1, SD 5.51 vs GPT: 71.3, SD 5.80; P=.001) and higher Flesch-Kincaid grade levels (MC: 7.58, SD 0.98 vs GPT: 6.70, SD 0.98; P<.001), indicating increased textual complexity. Similar patterns were observed for the adjusted Gunning fog index (P<.001), the ARI (P=.004), the number of words per sentence (P<.001), and several syllables per word (P=.002). No statistically significant differences were observed between generators for the Coleman-Liau index, letters per word, or proportion of complex words (Table 4).

XSL•FO
RenderX

Table . Readability analysis of texts. This table presents the central tendency and dispersion (mean, SD) of readability metrics, as well as the frequency (n) and percentage (%) for categorical variables.

	GPT base	Maria Ciência	*P* value
For children			
Overall grade level, mean (SD)	7.56 (1.03)	8.23 (0.96)	.002
Recommended age group (year), n (%)			.08
11 - 14	39 (83)	30 (63.8)	
15 - 18	8 (17)	16 (36.2)	
Legibility result, n (%)			—[a]
High	47 (100)	47 (100)	
Flesch reading ease, mean (SD)	71.3 (5.80)	67.1 (5.51)	.001
Gulpease index, mean (SD)	69.6 (4.26)	1104 (6883)	.32
Flesch-Kincaid grade level, mean (SD)	6.70 (0.98)	7.58 (0.98)	<.001
Adjusted Gunning fog index, mean (SD)	8.12 (0.85)	8.85 (0.79)	<.001
Automated readability index, mean (SD)	6.34 (1.24)	7.09 (1.18)	.004
Coleman-Liau index, mean (SD)	9.23 (1.21)	9.64 (1.05)	.09
Letters per word, mean (SD)	4.65 (0.21)	4.68 (0.20)	.51
Syllables per word, mean (SD)	1.98 (0.08)	2.03 (0.07)	.002
Words per sentence, mean (SD)	11.2 (1.44)	12.5 (1.45)	<.001
Complex words (%), mean (SD)	13.7 (3.44)	14.6 (2.44)	.15
For the health manager			
Overall grade level, mean (SD)	13.7 (1.70)	14.4 (2.05)	.11
Recommended age group (year), n (%)			.13
15 - 18	10 (21.3)	5 (10.6)	
Can be easily understood by university students	30 (63.8)	35 (74.5)	
For people with a college degree	7 (14.9)	3 (6.4)	
Extremely difficult text	0 (0)	4 (8.5)	
Legibility result, n (%)			.32
High	10 (21.3)	5 (10.6)	
Medium	34 (72.3)	37 (78.7)	
Low	3 (6.4)	5 (10.6)	
Flesch reading ease, mean (SD)	27.8 (8.13)	32.6 (8.61)	.008
Gulpease index, mean (SD)	53.7 (6.72)	50.0 (4.36)	.003
Flesch-Kincaid grade level, mean (SD)	13.5 (1.81)	14.0 (2.22)	.24
Adjusted Gunning fog index, mean (SD)	11.5 (2.20)	14.0 (2.81)	<.001
Automated readability index, mean (SD)	13.3 (2.06)	14.2 (2.64)	.08
Coleman-Liau index, mean (SD)	16.4 (1.21)	15.0 (1.11)	<.001
Letters per word, mean (SD)	5.95 (0.21)	5.59 (0.22)	<.001

XSL•FO
RenderX

	GPT base	Maria Ciência	*P* value
Syllables per word, mean (SD)	2.56 (0.10)	2.42 (0.11)	<.001
Words per sentence, mean (SD)	13.5 (4.17)	19.2 (6.09)	<.001
Complex words (%), mean (SD)	25.9 (3.44)	23.9 (3.08)	.006
For social media			
Overall grade level, mean (SD)	8.89 (2.05)	11.2 (1.59)	<.001
Recommended age group (years), n (%)			<.001
11 - 14	20 (42.6)	1 (2.1)	
15 - 18	24 (51)	40 (85.1)	
Can be easily understood by university students	3 (6.4)	6 (12.8)	
Legibility result, n (%)			.22
High	43 (91.5)	39 (83)	
Medium	4 (8.5)	8 (17)	
Flesch reading ease, mean (SD)	65.5 (12.8)	50.7 (8.95)	<.001
Gulpease index, mean (SD)	64.4 (6.92)	57.1 (4.27)	<.001
Flesch-Kincaid grade level, mean (SD)	8.02 (2.26)	10.7 (1.75)	<.001
Adjusted Gunning fog index, mean (SD)	9.49 (1.72)	10.9 (1.95)	<.001
Automated readability index, mean (SD)	7.91 (2.52)	10.7 (1.97)	<.001
Coleman-Liau index, mean (SD)	10.1 (2.20)	12.4 (1.50)	<.001
Letters per word, mean (SD)	4.76 (0.38)	5.14 (0.29)	<.001
Syllables per word, mean (SD)	2.03 (0.16)	2.20 (0.12)	<.001
Words per sentence, mean (SD)	13.5 (2.97)	16.0 (4.10)	.002
Complex words (%), mean (SD)	15 (4.83)	16.2 (3.12)	.17
For the general public			
Overall grade level, mean (SD)	10.5 (1.22)	9.96 (1.00)	.026
Recommended age group (year), n (%)			.52
11 - 14	3 (6.4)	2 (4.3)	
15 - 18	41 (87.2)	44 (93.6)	
Can be easily understood by university students	3 (6.4)	1 (2.1)	
Legibility result, n (%)			.31
High	44 (93.6)	46 (97.9)	
Medium	3 (6.4)	1 (2.1)	
Flesch reading ease, mean (SD)	53.1 (7.35)	58.7 (5.87)	<.001
Gulpease index, mean (SD)	59.5 (4.05)	60.5 (3.08)	.21
Flesch-Kincaid grade level, mean (SD)	9.92 (1.24)	9.33 (1.05)	.02
Adjusted Gunning fog index, mean (SD)	9.96 (1.23)	10.4 (1.11)	.07
Automated readability index, mean (SD)	9.83 (1.47)	9.02 (1.21)	.06

XSL•FO
RenderX

	GPT base	Maria Ciência	*P* value
Coleman-Liau index, mean (SD)	12.2 (1.39)	10.9 (1.00)	<.001
Letters per word, mean (SD)	5.14 (0.25)	4.86 (0.18)	<.001
Syllables per word, mean (SD)	2.20 (0.11)	2.10 (0.08)	<.001
Words per sentence, mean (SD)	14.0 (2.38)	15.1 (1.83)	.02
Complex words (%), mean (SD)	16.2 (4.51)	15.8 (3.06)	.62

[a]Not available.

The readability analyses of texts generated for health managers revealed overall comparable patterns between Maria Ciência and the base GPT-4.5 model, with few statistically significant differences across metrics. The overall grade level did not differ significantly between generators (*P*=.11), as well as the classification by recommended audience level (*P*=.13), and the overall legibility categories (*P*=.32). Despite these similarities, Maria Ciência texts exhibited higher Flesch reading ease scores compared to base GPT-4.5 (MC: 32.6, SD 8.61 vs GPT: 27.8, SD 8.13; *P*=.008), higher Gunning fog index (MC: 14.0, SD 2.81 vs GPT: 11.5, SD 2.20; *P*<.001), alongside lower Gulpease index values (MC: 50.0, SD 4.36 vs GPT: 53.7, SD 6.72; *P*=.003), if compared to base GPT 4.5 model (Table 4). At the lexical and syntactic levels, Maria Ciência texts contained fewer letters per word (*P*<.001) and fewer syllables per word (*P*<.001), but substantially longer sentences (*P*<.001). The proportion of complex words was slightly lower in Maria Ciência outputs (23.9%, SD 3.08% vs 25.9%, SD 3.44%; *P*=.006; Table 4).

For materials intended for social media, Maria Ciência outputs exhibited a substantially higher overall grade level compared to base GPT-4.5 (MC: 11.2, SD 1.59 vs GPT: 8.89, SD 2.05; *P*<.001). Consistent with this pattern, having high legibility across multiple recommended age group classifications differed significantly between generators (*P*<.001). While base GPT-4.5 outputs were more frequently classified as suitable for younger adolescents (11 - 14 y), the majority of Maria Ciência's texts were classified as appropriate for older adolescents (15 - 18 y). Despite these differences in textual complexity, overall legibility categories (high vs medium) were comparable between generators (*P*=.22), with most texts in both groups classified as having high legibility. Across multiple readability indices, Maria Ciência's texts consistently demonstrated increased linguistic complexity. These outputs showed lower Flesch reading ease scores (MC: 50.7, SD 8.95 vs GPT: 65.5, SD 12.8; *P*<.001) and lower Gulpease index values (MC: 57.1, SD 4.27 vs GPT: 64.4, SD 6.92; *P*<.001), alongside higher Flesch-Kincaid grade levels (MC: 10.7 (SD 1.75 vs GPT: 8.02, SD 2.26; *P*<.001). Similar trends were observed for the adjusted Gunning fog index, ARI, and Coleman-Liau index (*P*<.001). At the lexical and syntactic levels, Maria Ciência outputs contained longer words, reflected by higher letters-per-word and syllables-per-word ratios (both *P*<.001), as well as longer sentences (*P*=.002). The proportion of complex words did not differ significantly between generators (Table 4).

Readability analyses indicated modest but statistically significant differences between texts generated by Maria Ciência and those generated by the base GPT-4.5 model for materials aimed at the general population. Maria Ciência outputs exhibited a slightly lower overall grade level compared to base GPT-4.5 (MC: 9.96, SD 1.00 vs GPT: 10.5, SD 1.22; *P*=.03). Recommended age group classifications were similar between generators (*P*=.52), with the majority of texts in both groups classified as suitable for adolescents aged 15 - 18 years. Overall legibility categories (high vs medium) were also comparable (*P*=.31), with most texts classified as having high legibility. Several readability indices suggested improved linguistic accessibility in Maria Ciência outputs. These texts showed higher Flesch Reading Ease scores (MC: 58.7, SD 5.87 vs GPT: 53.1, SD 7.35; *P*<.001) and lower Flesch–Kincaid Grade Levels (MC: 9.33, SD 1.05 vs GPT: 9.92, SD 1.24; *P*=.02). The Coleman-Liau Index was also lower for Maria Ciência texts (MC: 10.9, SD 1.00 vs GPT: 12.2, SD 1.39; *P*<.001), indicating reduced lexical complexity. At the word level, Maria Ciência outputs contained fewer letters per word (MC: 4.86, SD 0.18 vs GPT: 5.14, SD 0.25; *P*<.001) and fewer syllables per word (MC: 2.10, SD 0.08 vs GPT: 2.20, SD 0.11; *P*<.001). In contrast, Maria Ciência texts exhibited a slightly higher number of words per sentence compared to base GPT-4.5 outputs (MC: 15.1, SD 1.83 vs 14.0, SD GPT: 2.38; *P*=.02). No statistically significant differences were observed between generators in the proportion of complex words (Table 4).

When directly compared with human-written texts, both base GPT-4.5 and Maria Ciência showed systematic and statistically significant differences across nearly all readability metrics. Human texts consistently exhibited higher overall grade levels (15.6, SD 1.96) than base GPT-4.5 (11.0, SD 1.15) and Maria Ciência (9.7, SD 0.67), indicating substantially greater linguistic complexity. This pattern was reinforced by multiple indices: human abstracts had lower Flesch reading ease scores and higher Flesch-Kincaid, Gunning fog, ARI, and Coleman-Liau indices, alongside longer sentences and a higher proportion of complex words (all with *P*<.01, Table 5). In contrast, both Maria Ciência and base GPT-4.5 generated texts that were easier to read, shorter at the sentence level, and lexically simpler. Importantly, while both systems differed from human writing in the same direction, Maria Ciência outputs tended to be consistently simpler than base GPT-4.5 across most metrics, including reading ease, grade level, sentence length, and lexical density (Table 5).

XSL•FO
RenderX

Table . Differences between Maria Ciência, base GPT 4.5, and human-written texts. This table presents the central tendency and dispersion (mean, SD) of readability metrics, as well as the frequency (n) and percentage (%) for categorical variables.

	GPT base°	Maria Ciência	Human	*P* value
Overall grade level, mean (SD)	11.0 (1.15)	9.70 (0.67)	15.6 (1.96)	<.001
Recommended age group, n (%)				<.001
11 - 14 year	0 (0)	1 (10)	0 (0)	
15 - 18 year	9 (90)	9 (90)	1 (10)	
Can be easily understood by university students	1 (10)	0 (0)	3 (30)	
For people with a college degree	0 (0)	0 (0)	6 (60)	
Legibility result, n (%)				<.001
High	9 (90)	10 (100)	1 (10)	
Medium	1 (10)	0 (0)	5 (50)	
Low	0 (0)	0 (0)	4 (40)	
Flesch reading ease, mean (SD)	50.6 (6.93)	58.7 (5.81)	32.4 (10.5)	<.001
Gulpease index, mean (SD)	57.8 (3.52)	60.9 (2.84)	47.1 (3.53)	<.001
Flesch-Kincaid grade level, mean (SD)	10.6 (1.10)	9.10 (0.77)	15.4 (1.92)	<.001
Adjusted Gunning fog index, mean (SD)	11.1 (1.43)	10.1 (0.88)	16.6 (2.06)	<.001
Automated readability index, mean (SD)	10.4 (1.32)	8.91 (0.95)	16.1 (2.10)	<.001
Coleman-Liau index, mean (SD)	12.3 (1.48)	10.9 (0.86)	14.2 (1.58)	<.001
Letters per word, mean (SD)	5.14 (0.28)	4.90 (0.17)	5.35 (0.28)	.002
Syllables per word, mean (SD)	2.20 (0.11)	2.08 (0.08)	2.32 (0.11)	<.001
Words per sentence, mean (SD)	15.5 (3.18)	14.6 (1.64)	26.0 (3.29)	<.001
Complex words (%), mean (SD)	18.6 (3.94)	15.6 (2.67)	20.6 (4.40)	.02

Discussion

This study provides exploratory evidence that Maria Ciência is capable of producing audience-adapted science communication materials that are perceived as clear, accessible, and linguistically appropriate across a range of stakeholders. Materials tailored for children and the general public were especially well received, while outputs for health managers showed greater variability, reflecting the distinct informational demands of this audience. The more critical feedback from original paper authors highlights the inherent challenge of balancing scientific precision with public accessibility.

Qualitative feedback from anonymous evaluators reinforced these trends, highlighting the clarity and perceived usefulness of the materials and offering constructive suggestions to further adapt tone and terminology for specific audiences. These results are encouraging, particularly in light of the urgent need to address health misinformation, which continues to erode public trust in science, hinder the implementation of health policies, and contribute to adverse health outcomes [1,2]. The COVID-19 pandemic brought the urgency of this issue into sharp focus, with widespread infodemics interfering with disease prevention efforts and amplifying avoidable harm [2], emphasizing the importance of tools that support effective science communication.

By enabling the production of trusted, audience-specific materials, Maria Ciência can complement existing strategies for combating misinformation, which traditionally rely on reactive fact-checking or broad public health campaigns [16]. Within the scope of this evaluation, Maria Ciência demonstrated superior conversational stability and contextual accuracy, particularly for sensitive audiences such as children, critical for fostering health literacy from an early age. Furthermore, previous studies have shown that GPT-based assistants

XSL•FO
RenderX

specifically trained or configured with domain-relevant and ethically curated materials achieve higher performance and contribute to greater user trust and acceptance of the platform in public health and educational contexts [5,6]. The approach adopted by Maria Ciência, combining prompt engineering with thematic supervision, is consistent with these findings and reinforces the value of domain-specific configuration for science communication tools.

Beyond model performance, Maria Ciência was designed to reduce practical barriers that limit researchers' engagement with public audiences. Despite growing recognition of science communication as a core responsibility, academic structures still prioritize publication in peer-reviewed journals over community outreach [3]. Researchers often lack time, institutional support, or training to translate their findings into accessible formats. The data presented here suggest that even minimal support from tools such as Maria Ciência can enable more scientists to participate meaningfully in public engagement by providing formats aligned with the needs of different audiences and reducing technical barriers to communication.

Complementing subjective evaluations, the readability analyses provide an objective lens through which to interpret these findings. Texts generated by Maria Ciência exhibited a deliberate balance between linguistic complexity and overall legibility. In some audience categories, modest increases in grade level, sentence length, or syllabic density were observed; however, these changes did not consistently translate into lower legibility classifications. Instead, they reflect structural adjustments aimed at preserving informational completeness rather than a loss of accessibility. Readability indices such as the Flesch reading ease, Flesch-Kincaid grade level, Gunning fog index, ARI, Coleman-Liau index, and Gulpease index are widely used in health communication research as proxies for accessibility and audience appropriateness [17,18]. Their combined use is particularly valuable for comparative analyses, as it allows systematic assessment of how different communication strategies modulate linguistic complexity while maintaining interpretability. Importantly, readability metrics are not end points: they function as proxies that can be linked to downstream outcomes such as comprehension, perceived credibility, and cognitive load, especially in settings where audiences must navigate competing narratives and misinformation. Lower linguistic complexity and shorter syntactic structures have been associated with improved comprehension and reduced cognitive load, particularly among audiences with limited health literacy, and may support trust-building when messages must compete with misleading narratives in digital environments [17,19,20]. Although readability alone cannot guarantee understanding or prevent misinformation, it provides an objective benchmark for whether outputs are likely to be approachable for nonspecialist readers and therefore serves as a useful complement to subjective quality ratings.

Direct comparisons with human-written science communication materials further contextualize these results. Human-authored texts consistently demonstrated substantially higher linguistic complexity across nearly all indices, including grade level, sentence length, and lexical density, reinforcing longstanding evidence that expert-driven communication frequently exceeds recommended readability thresholds for lay audiences. In contrast, both Maria Ciência and the base GPT-4.5 model produced texts that were systematically easier to read and more linguistically accessible. While both systems diverged from human writing in the same direction, Maria Ciência's outputs were consistently simpler than those generated by the base model across most metrics. Together, these findings suggest that Maria Ciência does not indiscriminately simplify content but rather modulates textual structure in an audience-dependent manner, supporting improved accessibility without compromising overall legibility.

Differences observed for health managers are also conceptually expected. Communication for managerial or policy-facing audiences typically prioritizes actionability, decision relevance, and structured synthesis rather than simplified language alone. As a result, evaluation criteria that work well for general audiences may undercapture what makes output useful for managers or policymakers, and future evaluations should incorporate audience-specific criteria tailored to decision-making contexts.

From an implementation perspective, these characteristics position Maria Ciência as a practical support tool for knowledge translation workflows. Potential applications include integration into journal dissemination processes, public health communication teams, and institutional outreach initiatives, where rapid generation of audience-specific summaries can support timely and responsible dissemination of peer-reviewed evidence. Importantly, such integration should remain coupled with human oversight to ensure scientific fidelity and contextual appropriateness. For example, a journal or research group could (1) identify a newly accepted open-access paper, (2) generate a public-facing summary and a social media post using Maria Ciência, (3) generate a brief manager-facing version highlighting actionable implications, (4) conduct a rapid author or communications-team review for scientific fidelity and tone, and (5) disseminate the finalized outputs through institutional websites, press offices, and social channels. Of note, although not evaluated in this paper, Maria Ciência is also capable of generating images that can help in the dissemination of knowledge. This hypothetical pipeline illustrates how AI-assisted generation can reduce production time while preserving accountability through human review.

Recent advances in AI also point to future directions for audience-centered health communication tools that extend beyond the scope of this study. Federated learning frameworks, for example, have been increasingly explored in health care to enable collaborative model improvement across institutions while preserving data privacy and regulatory compliance [4]. In the context of health communication, such approaches may become relevant for incorporating region- or institution-specific linguistic and cultural patterns without requiring direct sharing of sensitive data. Similarly, developments in self-supervised learning highlight the potential of leveraging large volumes of unlabeled data to improve contextual understanding and adaptability of language models. In health care and related domains, self-supervised approaches have been proposed to enhance prediction and pattern recognition while reducing

XSL•FO
RenderX

dependence on costly manual annotation [21]. Although not applied in the present work, these principles suggest possible future pathways for improving the adaptability and robustness of AI-assisted communication systems, particularly in settings where labeled training data are scarce or ethically constrained. Together, these developments emphasize that future research should explore how privacy-preserving and data-efficient learning paradigms can complement human oversight in responsible AI-driven health communication.

As AI-driven science communication tools gain increasing relevance worldwide, it is important to consider their potential for application across diverse international contexts. Although initially developed and evaluated in Brazilian Portuguese, the architecture and prompting framework of Maria Ciência, being built on the GPT platform, allow the tool to operate in over 50 languages with the same ethical and technical standards [14]. This provides immediate potential for multilingual deployment in global health communication efforts. However, while the tool can linguistically adapt to multiple languages, effective application across different regions also requires attention to cultural nuances, health literacy variations, and sociolinguistic differences. Linguistic translation alone is insufficient; local framing, terminology, and narrative styles must be carefully adapted to ensure meaningful engagement and to support health literacy and misinformation prevention across diverse public health ecosystems. Accordingly, future work should test culturally grounded adaptation workflows, ideally combining local expert review with community feedback.

Despite these strengths, this study has limitations. The number of evaluators per target audience was not standardized, and some subgroups, such as health managers and communication professionals, were underrepresented. Author evaluations, while informative, were based on subjective perceptions and not validated against formal scientific fidelity criteria. The evaluation also focused on perception and usability rather than long-term impacts on knowledge retention or behavioral outcomes. Third, children were not included as evaluators. As a result, evaluations of children-focused materials reflect adult perceptions and should be interpreted as preliminary. Furthermore, the tool was primarily tested in Brazilian Portuguese; while the underlying architecture is multilingual, additional research is needed to validate its performance in other languages and cultural contexts.

Even with these limitations, the evaluation provides a strong foundation to support the validity and utility of the tool. The systematic and transparent development process, combined with broad stakeholder engagement, underscores the potential of Maria Ciência to contribute meaningfully to health literacy and misinformation prevention. Its ability to outperform a baseline GPT model in contextual stability and accuracy further validates the importance of domain-specific configuration for public health applications. Future evaluations will adopt a stratified design in which materials are assessed exclusively by their intended target audiences, using validated instruments and balanced stakeholder representation.

In summary, Maria Ciência offers a promising avenue for enhancing knowledge translation and addressing the communication challenges posed by health misinformation. As international organizations and national governments have emphasized, combating the impacts of infodemics requires more than reactive correction. It necessitates proactive investment in tools and strategies that produce accessible, trusted, and culturally relevant information. By supporting the generation of audience-adapted materials and facilitating researcher engagement in public dialogue, Maria Ciência contributes to this broader effort to advance scientific literacy and strengthen public health resilience.

Acknowledgments

We gratefully acknowledge Ms Luisa Vinagre and Mr Rodrigo Pontes for their valuable contributions in generating the chatbot outputs used in this study. We also extend our sincere thanks to all anonymous respondents who participated in the evaluation forms, and to Ms Elze Leite for her administrative support throughout the project.

Funding

This study was supported by the Intramural Research Program of the Fundac¸a˜o Oswaldo Cruz (BBA). BS-P received a fellowship from the Fundação de Amparo à Pesquisa da Bahia (FAPESB), KV-S, JNC-O, GP-R, SDM, RRdCS, and MA-P, received fellowships from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil. BdBA is a sênior fellow from the CNPq.

Authors' Contributions

Conceptualization: MA-P, KV-S, BdBA
Data verification and curation: BS-P, KV-S
Formal analysis: MA-P, BS-P, GP-R, JNC-O, KV-S
Funding acquisition: GP-R, BdBA
Investigation: MA-P, BS-P, KV-S, SDM, JNC-O, RRdCS, BdBA
Methodology: MA-P, KV-S, SDM, JNC-O, GP-R
Project administration: MA-P
Resources: MA-P, BdBA

XSL•FO
RenderX

Software: MA-P, BS-P, GP-R, JNC-O, KV-S
Supervision: MA-P, BdBA
Writing – original draft: MA-P, KV-S, GP-R, BS-P
Writing – review and editing: all authors

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of comments, and comparative accuracy and context stability evaluation table.
[DOCX File, 26 KB - infodemiology_v6i1e78843_app1.docx]

References

1. Oliveira TD. Como enfrentar a desinformação científica? Desafios sociais, políticos e jurídicos intensificados no contexto da pandemia [Article in Portuguese]. Liinc Rev 2020 Dec 23;16(2):e5374. [doi: 10.18617/liinc.v16i2.5374]
2. Caceres MMF, Sosa JP, Lawrence JA, et al. The impact of misinformation on the COVID-19 pandemic. AIMS Public Health 2022;9(2):262-277. [doi: 10.3934/publichealth.2022018] [Medline: 35634019]
3. Araújo-Pereira M, Andrade BB. The heroic journey of young Brazilian scientists: challenges and opportunities. Lancet Reg Health Am 2024 May;33:100735. [doi: 10.1016/j.lana.2024.100735] [Medline: 38590323]
4. Abbas SR, Abbas Z, Zahir A, Lee SW. Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. Healthcare (Basel) 2024 Dec 22;12(24):2587. [doi: 10.3390/healthcare12242587] [Medline: 39766014]
5. Hakim VGA, Paiman NA, Rahman MHS. Genie‐on‐demand: a custom AI chatbot for enhancing learning performance, self‐efficacy, and technology acceptance in occupational health and safety for engineering education. Comp Appl Eng 2024 Nov;32(6):e22800. [doi: 10.1002/cae.22800]
6. Pu J, Hong J, Yu Q, et al. Accuracy, satisfaction, and impact of custom GPT in acquiring clinical knowledge: potential for AI-assisted medical education. Med Teach 2025 Sep;47(9):1502-1508. [doi: 10.1080/0142159X.2025.2458808] [Medline: 39893644]
7. Cunda BV. Na ponta das línguas: pequeno glossário para apoiar o enfrentamento do estigma e da discriminação [Bachelor's thesis] [Article in Portuguese]. : Universidade Federal do Rio Grande do Sul; 2023 URL: https://lume.ufrgs.br/handle/10183/267984 [accessed 2026-02-17] [Medline: 38071171]
8. UNAIDS terminology guidelines. UNAIDS. 2024. URL: https://www.unaids.org/en/resources/documents/2024/terminology_guidelines [accessed 2026-02-17]
9. Terminology guidelines. UNICEF. 2016. URL: https://www.unicef.org/documents/terminology-guidelines [accessed 2026-02-17]
10. Frontiers for young minds. Frontiers. URL: https://kids.frontiersin.org/ [accessed 2026-02-17]
11. Antonof G, Rodrigues M, Schmidt M. Diálogos: um guia para divulgação científica em linguagem simples [Article in Portuguese]. : Universidade de Brasilia; 2023 URL: https://bdm.unb.br/bitstream/10483/38689/2/2023_GabrielAntonoff_MateusRodrigues_MateusSilva_produto.pdf [accessed 2026-02-17]
12. Paraguassu LB, da Costa CV. ICICT simple language guide [Article in Portuguese]. : Oswaldo Cruz Foundation. Institute of Scientific and Technological Communication and Information in Health; 2023 URL: https://www.arca.fiocruz.br/handle/icict/61432 [accessed 2026-02-17]
13. Profissão Biotec. URL: https://profissaobiotec.com.br/ [accessed 2026-02-17]
14. How to change your language setting in ChatGPT. OpenAI. URL: https://help.openai.com/en/articles/8357869-how-to-change-your-language-setting-in-chatgpt [accessed 2026-02-17]
15. Moreno GDL, de Souza MPM, Hein N, Hein AK. ALT: um software para análise de legibilidade de textos em língua portuguesa [Article in Portuguese]. Policromias 2023;8(1):91-128. [doi: 10.61358/policromias.v8i1.54352]
16. Tiwary A, Rimal A, Paudyal B, Sigdel KR, Basnyat B. Poor communication by health care professionals may lead to life-threatening complications: examples from two case reports. Wellcome Open Res 2019;4:7. [doi: 10.12688/wellcomeopenres.15042.1] [Medline: 31448336]
17. Boutemen L, Miller AN. Readability of publicly available mental health information: a systematic review. Patient Educ Couns 2023 Jun;111:107682. [doi: 10.1016/j.pec.2023.107682] [Medline: 36944285]
18. Mendez SR, Galvez SMN, Emmons KM, Viswanath K. Reading grade level estimates of public health communication on social media vary due to text preprocessing. J Health Commun 2025;30(10-12):634-640. [doi: 10.1080/10810730.2025.2583460] [Medline: 41204693]
19. Eslami H. The effect of syntactic simplicity and complexity on the readability of the text. JLTR 2014 Sep 1;5(5):1185-1191. [doi: 10.4304/jltr.5.5.1185-1191]

XSL•FO
RenderX

20. Mac O, Ayre J, Bell K, McCaffery K, Muscat DM. Comparison of readability scores for written health information across formulas using automated vs manual measures. JAMA Netw Open 2022 Dec 1;5(12):e2246051. [doi: 10.1001/jamanetworkopen.2022.46051] [Medline: 36508219]
21. Abdulrazzaq MM, Ramaha NTA, Hameed AA, et al. Consequential advancements of self-supervised learning (SSL) in deep learning contexts. Mathematics 2024 Mar 2;12(5):758. [doi: 10.3390/math12050758]

Abbreviations

AI: artificial intelligence
ALT: analysis of language and text
ARI: automated readability index

Edited by M Haupt; submitted 10.Jun.2025; peer-reviewed by SW Lee, X Liang; revised version received 23.Dec.2025; accepted 23.Dec.2025; published 03.Mar.2026.

Please cite as:
Araújo-Pereira M, Villalva-Serra Jr K, Pires-Ramos G, Sousa-Peres B, Conceição-Oliveira JN, Maiche SD, Silva RRDC, Andrade BDB
Audience-Specific Health Communication: Mixed Methods Evaluation of the Maria Ciência AI-Assisted Knowledge Translation Tool
JMIR Infodemiology 2026;6:e78843
URL: https://infodemiology.jmir.org/2026/1/e78843
doi:10.2196/78843

© Mariana Araújo-Pereira, Klauss Villalva-Serra Jr, Gustavo Pires-Ramos, Beatriz Sousa-Peres, Joanã Nascimento Conceição-Oliveira, Sarah Dourado Maiche, Rebeca Rebouças da Cunha Silva, Bruno de Bezerril Andrade. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 3.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Performance of Large Language Models in the Cognitive Analysis of Misinformation: Evaluation Study

Dominika Nadia Wojtczak[1], PhD; Cheryl McQuire[2], PhD; Luisa Zuccolo[3,4], PhD; Claudia Peersman[1], PhD; Ryan McConville[5], PhD

[1]School of Computer Science, University of Bristol, Bristol, United Kingdom

[2]Bristol Medical School, University of Bristol, Bristol, United Kingdom

[3]Health Data Science Centre, Human Technopole, Milan, Italy

[4]Bristol Medical School, Department of Population Health Sciences, University of Bristol, Bristol, United Kingdom

[5]School of Engineering Mathematics and Technology, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Dominika Nadia Wojtczak, PhD
School of Computer Science
University of Bristol
Merchant Venturers Building
75 Woodland Rd
Bristol, BS8 1UB
United Kingdom
Phone: 44 117 928 9000
Email: dominika.nadia.wojtczak@bristol.ac.uk

Abstract

Background: Public discourse is significantly impacted by the rapid spread of misinformation on social media platforms. Human moderators, while capable of performing well, face many challenges due to scalability. While large language models (LLMs) show great potential across various language tasks, their capacity for cognitive and contextual analysis, in detecting and interpreting misinformation, remains less evaluated.

Objective: This study evaluates the effectiveness of LLMs in detecting and interpreting misinformation compared to human annotators, focusing on tasks requiring cognitive analysis and complex judgment. Additionally, we analyze the influence of different prompt engineering strategies on model performance and discuss ethical considerations for using LLMs in content moderation systems.

Methods: We evaluated 4 OpenAI models against a panel of human annotators using a subset of posts from the MuMiN dataset. Each model and human annotator responded to structured questions on misinformation, following an established cognitive framework. Both human annotators and LLMs also provided scores indicating how confident they were in their responses. Various prompting strategies were used in this research, including: 0-shot, few-shot, and chain-of-thought, with performance evaluated through precision, recall, F_1-score, and accuracy. We used statistical tests, including the McNemar test, to quantitatively assess differences between LLM and human ratings of misinformation.

Results: GPT-4 Turbo with chain-of-thought prompting achieved the highest performance of all LLMs for detecting misinformation, with an accuracy of 67.2% and an F_1-score of 78.3%, but was outperformed by human annotators, who achieved 70.1% accuracy and an F_1-score of 81%. LLMs performed well in tasks involving logical reasoning and straightforward misinformation detection, but struggled with complex judgments, including detecting sarcasm, understanding misinformation, and analyzing user intent. LLM confidence scores positively correlated with accuracy in simpler tasks (r=0.72, P<.01) but were less reliable in subjective and complex contextual evaluations.

Conclusions: LLMs show significant potential for automating misinformation detection. Their limitations in understanding and interpreting these posts highlight the current necessity of human oversight. A hybrid framework combining LLMs for preliminary screening with human moderators for more complex evaluation presents a promising future direction. Future research could prioritize the fine-tuning of LLMs using datasets that emphasize cognitive and emotional linguistic features, alongside the development of advanced prompting techniques.

XSL•FO
RenderX

(JMIR Infodemiology 2026;6:e72524) doi:10.2196/72524

KEYWORDS

online misinformation; LLM; fake news; cognitive analysis; panel of annotators; ground truth; intent; reasoning; cognitive reasoning; large language model

Introduction

Online Misinformation

Political leaflets in ancient Rome and the Great Moon Hoax of 1835 remind us that societies have long struggled with falsehoods; however, the scale, speed, and sociotechnical context of the problem have changed radically in the social media era [1].

Social media platforms have become major hubs of misinformation, significantly influencing public opinion, political discourse, and public health decisions [2,3]. Platforms such as X (formerly known as Twitter, X Corp) provide a global space for communication about current affairs, allowing users to share their perspectives widely. Still, they also enable the rapid spread of misinformation, which can distort perceptions of reality, damage trust, and contribute to polarized echo chambers leading to the viral spread of online misinformation [4,5]. Users may explore credibility on social platforms through social factors that influence their perception of misinformation, in addition to factual verification of posts [6].

The current research distinguishes misinformation (false content shared without intent to deceive), disinformation (false content shared with intent), and malinformation (genuine information taken out of context that is shared maliciously to harm individuals, groups, or society). These distinctions highlight the variety of false content, which can differ in origin, intent, and potential harm [7-9]. Nonetheless, detecting these features at scale remains challenging. The MuMiN definition of misinformation is used in this study, which classifies any verified (and mentioned) falsehood as misinformation [10]. While it remains crucial for maintaining conceptual clarity, consolidating all misleading content under a single label supports the large-scale annotation and detection efforts. We adopt MuMiN's definition because it tries to provide a similar method to real-world moderation pipelines, where the primary question is simply: is this claim factually wrong? And therefore, the same question is what our paper poses to both human annotators and large language models (LLMs). This methodology forms the foundation of our dataset development [11-13]. Moreover, we do not differentiate these categories in this study due to the lack of intent annotations in the MuMiN dataset. However, our cognitive framework includes perceived intent as a feature, and future research may more clearly separate these concepts.

Spreading misinformation can result in serious risks to society [14]. For example, false narratives can manipulate voter behavior [15], increase social divisions [16,17], and spread misinformation about diseases, treatments, or vaccines that can lead to dangerous actions, as seen during the COVID-19 pandemic [2] or increase social divisions as before it. The rapid spread of misinformation, which can be intensified by algorithms that promote engaging content, poses a significant challenge for both platforms and policy regulators.

Traditional methods of content moderation relying mainly on human moderators are becoming insufficient as misinformation grows in both volume and complexity. The amount of content makes it impossible for human moderators to manually evaluate every potential instance of misinformation. Furthermore, the nature of this content can also be harmful to human moderators' well-being [18].

About LLMs

Given these challenges, approaches involving generative artificial intelligence (AI) have been explored to address the issue of misinformation detection at scale. LLMs such as GPT-4 have gained attention for their ability to generate responses to a wide range of prompts. These models can perform various language tasks, including sentiment analysis and text summarization, which has led to their application in content moderation and misinformation detection.

Nevertheless, despite their advancements, these models have limitations. Although LLMs exhibit extensive knowledge, they lack an understanding of commonsense knowledge [19]. Therefore, one of the main challenges is whether they can truly replicate the judgments of human annotators, especially in areas that require understanding context in text analysis. Human moderators can be experienced at identifying forms of manipulation, sarcasm, and irony, and understanding the social and cultural contexts in which misinformation exists. In contrast, language models may struggle with these tasks because, while they can process vast amounts of data, they lack the understanding of context and intent that human moderators have [20,21].

Additionally, these models can be biased by the data they are trained on, leading to potentially inaccurate judgments, similar to the biases that can affect human moderators. Concerns also arise regarding their ability to distinguish between fact and opinion, especially when identifying intent or evaluating source credibility. However, this challenge is not unique to LLMs. Human moderators also struggle with these complexities, as biases and subjective interpretations can influence their judgments. Therefore, understanding the limitations and potential biases of both humans and LLMs is important in developing effective misinformation detection strategies [22,23].

Moreover, detecting and analyzing misinformation is not only about identifying false statements. It also involves understanding the intent behind the content and the linguistic features that may influence how information is perceived and shared [24]. Cognitive analysis of social media users' posts, which examines the psychological and social factors affecting user behavior, is essential in this context. Traditional computational approaches have struggled to evaluate these aspects due to often focusing

XSL•FO
RenderX

on surface-level features without exploring the underlying motivations or implications. The integration of advanced language models presents an opportunity to enhance this understanding, but it also poses many challenges that need to be addressed.

While previous studies have examined the capabilities of LLMs in various aspects of misinformation detection, there is still a gap in understanding how these models perform in cognitive reasoning tasks compared to human annotators. Specifically, there is limited research on the ability of LLMs to interpret user intent, social dynamics, and other cognitive aspects inherent in misinformation. Our study aims to fill this gap by comparing the performance of LLMs and human annotators in these complex tasks, using different prompting strategies to evaluate their effectiveness.

Research Objectives and Questions

The primary objective of this study is to investigate the effectiveness of LLMs in detecting and interpreting misinformation and how their outputs compare to human annotators, with a specific focus on tasks requiring cognitive analysis and subjective judgment. By comparing the performance of these models using different prompting strategies, we aim to address the following research questions (RQs):

1. Can LLMs replicate human annotators' judgments in the context of misinformation detection? (RQ1)
2. What are the comparative strengths and weaknesses of different prompting strategies in improving LLM performance on misinformation detection tasks? (RQ2)
3. To what extent do LLM outputs align with expert panel ratings on items that relate to our cognitive framework? (RQ3)

To address RQ3, we established a cognitive framework consisting of 3 parts: content and cognitive analysis, engagement and perception, and personal perception. We build upon established cognitive and social-psychological models to establish a cognitive framework that conceptually relies on the Reasoned-Action-Approach [25].

Building on these insights, we develop a set of RQs based on a cognitive framework encompassing 3 critical elements: knowledge, intention, and action. This structure clarifies how users encounter, interpret, and ultimately propagate misinformation. Adapted from earlier theoretical models, our framework posits that a user's initial knowledge, shaped by cognitive biases and prior beliefs, guides their intention (whether to share, ignore, or challenge content), which in turn informs their subsequent action (ie, reposting, commenting, or liking). By integrating additional factors such as stance, in-group or out-group identity, and reflection, this cognitive framework provides a comprehensive perspective on how psychological motivations, social influences, and contextual features interact within online environments.

1. Content and cognitive questions support our goal of analyzing intent and cognitive areas to identify where LLMs underperform compared to human annotators. This approach aligns with content analysis methodologies [26], which are fundamental to identifying the ideological leanings or affiliations of users. For example, understanding whether a user's post aligns with specific ideologies or discusses controversial topics allows us to analyze areas of polarization and potential misinformation triggers [4,14]. Additionally, questions exploring the purpose of the user posts (such as inform, persuade, and entertain) provide insights into user motivations and how content might be received or shared [27-29].
2. Engagement and perception questions enable evaluation of virality and influence, which are critical for understanding the spread of misinformation. These included items relating to engagement patterns, such as virality potential and interaction types (likes, shares, and comments) [30]. These factors provide context for identifying why certain content has virality and how misinformation propagates [31]. Moreover, we examined how users' posts may influence audience attitudes and behavior by evaluating user stance, persuasive elements, and hidden messages [32]. We included questions exploring manipulative intent or opinion-changing attempts to investigate ethical concerns and the potential impact of misinformation [27].
3. Personal reaction questions address the subjective nature of misinformation analysis. This enables us to identify the interpretive capabilities of LLMs. Questions examining personal reactions and discrepancies between intent and language can show discrepancies and biases [12]. These insights are important for understanding audience response and identifying risks [33]. For instance, exploring whether content appears trustworthy provides context for evaluating its societal implications.

Related Work: Traditional Approaches to Misinformation Detection

Detecting misinformation in online systems has been a prevailing challenge. Early methods primarily used feature-based machine learning models relying on linguistic characteristics, sentiment analysis, user metadata, and network patterns of information spread. These supervised models required data labeled by humans and domain knowledge for training. While they could detect indicators of misinformation, they struggled to interpret context, intent, and features such as humor or sarcasm that contribute to misleading content [11,13,34].

A significant limitation of these traditional methods is their lack of generalizability across different contexts. Heavily dependent on training and specific features in classification models, they often become outdated as social media platforms and misinformation rapidly evolve. Moreover, variations in misinformation across languages and different social media platforms presented challenges, as models trained on 1 dataset may fail on others [11]. These drawbacks have led researchers to develop more robust and adaptable solutions capable of handling the dynamic nature of social media and the diverse features of misinformation.

Human Annotation and Moderation

Despite the scalability of automated systems, human expertise remains essential for complex tasks requiring an understanding of context and intent in both model training and moderation.

XSL•FO
RenderX

Human moderation to label social media's extensive content is expensive and time-consuming [18]. Even experienced annotators can disagree on classifying content when definitions of misinformation are unclear or intent is difficult to determine, raising concerns about the reliability of human-annotated datasets used as ground truth. To mitigate these concerns, methods such as consensus building among annotators, clear guidelines, and training have been implemented to improve the reliability of human-annotated datasets [26,35].

Evolution of LLMs in Misinformation Detection

The development of LLMs has significantly advanced natural language processing. Early models such as GPT-2 [36] showed that unsupervised pretraining on large text corpora, followed by task-specific fine-tuning, could enhance performance in text classification and summarization. Successors such as GPT-3 and GPT-4 [19,37] further demonstrated the potential of LLMs in misinformation detection. Models such as GPT-3 introduced the concept of few-shot learning and helped demonstrate that even a small number of examples can effectively guide large-scale model predictions [19]. However, research also shows that the ordering of prompts can substantially impact few-shot performance, which shows a particular sensitivity in LLMs [38]. These models are suited for detecting misinformation due to their ability to process extensive amounts of text and identify latent patterns. They perform well even without extensive labeled data, making them adaptable to new forms of misinformation, which is a significant advantage over their predecessors to avoid constant retraining [37].

Several studies have applied LLMs to misinformation detection. Wang [39] and Wang et al [40] developed models to detect machine-generated fake news by capturing the structure and style of manipulations within text. Other research has shown that LLMs can outperform traditional models in detecting rumors, conspiracies, and politically motivated misinformation by recognizing deeper semantic patterns and linguistic features [41]. These advancements are relevant to our study as we explore the capabilities of LLMs in detecting misinformation and understanding the underlying cognitive and linguistic features, comparing their performance to human annotators.

However, LLMs can generate hallucinations, which can be defined as possible but incorrect or misleading information [42]. Therefore, models might inadvertently reinforce the misinformation they aim to detect. Moreover, careful prompt design is required to ensure outputs align with initial requirements. Recent advances in prompting strategies, such as chain-of-thought prompting [43], have shown promise in enhancing reasoning capabilities, but their effectiveness compared to human annotators remains to be determined.

Comparative Studies of Human Annotators and Advanced Models

Research increasingly compares LLMs' performance with human annotators, particularly in tasks requiring subjective judgment and cognitive analysis. A key advantage of these models is their ability to process large datasets rapidly, analyzing thousands of social media users' posts in real time to identify misinformation and its linguistic features. However, human annotators are still better at interpreting context and other linguistic features. For example, Thomas et al [20] found that while advanced models can assist human moderators in content moderation tasks, they primarily assist rather than replace human decisions. It remains uncertain whether models can fully replicate human-level analysis in tasks requiring subjective judgment, such as exploring intent or interpreting sarcasm [22].

Some studies show that advanced models have the potential to independently perform tasks traditionally requiring human expertise. Wang et al [40] used BERT and fine-tuned RoBERTa to detect AI-generated news by GPT. The model performed comparably to humans in straightforward factual evaluations but fell short in analyzing advanced linguistic tasks, including interpreting rumors. Similarly, Thomas et al [20] conducted a study demonstrating that LLM-assisted human raters had increased accuracy in content moderation. However, the study only focused on assisting humans rather than directly comparing LLM outputs to human annotations. We interpret this as LLMs may show promise for automating large-scale misinformation detection, but their performance on tasks requiring deeper contextual understanding and subjective judgment remains unclear.

Moreover, recent work has also pointed out the application of LLMs for explainability across a range of health-related domains. One study evaluated ChatGPT's (OpenAI) reliability in self-diagnosis contexts using medical licensing-style context prompts assessed by both experts and nonexperts. The results showed that while some responses were robust to input variation, only a minority of outputs were consistently rated as factually correct, raising concerns about ambiguity and misinformation in unmoderated use cases [44]. Other research has compared LLM performance in detecting health-related rumors, showing high surface-level accuracy in binary classification tasks, but showing certain limitations in the depth and consistency of explanations, particularly across languages and cultural contexts [45].

Finally, these studies highlight both the potential and limitations of LLMs for misinformation detection, emphasizing the importance of retrieval grounding, careful prompt design, and transparent justification generation in high-stakes applications such as public health.

Methods

Ethical Considerations

Our study contributes to the ongoing discussion on the application of computational methods in content moderation. The potential of LLMs to scale misinformation detection is significant, but their deployment raises several ethical concerns. First, there is the risk that these models might reinforce existing biases, given their sensitivity to biases in training data, similar to human biases. Second, the lack of transparency in LLM decision-making processes can lead to challenges in accountability. This raises a fundamental question: how do we ensure that these systems make fair and unbiased judgments, especially when dealing with controversial content? [46].

XSL•FO
RenderX

This study received approval from the University of Bristol Faculty Ethics Committee (ref 18316).

Dataset

We used a subset of users' posts from the MuMiN dataset [10], which is a large-scale, multilingual, and multimodal dataset for misinformation detection. We selected a balanced set of users' posts containing both misinformation (88%) and factual (12%) content to ensure a diverse representation of topics.

Recruitment and Selection of Human Annotators

We recruited 25 participants (both undergraduate and postgraduate) from the University of Bristol (both undergraduate and postgraduate) through promoting our research through emails. The selection of annotators was based on responses to a screening questionnaire designed to capture participants' background and suitability for the task. The questionnaire collected information on factors including participants' experience with X (ie, the frequency of usage and time spent on this social media platform), age, educational background, political orientation, and familiarity with social media platforms. We selected participants who met specific criteria out of an initial 560 respondents, including active users with a minimum of 1 year of experience using X, a range of ages, educational backgrounds, and political affiliations to ensure diversity, varied political orientations to capture different perspectives, and demonstrated familiarity with social media content.

Each annotator evaluated 60 users' posts and provided three pieces of information for each post: (1) classification decisions on misinformation and cognitive features (5-point Likert-type questionnaire); (2) a self-reported confidence score indicating how certain they felt about their classification, using a 5-point scale (1=very uncertain to 5=very certain); and (3) open-ended justifications for their answers. The annotation questions are provided in Table S1 in Multimedia Appendix 1, and the annotation form instructions and example questions are provided in Section S1 in Multimedia Appendix 1.

Partial responses were excluded from the analysis to maintain data integrity.

LLM Specifications

Overview

We designed an ablation study in which we compared various versions of OpenAI's GPT series of models, specifically GPT-3.5, GPT-4, GPT-4o, and GPT-4o mini (Table 1) to human annotators. We also explored various prompting strategies and how these affect the performance of the models in our study. We measured performance using precision, recall, specificity, and accuracy. These models were set up to mimic the annotation form provided to human annotators.

Table 1. Summary of GPT model versions—overview of GPT model versions, ranging from GPT-3.5 to GPT-4 versions. This table indicates the model release dates and versions.

Model	Version
GPT-3.5	gpt-3.5-turbo-1106
GPT-4 Turbo	gpt-4-turbo-2024-04-09
GPT-4o	gpt-4o-2024-08-06
GPT-4o Min	gpt-4o-mini-2024-07-18

Prompting Strategies for LLMs

Recent studies show that refined prompt engineering can substantially improve model outputs [41]. We used 0-shot, few-shot, and chain-of-thought prompting techniques (described in Table 2; example prompts are provided in Table S2 in Multimedia Appendix 1) to generate responses with little or no prior information, checking how well LLMs could apply their existing knowledge to generate responses. However, other prompt strategies, such as batch prompting or a prompt pattern catalog, are not used in this paper [41,47,48].

Table 2. Summary of prompting techniques—overview of 3 primary prompting techniques (0-shot, few-shot, and chain-of-thought). This table highlights their descriptions and how each method guides an LLM'sa response generation.

Prompting technique	Description
Zero-shot prompting	Give the LLM a single instruction without any examples related to the task. This method tests the model's ability to use what it already knows to respond to new prompts, showing how well it can apply its existing knowledge to generate responses [18].
Few-shot prompting	Provide the LLM with examples related to the task before giving it the test prompt. This approach is designed to help shape the model's response by using examples to guide its reasoning. It depends on the model's capacity to adapt its answers based on the specifics provided.
Chain-of-thought prompting	This method involves creating instructions that help the LLM to think step-by-step, similar to how annotators would logically tackle these questions. The method helps the LLM to explain each part of its thinking process, improving the accuracy and relevance of the answers it provides.

[a]LLM: large language model.

XSL•FO
RenderX

Temperature Control for LLMs

To keep the results consistent and predictable, we set the temperature parameter to 0 for all tests. This helped reduce randomness in the responses that were produced, making the outputs more reliable.

Ground Truth for Misinformation Detection

In our study, we use binary MuMiN's misinformation label as the ground truth for the first stage of this study. These MuMiN labels were established by results from multiple fact-checkers, which were standardized by a fine-tuned RoBERTa model, and therefore provide a verified reference for whether a post is considered misinformation. We compared LLM and human annotator performance against this pre-existing ground truth misinformation label.

Ground Truth for Cognitive Analysis

Overview

Human annotator ratings were the reference standard against which we compared LLM performance for cognitive analysis. Given the fact that external fact-checking mechanisms do not exist for cognitive questions, and that there is inherent subjectivity and variability among human raters in such tasks, four different approaches or panel configurations were used to approximate ground truth: (1) precision-based selection, (2) conformity-based selection, (3) clustering technique, and (4) hybrid approach (described further below). Each of these was compared to the various LLM specifications to indicate model performance.

Precision-Based Selection

We selected annotators with high precision in detecting misinformation, benchmarked against the MuMiN misinformation label. Precision was measured as the proportion of correctly categorized users' posts out of the total reviewed. Therefore, we added a "misinformation label" column to our dataset. This label was determined based on annotators' scores: if an annotator scored a user's post 4 or higher on a 5-point Likert scale, the user's post was labeled as misinformation (misinformation label=1); otherwise, it was not labeled as misinformation (misinformation label=0). We then calculated the precision for each annotator, which was the proportion of users' posts they correctly identified as misinformation out of all the users' posts they reviewed. Annotators with a precision score above 65% of the posts that they reviewed were considered reliable in identifying misinformation in most cases.

Conformity-Based Selection

For this method, we assume that frequent alignment with the majority reflects accuracy. We calculated the majority response (mode) for each user's post based on all annotator responses. Therefore, we have chosen annotators who consistently aligned with the majority response for each X post (agreement >65%).

Clustering Technique

K-means clustering was used to segment annotators' responses, self-reported confidence scores, and textual justifications (vectorized using GPT 4o). We used the K-means clustering algorithm, determining the optimal number of clusters using the elbow method. Each cluster's accuracy was evaluated by comparing the proportion of correct versus incorrect misinformation identifications within the cluster. Clusters with a majority of correct identifications were labeled as high-accuracy. The elbow method was used to determine the optimal number of clusters for grouping our annotators based on their responses and characteristics. This study indicated that 3 clusters (k=3) were the most suitable choice for balancing within-cluster similarity and between-cluster differences. The accuracy of these clusters varied, reflecting the diversity in annotator reliability and perspective (Table 3).

Table 3. Cluster performance—this table presents the accuracy of 3 annotator clusters in identifying misinformation.

Cluster	Results
0	Moderate accuracy at 54.07%, indicating a mixed level of performance (no better than chance)
1	The highest accuracy at 71.49%, suggesting a strong ability to identify misinformation consistently
2	The lowest accuracy at 39.12%, which might reflect either a lack of understanding of the misinformation criteria or a differing interpretation compared to other annotators

Hybrid Approach

This analysis used the insights from the 3 previous methods to explore whether a combined approach would enhance the accuracy of the misinformation labels. By integrating the precision-based, conformity-based, and clustering results, we aimed to leverage the strengths of each to compensate for its individual limitations. Therefore, the intersection of annotators identified across the 3 methods was used to form the final panel.

The evaluation of accuracy across different panels of annotators was important in determining the effectiveness of each method for identifying misinformation. In our comparative analysis of annotator panels, we have summarized findings that show differentiated performance in Table 4. These findings indicate a combined-methods approach as the most effective strategy, and we will be using this as the main panel of (n=18) human annotators, against which we will compare LLM performance.

XSL•FO
RenderX

Table 4. Performance evaluation of different selection methods—this table displays performance metrics for 4 different selection methods of the expert panel.

Study	Description	Precision	Recall	F_1-score	Note
1	Precision-based selection	0.714	0.892	0.793	Panel 1 achieved high recall but lower precision, indicating strong detection capabilities accompanied by a higher occurrence of false positives.
2	Conformity-based selection	0.710	0.872	0.783	Panel 2 displayed reduced performance across all metrics, suggesting that conformity to majority opinions does not necessarily improve detection accuracy.
3	Clustering-based method	0.727	0.903	0.806	Panel 3 improved both precision and recall, showing the effectiveness of a profiling-based approach that leverages individual strengths.
4	Hybrid approach	0.725	0.918	0.810	Panel 4 recorded the highest recall and F_1-scores, effectively minimizing false negatives and demonstrating superior overall performance in misinformation detection.

Misinformation Detection Performance

We evaluated each LLM and human annotators on metrics including precision, recall, accuracy, and F_1-score in identifying misinformation in our MuMiN's ground truth. The results from this evaluation helped us select a final panel of human annotators, which we then used to establish the "ground truth" reference for the cognitive analyses study as described in the next section.

Comparative Analysis of LLMs

A comprehensive comparison of various LLM models to evaluate their performance across a range of prompting mechanisms was conducted. Using the 0-shot prompt approach as the baseline, the McNemar test was applied to statistically assess the performance of different prompt strategies within each model. This investigation focused specifically on the models' capabilities to handle complex cognitive questions. Performance metrics were used, such as the F_1-score, precision, and recall, to provide a detailed assessment of each model's ability to interpret the same questions as the panel annotators' baseline. Hence, the following performance metrics were used: accuracy, precision, recall, and F_1-score to evaluate both human annotators and LLMs.

- 
- 
- 
- 

Categories for Cognitive Analysis of Misinformation

Overview

We explored the capabilities of LLMs compared to human annotators using specific cognitive categories (Table 5). These categories were designed to capture various aspects of human cognitive and social dynamics that help understand and identify misinformation and intent.

We aimed to find how LLMs performed compared to human annotators in each cognitive category. This categorization allowed us to analyze the technical aspects, such as linguistic and cognitive features, which contribute to the spread and understanding of misinformation, for example, how individuals interpret misinformation, and how we can evaluate intent and social dynamics of how echo chambers could be formed [49].

Table 5. Cognitive categories for misinformation analysis—this table presents the cognitive framework used for the analysis of misinformation in this thesis. These categories enable a detailed comprehension of misinformation ranging from individual cognitive processes to social interactions.

Category	Description
Content analysis	Analyzing the factual accuracy, logical consistency, and credibility of the content itself
Perception and intent	Assessing the underlying intentions of the content creator and how the message is likely to be perceived by other users. This involves interpreting sarcasm, irony, and implicit meanings
User engagement and interaction	Analyzing how users interact with the content, including likes, shares, comments, and the potential for virality. This reflects the content's influence on different users
Personal reaction	Evaluating the emotional responses that the content may provide in individuals, such as fear, anger, empathy, or trust
Behavioral analysis	Investigating patterns in user behaviors that might indicate bot activity or organized activity
Social and group dynamics	Understanding how the content affects or reflects group behaviors, social norms, sentiments, and echo chambers or polarization

XSL•FO
RenderX

LLM Ratings

We conducted a comparative analysis of several LLMs on a multilabel classification task involving cognitive questions related to posts from the MuMiN dataset. The models were tasked with generating responses to these cognitive questions, and their responses were evaluated based on multilabel classification criteria. The models analyzed were GPT-4 Turbo, GPT-4o, GPT-4o mini, and GPT-3.5. We provided the same questions that annotators answered, and to enhance fairness. We configured all models with a temperature setting of 0. This setting minimizes randomness in the model's answer generation processes, allowing for consistent responses across multiple runs.

Confidence Scores

This section highlights the relationship between the confidence scores provided by LLMs and their accuracy. We analyzed the self-reported confidence metrics of the models. Confidence scores represent the probability assigned by the LLM to its chosen prediction, which indicates the model's confidence about the correctness of its response. While these scores are not direct measures of reliability, they provide insights into the model's self-assessment capabilities.

Our analysis examines how well confidence scores aligned with prediction accuracy across different cognitive categories. Furthermore, we calculated the Pearson correlation coefficient between the confidence levels and the correctness of the predictions for the following categories: content analysis, user engagement and interaction, personal reaction and perception, and intent recognition.

Case Study Selection and Analysis

We conducted a focused case study analysis on selected posts in order to show differences between LLMs and human annotators. We identified posts from the annotated MuMiN subset that have shown particularly interesting features such as sarcasm, subtle health misinformation, or ambiguous intent. We then tried to examine how both LLMs and human annotators labeled these posts regarding their misinformation label and the cognitive categories of perception and intent, user engagement, and personal reaction.

Results

Detecting Misinformation: Comparing LLM and Human Annotator Performance

Overview

Table 6 shows the performance metrics of different GPT models against the ground truth misinformation labels that we extracted from the MuMiN dataset. Among the models we tested, GPT-4 Turbo achieved the highest accuracy of detecting misinformation: 67.2%, with a precision of 75%, a recall of 81.8%, and an F_1-score of 78.2%. In comparison, GPT-3.5 achieved an accuracy of 63.93%, precision of 84.3%, recall of 61.3%, and an F_1-score of 71%.

Comparatively, the panel of annotators achieved an accuracy of 70.1%, precision of 72.5%, recall of 91.8%, and an F_1-score of 81%. These findings have important implications for future content moderation and cognitive understanding of LLMs. While human annotators showed slightly higher accuracy and recall, GPT-4 Turbo's comparable performance in terms of precision and F_1-score suggests it can be used as a scalable alternative for moderation, especially in areas requiring high precision, such as misinformation detection. This also points to the potential for hybrid systems, where GPT-4 handles large-scale initial sets of users' posts, and human moderators focus on supervision and validation. Such an approach could optimize resource efficiency while maintaining high-quality outputs. This result underscores the evolving role of human moderators in complementing AI systems and calls for further exploration into collaborative frameworks that maximize the strengths of both.

Table 6. Performance metrics of different GPT models against MuMiN's ground truth—this table shows the accuracy, precision, recall, and F1-scores for 4 LLMa-based models and a panel of human annotators (expert panel) when compared against MuMiN's ground truth.

Model	Accuracy (%)	Precision (%)	Recall (%)	F_1-score (%)
GPT-3.5	63.93	84.38	61.36	71.05
GPT-4 Turbo	67.21	75	81.82	78.26
GPT-4o Mini	60.66	81.25	59.09	68.42
GPT-4o	62.3	76.92	68.18	72.29
Panel of annotators	70.1	72.5	91.8	81

[a]LLM: large language model.

Analysis of Prompting Strategies

We evaluated the effect of different prompting strategies (0-shot, few-shot, and chain-of-thought) on the performance of the GPT models. Table 7 presents the performance metrics for GPT-4 Turbo under these prompting conditions.

The chain-of-thought prompting strategy yielded the best performance, with an accuracy of 72.1%, precision of 78%, recall of 85%, and an F_1-score of 81%. Such findings suggest that providing the model with a structured reasoning path enhances its ability to interpret complex cognitive questions.

We applied the McNemar test to explore the statistical significance of the differences between prompting strategies. The test results indicated that the performance improvement of chain-of-thought prompting over 0-shot prompting was statistically significant (P<.05).

XSL•FO
RenderX

Table 7. Performance of GPT-4 Turbo for different prompting strategies—this table presents GPT-4 Turbo's performance under 3 prompting strategies (0-shot, few-shot, and chain-of-thought) across accuracy, precision, recall, and F1-score.

Prompting strategy	Accuracy (%)	Precision (%)	Recall (%)	F_1-score (%)
Zero-shot	65.57	73	79	75.9
Few-shot	68.85	76.5	82	79.13
Chain-of-thought	72.13	78	85	81.33

Confidence Scores of LLMs and Accuracy

Our findings illustrate a positive correlation between confidence scores and prediction accuracy in the areas of content analysis (ρ=0.72, r<0.01) and user engagement and interaction (ρ=0.65, r<0.05). These results suggest that in straightforward tasks with simpler linguistic patterns, higher confidence scores correspond to greater accuracy. For example, they were both confident and correct in cases where the models analyzed clear misinformation statements.

In contrast, the correlation was weaker in the area of personal reaction and perception (ρ=0.38, r=0.12) and intent recognition (ρ=0.33, r=0.15). These categories involve subjective and more cognitive interpretation, such as understanding sarcasm or underlying intent, where the models expressed overconfidence in incorrect predictions. This difference indicates that while confidence scores can be used for reliability in certain contexts, they fail to benefit in more complex analyses.

Cognitive Categories

Overview

We further analyzed the models' performance across different cognitive science categories. Table 8 compares the best GPT-4 Turbo and worst-performing LLM GPT-3.5 in terms of accuracy, precision, recall, and F_1-score for each category. Overall, GPT-4 Turbo outperformed GPT-3.5 across all categories, particularly in user engagement and interaction and social and group dynamics. However, performance in categories including perception and intent and personal reaction remained low for both models, indicating challenges in capturing human interpretations in these areas.

Table 8. Comparison of GPT-4 Turbo versus GPT-3.5 for cognitive categories—this compares GPT-4 Turbo and GPT-3.5 across 6 cognitive categories, showing higher accuracy, precision, recall, and F1-scores for GPT-4 Turbo in every category. Overall, GPT-4 Turbo consistently outperforms GPT-3.5 based on these metrics.

Category and model	Accuracy (%)	Precision (%)	Recall (%)	F_1-score (%)
Content analysis				
GPT-4 Turbo	41.4	43.3	41.4	42
GPT-3.5	20.9	46	20.9	22.6
Perception and intent				
GPT-4 Turbo	23.8	47.7	23.8	24.7
GPT-3.5	11.1	36.2	11.1	10.5
User engagement and interaction				
GPT-4 Turbo	63.4	69.6	63.4	66
GPT-3.5	58.7	65.4	58.7	61.8
Personal reaction				
GPT-4 Turbo	24	42	24	30
GPT-3.5	19.7	41.6	19.7	26.6
Behavioral analysis				
GPT-4 Turbo	31.1	26.3	31.1	28.5
GPT-3.5	21.9	25.6	21.9	23.6
Social and group dynamics				
GPT-4 Turbo	45.9	67.4	45.9	54.9
GPT-3.5	13.9	36.4	13.9	20

Comparison With Human Annotators

We compared each LLM's outputs to the labels from our final panel of 18 annotators for each cognitive category.

Our results show that human annotators consistently outperformed the models in tasks requiring deeper contextual understanding and subjective judgment. For instance, in perception and intent, the panel consensus achieved an accuracy

XSL•FO
RenderX

of approximately 60% when compared within themselves (eg, comparing each annotator's label to the vote within the final panel), whereas GPT-4 Turbo's accuracy measured against that same panel was only 23.8%. This gap highlights the complexity of interpreting context, sarcasm, or hidden intent.

Importantly, "accuracy" for human annotators in these cognitive labels was based on how often each individual aligned with the final panel. This ensures consistency in evaluating both LLM and human performance, though we acknowledge that the final panel's answers themselves may still contain subjectivity.

Discussion

Principal Results

Based on the statistical significance established in our analysis, we explore the implications of these findings in the following discussion. As intent remains difficult to establish in online content, both human and LLM evaluations must be treated as interpretive rather than conclusive. Detecting misinformation on X is very challenging due to complex content, social interactions, and huge amounts of data. We conducted a comparative study on how well LLMs and human annotators detect misinformation, focusing on how they understand behavior and context. While our cognitive framework aims to explore how LLMs and humans interpret features such as intent and perception, we acknowledge the limits of establishing these features only from textual content. Our findings should be interpreted as an investigation and comparison with a panel of experts. These limitations reflect broader challenges also faced by human moderators and platforms when interpreting user intent under content moderation policies.

A hybrid approach, where computational models work alongside human moderators, may offer the most effective solution. Such a system could improve the scalability of content moderation while relying on human expertise for more complex cases that require ethical consideration [11]. Our results reveal important insights into what LLMs can and cannot do, suggesting ways to combine machine learning models with human judgment in content moderation systems.

Comparing Performance Across Different LLMs

Our analysis revealed that the chosen model of LLM significantly impacted performance in misinformation detection tasks. GPT-4 Turbo outperformed chosen models across all metrics, demonstrating higher accuracy, precision, recall, and F_1-score. For instance, GPT-4 Turbo achieved an accuracy of 67.21%, while GPT-4o and GPT-4o mini achieved 62.30% and 60.66%, respectively. This trend indicates that larger models have enhanced capabilities in pattern recognition and contextual understanding.

Moreover, recent benchmark studies have raised important questions about the reasoning capabilities of LLMs. Shojaee et al [50] observed a decline in model performance on structured reasoning tasks at higher levels of complexity, which was interpreted as a sign of cognitive limitation. However, further research by Lawsen [51] and Varela et al [52] suggested that such failures may instead reflect constraints in experimental design, such as unsolvable task configurations or token-length limitations, rather than fundamental reasoning deficits. These findings echo patterns in our own study, where LLMs performed well on certain structured misinformation tasks but diverged from panel annotators on more interpretive items, such as user intent and sarcasm. Importantly, these areas are challenging not only for models but also for human raters, as judgments about meaning, motive, or tone can be inherently subjective. As such, these disagreements may not reflect clear errors but rather underline the difficulty of establishing a consistent ground truth for cognitive analysis. This highlights the need for further research into how both humans and models interpret subtle linguistic cues, and how such interpretations can be evaluated more reliably.

Strengths and Limitations of LLMs

Performance of Chain-of-Thought Prompting

Chain-of-thought prompting generally performed better than other prompting methods, including 0-shot and few-shot across various cognitive areas, but its success varied based on task complexity.

In areas that rely on clear pattern recognition, such as analyzing user engagement and interactions, chain-of-thought prompting has significantly improved both accuracy and recall. This aligns with researchers who suggest treating LLMs as "program executors" where step-by-step reasoning helps reduce errors [53]. By thinking through each step, LLMs were better at identifying strategies that kept users engaged, leading to more precise results. However, chain-of-thought can also lead to "greedy reasoning," where the model is based on incomplete logic [54].

Yet, in more complicated areas, including understanding perceptions and intentions, the advantages of chain-of-thought prompting were not as strong. Although there was still some improvement, compared to other prompting strategies, the LLMs did not perform as well as they did in the simpler cognitive categories. This indicates that while breaking down the reasoning process helped the LLM organize its tasks to do and improve its reasoning, it does not completely overcome the LLM's challenges in understanding deeper behavioral questions.

Future studies might look into more advanced prompting methods or combine chain-of-thought with other contextual information (such as prior X posts or followers of each user) to further improve performance in these complex areas.

To conclude, our results indicate that when using chain-of-thought prompting strategies, GPT-4 Turbo nearly matches human performance in several tasks, especially those involving logical reasoning. For example, the model was good at detecting clearly false statements and identifying patterns of how misinformation spreads, using reasoning similar to that of human annotators. However, the models have limitations, especially in areas that require a deeper understanding of behavior and context.

XSL•FO
RenderX

Human Annotators' Strengths: Understanding and Judgment

Human annotators consistently outperformed the models in tasks requiring deep contextual understanding and subjective reasoning. They were better at interpreting aspects of user intent and social dynamics, especially when misinformation was contextual or was written to sound truthful. This supports previous research that emphasizes the importance of human insight in interpreting context, behavior, and intent in social media content [6,18,20].

For example, when users discussed misinformation about diet strategies, human annotators were more skilled at determining the intent behind the message, while models often misread the content as factual or harmless. This highlights the limitations of language models in tasks needing more than basic analysis tasks, where social, emotional, and cultural knowledge are crucial for accurate content classification.

GPT-4 Turbo found tasks involving analyzing intent and linguistic features difficult, even with advanced prompting strategies. For instance, GPT-4 Turbo struggled to detect misinformation in a user's post that used satire to critique health misinformation, misclassifying it as genuine content. This aligns with previous research showing that while language models handle surface-level analysis well, they often struggle with tasks needing deeper thinking or social understanding. For instance, the models were less effective at identifying misinformation when the intent behind the message was unclear or when context significantly changed the meaning.

Comparing Performance Across Cognitive Categories

Our analysis, Table 8, of performance across specific cognitive categories revealed that models such as GPT-4 Turbo performed similarly to human annotators in straightforward categories such as user engagement and interaction and social and group dynamics, where finding obvious patterns was sufficient. However, GPT-4 Turbo's performance dropped significantly in categories requiring interpretation of perception and intent and personal reaction, where human emotional judgment and cultural awareness were important.

For instance, in the user engagement and interaction category, the models effectively identified users' posts designed to provoke engagement through provocative or polarizing content, as shown by their relatively high recall scores. However, GPT-4 Turbo struggled in the perception and intent category, which required understanding the underlying intentions of the user who authored the post. Human annotators were better at recognizing when a user's post was intended to mislead, even if the content was not obviously false.

Limitations of LLMs in Contextual Interpretation

We found that LLMs often struggle when it comes to accurately categorizing content related to perception and intent or personal reactions (Table 8).

These areas require a deep understanding of complex human emotions, hidden intentions, and the contexts that are not always clearly stated in the text. For example, telling the difference between someone genuinely expressing concern and making a sarcastic comment involves picking up on tone and underlying motives, which goes beyond just processing the words.

Research has consistently shown that these models have difficulty with tasks that require a theory of mind or cognitive reasoning [55]. While they are good at recognizing and generating text based on patterns they have learned, they do not possess the natural cognitive abilities that humans use to accurately infer intentions and emotional states.

Additionally, LLMs often miss cultural and contextual subtleties, which makes their performance even worse in these categories. For instance, sarcasm that depends on context or idioms specific to a culture can easily be misunderstood, leading to less accurate and reliable classifications. Recent studies support this, highlighting that models frequently overlook the deeper meanings and practical aspects of human communication. Understanding these challenges is crucial for improving how LLMs analyze cognitive tasks. It shows the importance of incorporating more advanced learning methods, such as using multiple types of data and better context-aware techniques, to create a richer understanding of complex human reasoning.

Finally, we conducted a manual review of LLM outputs as part of our evaluation to analyze both the Likert-scale classifications and the accompanying open-ended justifications. This hallucination audit aimed to identify cases where the model's explanations introduced information not present in the original user post, a phenomenon known as hallucination. While hallucinations were more evident in the open-ended responses, where the LLM occasionally inferred motivations or context beyond the given text, they may also influence binary or scaled answers, particularly when these judgments rely on false information. These examples suggest that hallucinations may subtly shape LLM classifications. This limitation offers a possible explanation for divergences between LLM and human annotator decisions and underscores the need for caution when interpreting LLM outputs in high-stakes settings such as misinformation detection.

Case Studies

Overview

We present 3 detailed case studies involving users' posts where all LLMs underperformed in order to show the limitations of LLMs compared to human annotators in detecting misinformation (Table S3 in Multimedia Appendix 1). Our analysis of human annotators' and LLMs is based on the provided justification by them.

Example 1: Political Sarcasm

Human annotators' justifications indicated that they interpreted the selected user post as discussing misinformation, primarily due to its reference to unverified and potentially false allegations concerning a political figure. Several annotators noted a sarcastic tone and perceived the content as suggestive of corruption, which they considered potentially misleading in the absence of credible evidence. This interpretation reflects the broader patterns observed in the Results section, where human annotators showed some sensitivity to rhetorical features such as sarcasm and indirect implication.

XSL•FO
RenderX

In contrast, the LLM did not flag the post as misinformation (example 1). Instead, it appeared to interpret the content as a general political opinion or criticism, without identifying the underlying tone or indirect implications. This may illustrate some of the model's limitations in capturing specific language features, including sarcasm or rhetorical framing. While the model's output followed a coherent line of reasoning based on the surface content, it may have lacked the contextual depth to infer possible intent.

This example suggests that LLMs may face challenges in identifying indirect or implied misinformation, particularly in politically charged discourse where tone, context, and social knowledge are key to interpretation. It also points to the continuing value of human moderation in complex analytical tasks, especially where judgments about user intent or potential societal impact are involved.

Example 2: Health Misrepresentation

Human annotators tended to classify this post as discussing misinformation, primarily due to its promotion of unverified and potentially unsafe alternative treatments for serious medical conditions. Several annotators noted that suggesting raw vegetable juice as a cure lacked scientific grounding and could, in their view, discourage individuals from seeking conventional medical care. Their responses reflected concern about the potential implications of such claims in public health contexts.

In comparison, the LLM did not identify the post as misinformation. It appeared to interpret the content as a personal narrative or opinion without further analyzing further public health implications of the treatment being described. This may suggest a limitation in the model's ability to evaluate persuasive intent or emotionally framed content.

As shown in example 2, this case illustrates some of the challenges LLMs may face when processing health-related narratives that blur the line between personal testimony and misinformation. The model's output, while internally coherent, did not reflect the same level of caution observed in human annotator responses. This difference underscores the importance of continued human involvement in moderation processes, particularly in domains where the potential for harm is closely tied to misinterpretation of health advice.

Example 3: Health

In example 3, human annotators generally classified this post as discussing misinformation or potentially false claims related to pregnancy. Some annotators expressed concern that the information shared in the post could cause confusion or cause concern among users, particularly those seeking guidance on maternal health. Their interpretations reflected a cautious approach to unverified claims in sensitive health contexts.

The LLM, however, did not flag the post as misinformation. It appeared to interpret the content as an expression of personal surprise or curiosity, without further evaluating the credibility of the underlying claim. This may indicate a limitation in the model's ability to assess the implications of health-related statements that are framed informally or anecdotally.

Therefore, LLMs could struggle to identify potentially misleading health content when it is hidden through indirect or emotionally neutral language, based on this example. While human annotators tended to consider the broader context and possible impact on readers, the model focused more narrowly on surface features of the text. Our results imply that additional safeguards or contextual prompts may be needed when using LLMs to evaluate health-related social media content.

Relationship Between Accuracy and Confidence of the LLMs

We also looked into how confident the models were about their predictions and whether that confidence matched their actual accuracy. By checking the confidence scores the LLMs provided, we wanted to see if higher confidence meant more accurate results, which would help us understand how much we can trust the models' self-assessments.

This indicates there is a positive link between confidence and accuracy in most areas, especially in content analysis and user engagement and interaction. In these fields, the models were more likely to be correct when they were more confident in their answers.

Nevertheless, this relationship was not as strong in areas such as personal reaction and perception, and intent. In these more complex categories, the models sometimes felt very confident but were still wrong. This suggests that while confidence scores can be a good sign of reliability in some situations, they are not always consistent across all types of cognitive analysis.

These findings highlight the need to develop better ways to measure confidence that take into account the complexity and subjective nature of certain behaviors. Improving these confidence metrics could make LLMs more trustworthy overall.

Implications for Misinformation Detection Systems

These findings have important implications for developing and using automated misinformation detection systems. However, misinformation varies in severity from harmful health claims (ie, drinking bleach to cure COVID-19) to political satire. Our findings suggest LLMs perform better in detecting clear factual inaccuracies, especially when tied to immediate public health risk. Conversely, failures in interpreting sarcasm in political discourse may represent lower immediate harm. These should inform platforms' moderation pipelines, which can deploy LLMs as a first filter for high-risk content, while reserving complex social interpretations for human moderators.

Given limitations as discussed, we suggest that hybrid systems, combining the scalability of language models with the understanding of human annotators, may offer a strong solution for misinformation detection on online platforms.

The models could serve as the first line of defense, flagging potential misinformation based on language patterns, while human reviewers focus on more challenging cases requiring deep contextual understanding and ethical judgment. This approach would improve the accuracy of misinformation detection and ensure that it remains flexible to the diverse and evolving nature of online content. Additionally, future work should consist of fine-tuning LLMs using an extensive dataset

XSL•FO
RenderX

that includes labels of cognitive linguistic features. This task could help LLMs improve their performance and focus on these details.

Moreover, prompting strategies such as chain-of-thought have shown promise in narrowing the gap between model and human performance, suggesting that refining these techniques could lead to better results. Incorporating such strategies into misinformation detection pipelines could improve the interpretative abilities of language models, helping them more closely mimic human reasoning in complex cases.

Strengths and Limitations of This Study

One of the key strengths of this research is its multimethod approach to determining ground truth for both misinformation detection and cognitive categories. The MuMiN dataset provides fact-checking labels for misinformation, and a 3-part method (precision-based selection, conformity-based selection, and clustering) offers a way to identify a reliable panel for the more subjective cognitive labels. This process increases confidence in this study's findings about LLMs' capabilities compared with human judgment.

However, this study has some limitations. First, depending on human annotators to assign cognitive labels can introduce subjectivity, even with efforts such as building panels of experts. The final panel may not be a perfect source of truth, since humans can still make mistakes and show bias based on their beliefs and background. Second, the MuMiN dataset might not represent every social media environment, which limits how broadly the results can be applied. Third, real-world examples can involve complex issues such as sarcasm or subtle health claims that may require cultural knowledge.

Conclusions

Our evaluation of LLMs to analyze their analysis of misinformation compared to human annotators' outputs revealed insights into their abilities to interpret behavior and context on social media users' posts.

We examined the impact of different prompting strategies (including 0-shot, few-shot, and chain-of-thought prompting) on model performance across various cognitive categories. Our findings indicate that while LLMs such as GPT-4 Turbo show promising capabilities and can approach human performance in tasks involving logical reasoning and pattern recognition, they are still behind in areas requiring a deep understanding of human behavior, context, and cultural understanding. However, chain-of-thought prompting improves model performance by guiding models through reasoning processes, suggesting that such strategies are beneficial. By conducting this comprehensive evaluation, our study significantly contributes to the understanding of LLM capabilities in the area of misinformation detection and cognitive analysis. These insights highlight the potential of LLMs as tools to support human moderators but also show the need for continued human involvement in misinformation detection.

Future Work

Future work could focus on fine-tuning LLMs with datasets that emphasize cognitive and emotional linguistic features to enhance their understanding of complex human behaviors. Exploring new prompting strategies, such as integrating chain-of-thought with additional contextual features of multimodal data, may further improve model performance. Finally, users' engagement with social media posts may be influenced by both textual and visual elements, making misinformation detection more complex. Additionally, integrating ethical considerations into automated systems will be crucial for developing responsible and effective misinformation detection tools.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data on annotations, prompts, and malinformation.
[DOCX File , 170 KB - infodemiology_v6i1e72524_app1.docx]

References

1. Phillips T. Truth: A Brief History of Total Bullsh*t. Mexico: Hanover Square Pr; 2020.
2. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
3. van Doorn M. Advancing the debate on the consequences of misinformation: clarifying why it's not (just) about false beliefs. Inquiry 2023:1-27. [doi: 10.1080/0020174x.2023.2289137]
4. Garimella K, Morales G, Gionis A, Mathioudakis M. Political discourse on social mediacho chambers, gatekeepers, and the price of bipartisanship. 2018 Presented at: Proc World Wide Web Conf; April 23-27, 2018; Lyon, France p. 913-922. [doi: 10.1145/3178876.3186139]
5. Törnberg P. Echo chambers and viral misinformation: modeling fake news as complex contagion. PLOS One 2018;13(9):e0203958 [FREE Full text] [doi: 10.1371/journal.pone.0203958] [Medline: 30235239]
6. Metzger M, Flanagin A, Medders R. Social and heuristic approaches to credibility evaluation online. J Commun 2010;60(3):439. [doi: 10.1111/j.1460-2466.2010.01488.x]

XSL•FO
RenderX

7. Santos-d'Amorim K, de Oliveira Miranda MKF. Misinformation, disinformation, and malinformation: clarifying the definitions and examples in disinfodemic times. Encontros Bibli Revi Eletrônica Biblioteconomia Ciênc Inf 2021:26.
8. Hussain M, Soomro T. Social media: an exploratory study of information, misinformation, disinformation, and malinformation. Appl Comput Syst 2023;28(1):13-20. [doi: 10.2478/acss-2023-0002]
9. Wardle C, Derakhshan H. Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking. Strasbourg, France: Council of Europe Strasbourg; 2017.
10. Nielsen D, McConville R. MuMiN: A large-scale multilingual multimodal fact-checked misinformation social network dataset. : ACM; 2022 Presented at: SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 11-15, 2022; Madrid, Spain p. 3141-3153. [doi: 10.1145/3477495.3531744]
11. Oshikawa R, Qian J, Wang W. A survey on natural language processing for fake news detection. : European Language Resources Association; 2020 Presented at: Proceedings of the Twelfth Language Resources and Evaluation Conference; August 07, 2025; Marseille, France p. 6086-6093.
12. Altay S, Acerbi A. People believe misinformation is a threat because they assume others are gullible. New Media Soc 2023;26(11):6440-6461. [doi: 10.1177/14614448231153379]
13. Shu K, Sliva A, Wang S, Tang J, Liu H. Fake news detection on social media. SIGKDD Explor. Newsl 2017;19(1):22-36. [doi: 10.1145/3137597.3137600]
14. Zhang X. Will political disagreement silence political expression? The role of information repertoire filtration and discussion network heterogeneity. Hum Commun Res 2023;49(2):139-148. [doi: 10.1093/hcr/hqad009]
15. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. Science 2019;363(6425):374-378. [doi: 10.1126/science.aau2706] [Medline: 30679368]
16. Ausat AMA. The role of social media in shaping public opinion and its influence on economic decisions. Technol Soc Perspect TACIT 2023;1(1):35-44. [doi: 10.61100/tacit.v1i1.37]
17. Arif A, Stewart LG, Starbird K. Acting the part: examining information operations within #BlackLivesMatter discourse. Proc ACM Hum-Comput Interact 2018;2(CSCW):1-27. [doi: 10.1145/3274289]
18. Steiger M, Bharucha T, Venkatagiri S, Riedl M, Lease M. The psychological well-being of content moderators: The emotional labor of commercial moderation and avenues for improving support. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 1-14. [doi: 10.1145/3411764.3445092]
19. Brown T, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. 2020 Presented at: NIPS'20: 34th International Conference on Neural Information Processing Systems; December 6-12, 2020; Vancouver, BC, Canada p. 1877-1901.
20. Thomas K, Kelley P, Tao D, Meiklejohn S, Vallis O, Tan S, et al. Supporting human raters with the detection of harmful content using large language models. 2024 Presented at: IEEE Symposium on Security and Privacy (SP); May 12-15, 2025; San Francisco, California. [doi: 10.1109/sp61157.2025.00082]
21. Ziegler D, Stiennon N, Wu J, Brown T, Radford A, Amodei D, et al. Fine-tuning language models from human preferences. arXiv Preprint posted online on January 8, 2020. [doi: 10.48550/arXiv.1909.08593]
22. Cho H, Liu S, Shi T, Jain D, Rizk B, Huang Y. Can language model moderators improve the health of online discourse? 2024 Presented at: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers); June 16–21, 2024; Mexico City, Mexico p. 7478-7496. [doi: 10.18653/v1/2024.naacl-long.415]
23. Ma H, Zhang C, Fu H, Zhao P, Wu B. Adapting large language models for content moderation: pitfalls in data engineering and supervised fine-tuning. arXiv Preprint posted online on March 7, 2024. [doi: 10.48550/arXiv.2310.03400]
24. Wildner-Bassett ME, Tannen D. Framing in Discourse. England: Oxford University Press; 1995:293.
25. Ajzen I. Martin Fishbein's legacy: the reasoned action approach. Ann Am Acad Pol Soc Sc 2012;640(1):11-27.
26. Krippendorff K. Content Analysis: An Introduction to its Methodology. Los Angeles: Sage Publications; 2018.
27. Petty R, Cacioppo J. Communication and Persuasion: Central and Peripheral Routes to Attitude Change. Berlin: Springer-Verlag; 1986.
28. Lovejoy K, Saxton G. Information, community, and action: how nonprofit organisations use social media. J Comput-Mediat Commun 2012;17(3):337-353. [doi: 10.1111/j.1083-6101.2012.01576.x]
29. Marwick AE, boyd D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media Soc 2010;13(1):114-133. [doi: 10.1177/1461444810365313]
30. Goel S, Anderson A, Hofman J, Watts DJ. The structural virality of online diffusion. Manage Sci 2016;62(1):180-196. [doi: 10.1287/mnsc.2015.2158]
31. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
32. Cialdini RB. Influence: Science and Practice. Boston, Massachusetts: Allyn & Bacon; 2001.
33. Stubenvoll M, Heiss R, Matthes J. Media trust under threat: antecedents and consequences of misinformation perceptions on social media. Int J Commun 2021;15:22.

XSL•FO
RenderX

34. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R. Automatic detection of fake news. 2018 Presented at: Proceedings of the 27th International Conference on Computational Linguistics; August 20-26, 2018; Santa Fe, New Mexico, USA p. 3391-3401.
35. Dumitrache A, Inel O, Timmermans B, Ortiz C, Sips R, Aroyo L, et al. Empirical methodology for crowdsourcing ground truth. SW 2021;12(3):403-421. [doi: 10.3233/sw-200415]
36. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. United States: OpenAI; 2019.
37. GPT-4 technical report. OpenAI. 2023. URL: https://openai.com/index/gpt-4-research/ [accessed 2025-08-07]
38. Lu Y, Bartolo M, Moore A, Riedel S, Stenetorp P. Fantastically ordered prompts and where to find them: overcoming few-shot prompt order sensitivity. : Association for Computational Linguistics; 2021 Presented at: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); May 22-27, 2022; Dublin, Ireland p. 8086-8098. [doi: 10.18653/v1/2022.acl-long.556]
39. Wang WY. LIAR, liar pants on fire: A new benchmark dataset for fake news detection. : Association for Computational Linguistics; 2017 Presented at: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers); July 30, 2017-August 4, 2017; Vancouver, Canada p. 422-426. [doi: 10.18653/v1/p17-2067]
40. Wang Z, Cheng J, Cui C, Yu C. Implementing BERT and fine-tuned RoBERTa to detect AI generated news by ChatGPT. arXiv Preprint posted online on Jun 9, 2023. [doi: 10.48550/arXiv.2306.07401]
41. Chen L, Zaharia M, Zou J. FrugalGPT: how to use large language models while reducing cost and improving performance. arXiv Preprint posted online on May 9, 2023. [doi: 10.48550/arXiv.2305.05176]
42. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. ACM Comput. Surv 2023;55(12):1-38. [doi: 10.1145/3571730]
43. Wei J, Wang X, Schuurmans D. Chain-of-thought prompting elicits reasoning in large language models. arXiv Preprint posted online on January 10, 2023. [doi: 10.48550/arXiv.2201.11903]
44. Zada T, Tam N, Barnard F, Van Sittert M, Bhat V, Rambhatla S. Medical misinformation in AI-assisted self-diagnosis: development of a method (EvalPrompt) for analyzing large language models. JMIR Form Res 2025;9:e66207 [FREE Full text] [doi: 10.2196/66207] [Medline: 40063849]
45. Luo Y, Miao Y, Zhao Y, Li J, Chen Y, Yue Y, et al. Comparing the accuracy of two generated large language models in identifying health-related rumors or misconceptions and the applicability in health science popularization: proof-of-concept study. JMIR Form Res 2024;8:e63188 [FREE Full text] [doi: 10.2196/63188] [Medline: 39622076]
46. Li Y, Xie Y. Is a picture worth a thousand words? An empirical study of image content and social media engagement. J Mark Res 2019;57(1):1-19. [doi: 10.1177/0022243719881113]
47. Cheng Z, Kasai J, Yu T. Batch prompting: efficient inference with large language model APIs. 2023 Presented at: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track; December 6-10, 2023; Resorts World Convention Centre p. 792-810. [doi: 10.18653/v1/2023.emnlp-industry.74]
48. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, et al. A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv Preprint posted online on February 21, 2023. [doi: 10.48550/arXiv.2302.11382]
49. Tajfel H, Turner J. The social identity theory of intergroup behaviour. In: Political Psychology. London, England, United Kingdom: Psychology Press; 1986:7-24.
50. Shojaee P, Mirzadeh I, Alizadeh K, Horton M, Bengio S, Farajtabar M. The illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity. Machine Learning Research. 2025. URL: https://machinelearning.apple.com/research/illusion-of-thinking [accessed 2025-08-09]
51. Lawsen A. Comment on the illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity. arXiv [FREE Full text] Preprint posted online on June 16, 2025 [doi: 10.48550/arXiv.2506.09250]
52. Varela I, Romero-Sorozabal P, Rocon E, Cebrian M. Rethinking the illusion of thinking. arXiv [FREE Full text] Preprint posted online on July 1, 2025 [doi: 10.48550/arXiv.2507.01231]
53. Pi X, Liu Q, Chen B, Ziyadi M, Lin Z, Fu Q, et al. Reasoning like program executors. 2022 Presented at: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; December 7-11, 2022; Abu Dhabi. [doi: 10.18653/v1/2022.emnlp-main.48]
54. Saparov A, He H. Language models are greedy reasoners: a systematic formal analysis of chain-of-thought. arXiv Preprint posted online on March 2, 2023. [doi: 10.48550/arXiv.2210.01240]
55. Wang X, Li X, Yin Z, Wu Y, Liu J. Emotional intelligence of large language models. J Pac Rim Psychol 2023;17. [doi: 10.1177/18344909231213958]

Abbreviations

AI: artificial intelligence
LLM: large language model
RQ: research question

XSL•FO
RenderX

Edited by M Haupt; submitted 11.Feb.2025; peer-reviewed by W Ahmed, C Wardle; comments to author 25.Jun.2025; revised version received 20.Jul.2025; accepted 29.Jul.2025; published 18.May.2026.

Please cite as:
Wojtczak DN, McQuire C, Zuccolo L, Peersman C, McConville R
Performance of Large Language Models in the Cognitive Analysis of Misinformation: Evaluation Study
JMIR Infodemiology 2026;6:e72524
URL: https://infodemiology.jmir.org/2026/1/e72524
doi:10.2196/72524
PMID:42149639

©Dominika Nadia Wojtczak, Cheryl McQuire, Luisa Zuccolo, Claudia Peersman, Ryan McConville. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 18.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Quality, Reliability, and Dissemination of In Vitro Fertilization–Related Videos on Chinese Social Media: Cross-Sectional Analysis of 300 Short Videos

Xueyan Bai[1*], MD; Feng Guo[2*], BA; Dapeng Chu[1*], PhD

[1]Beijing Chao-Yang Hospital, Capital Medical University, Main Campus, Beijing Chao-Yang Hospital, 8 Gongren Tiyuchang Nanlu, Chaoyang District, Beijing, 100020, Beijing, China

[2]APUCH Innovation, Beijing, China

[*]all authors contributed equally

Corresponding Author:
Dapeng Chu, PhD
Beijing Chao-Yang Hospital, Capital Medical University, Main Campus, Beijing Chao-Yang Hospital, 8 Gongren Tiyuchang Nanlu, Chaoyang District, Beijing, 100020, Beijing, China

Related Article:

Correction of: https://infodemiology.jmir.org/2026/1/e83900

(JMIR Infodemiology 2026;6:e95354) doi:10.2196/95354

In "Quality, Reliability, and Dissemination of In Vitro Fertilization–Related Videos on Chinese Social Media: Cross-Sectional Analysis of 300 Short Videos" [1], the authors made a change to the order of authorship.

The authorship order was originally presented as:

Dapeng Chu, Xueyan Bai, Feng Guo

This has been changed to the following:

Xueyan Bai, Feng Guo, Dapeng Chu

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. Bai X, Guo F, Chu D. Quality, reliability, and dissemination of in vitro fertilization–related videos on Chinese social media: cross-sectional analysis of 300 short videos. JMIR Infodemiology 2026 Jan 28;6:e83900. [doi: 10.2196/83900] [Medline: 41604533]

Submitted 15.Mar.2026; this is a non–peer-reviewed article;accepted 27.Mar.2026; published 22.Apr.2026.

Please cite as:
Bai X, Guo F, Chu D
Correction: Quality, Reliability, and Dissemination of In Vitro Fertilization–Related Videos on Chinese Social Media: Cross-Sectional Analysis of 300 Short Videos
JMIR Infodemiology 2026;6:e95354
URL: https://infodemiology.jmir.org/2026/1/e95354
doi:10.2196/95354

© Xueyan Bai, Feng Guo, Dapeng Chu. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 22.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Japanese-Language AI Agent System for Human Papillomavirus Vaccine Infoveillance and Public Communication: Development and Feasibility Evaluation

Junyu Liu[1], ME; Siwen Yang[2], BEng; Dexiu Ma[3], MS; Qian Niu[4], PhD; Zequn Zhang[5], MS; Momoko Nagai-Tanima[1], PhD; Tomoki Aoyama[1], MD, PhD

[1]Kyoto University, Graduate School of Medicine, Yoshida-honmachi, Sakyo-ku, Kyoto, Japan

[2]David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada

[3]Department of Computer Science, Whitacre College of Engineering, Texas Tech University, Lubbock, TX, United States

[4]Graduate School of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

[5]Department of EEIS, University of Science and Technology of China, Hefei, Anhui, China

Corresponding Author:
Tomoki Aoyama, MD, PhD
Kyoto University, Graduate School of Medicine, Yoshida-honmachi, Sakyo-ku, Kyoto, Japan

Related Article:

Correction of: https://infodemiology.jmir.org/2026/1/e90295

(JMIR Infodemiology 2026;6:e109982) doi:10.2196/109982

In "Japanese-Language AI Agent System for Human Papillomavirus Vaccine Infoveillance and Public Communication: Development and Feasibility Evaluation" [1], we identified several items requiring correction. They are set out below in the order in which they appear in the article, providing the original published text and the corrected text. None of the corrections change any finding or conclusion of the article.

First, in the author byline, the degree designations of three authors were incorrect. The byline originally read:

> *Junyu Liu, PhD; Siwen Yang, BEng; Dexiu Ma, PhD; Qian Niu, PhD; Zequn Zhang, PhD; Momoko Nagai-Tanima, PhD; Tomoki Aoyama, MD, PhD*

The corrected byline now reads:

> *Junyu Liu, ME; Siwen Yang, BEng; Dexiu Ma, MS; Qian Niu, PhD; Zequn Zhang, MS; Momoko Nagai-Tanima, PhD; Tomoki Aoyama, MD, PhD*

Second, in the section Report Generation Evaluation Methodology in the Methods section, the following text was corrected:

> *Both protocols use LLM-based judges with standardized scoring rubrics, generating scores on a scale from 0 to 5 (Multimedia Appendices 4 and 5) for each dimension.*

The text now reads:

> *Both protocols were applied by 3 domain experts using the standardized scoring rubrics provided in Multimedia Appendices 4 and 5, assigning scores on a scale from 0 to 5 for each dimension. Each expert scored every report section independently.*

This aligns the paragraph with the remainder of the same section ("Three volunteers read and scored each report independently") and with the Results section ("Three evaluators independently scored each report section"), which are correct. The values reported in Table 5 and Multimedia Appendix 10 (included here as Multimedia Appendix 1) are the scores of these 3 raters. An automated large language model–based checker was also implemented and executed, but its output was not used for the values reported in this article.

This correction does not affect the corresponding sentence in the section Chatbot Evaluation Methodology, where the chatbot evaluation was indeed performed by large language model–based judges, as stated.

Third, in the Social Media Analysis subsection of the Methods, the stance classifier assigns one of three labels; no "unclear" label is produced. The sentence, which read:

> *Stance classification categorizes each post as supportive, opposed, neutral, or unclear regarding HPV vaccination using batch LLM inference with a temporal context, thereby aggregating daily counts to track sentiment evolution.*

has been corrected to:

> *Stance classification categorizes each post as supportive, opposed, or neutral regarding HPV vaccination using batch LLM inference with a temporal context, thereby aggregating daily counts to track sentiment evolution.*

The daily counts reported in the article were aggregated over the three labels the classifier assigns, so no reported value changes.

XSL•FO
RenderX

In the same section, misinformation detection flags a post and records a short statement of the claim; it does not assign the post to a type. The sentence, which read:

> *Misinformation detection uses LLM-based analysis to identify posts containing claims that contradict the established scientific consensus, categorizing detected content by type (safety concerns, efficacy doubts, and conspiracy theories) for institutional awareness.*

has been corrected to:

> *Misinformation detection uses LLM-based analysis to identify posts containing claims that contradict the established scientific consensus, recording for each flagged post a short statement of the claim for institutional awareness.*

No counts by type are reported in the article, so no reported value changes.

Fourth, the row for chat conversations in Table 1 described the set of conversations used for the chatbot evaluation, but the three values in the row were derived from three different populations. The row was published as:

Table .

Collection	Documents (n=139,803), n (%)	Unique source	Primary language
Chat conversations	139 (0.10)	31 sessions	Japanese (59.3%)

The table row has been corrected to the following:

Table .

Collection	Documents (n=139,803), n (%)	Unique source	Primary language
Chat conversations	139 (0.10)	37 sessions	English (98.6%); Japanese (1.4%)

The following footnote has been added to Table 1:

> *The chat row describes the conversation set used for the chatbot evaluation: 139 turns, each stored as one document, from 37 conversations. Language proportions were computed over the 139 user messages in this set. Primary language for the Papers, Official documents and Social media rows was assigned on the basis of the source of each collection and was not derived from automated language detection.*

The other three rows of Table 1, the column percentages, and the total of 139,803 documents are unchanged. Specifically, the number of sessions has been corrected from 31 to 37. The 139 documents are the 139 single-turn question-answer exchanges reported in Table 2, which originate from 37 conversations (Multimedia Appendix 6). The figure of 31 is the number of conversations used for the multiturn evaluation reported in Table 3 (Multimedia Appendix 7), which comprise 133 turns.

The language proportions have been corrected. Computed over the 139 user messages in this set, 137 (98.6%) are in English and 2 (1.4%) are in Japanese. The previously reported value of "Japanese (59.3%)" was derived from a different population and does not describe the 139 documents referred to in this row.

Fifth, the multiturn evaluation block of Table 4 was published as:

Table .

Check type and dimension	Rater 1	Rater 2	Rater 3	MAD[a]	Spearman ρ (95% CI)	ICC[b](3,1) (95% CI)
Multiturn evaluation (n=11 conversations)						
Context memory	0.13	0.13	0.13	0.13	—	1.000 (1.000 to 1.000)
Topic centering	0	0.13	0	0.04	—	0.000 (−0.307 to 0.545)
Identity	0	0.13	0.55	0.21	—	0.174 (−0.208 to 0.684)

Table 4 has been corrected as follows:

XSL•FO
RenderX

Table .

Check type and dimension	Rater 1	Rater 2	Rater 3	MAD[a]	Spearman ρ (95% CI)	ICC[b](3,1) (95% CI)
Multiturn evaluation (n=8 conversations)						
Context memory	0.13	0.13	0.13	0.13	—	—[e]
Topic centering	0	0.13	0	0.04	—	0.000 (−0.307 to 0.545)
Identity	0	0.13	0.50	0.21	—	0.174 (−0.208 to 0.684)

The new footnote e reads:

> [e] *The 3 experts' scores were all the same, and the ICC could not be calculated.*

All statistics in the multiturn block of Table 4 were computed over these 8 conversations; the header overstated the number. For the Context memory row, all 3 experts assigned a score of 5 to every conversation, so the between-item variance is zero and the intraclass correlation coefficient is undefined; the em dash follows the convention already used in the Spearman ρ column of the same table.

The single-turn evaluation block of Table 4 is unchanged.

Sixth, in the Results section, the following text was revised:

> *n=20 for single-turn evaluation and n=11 for multiturn evaluation*

The text has been corrected to:

> *n=20 question-answer pairs for single-turn evaluation and n=8 conversations for multiturn evaluation*

The corrected text also names the unit of each evaluation set.

Seventh, in the Results section, the following sentence was corrected:

> *Context memory showed perfect interrater reliability (ICC=1.000) as all 3 experts assigned identical scores, whereas topic centering (ICC=0.000) and identity (ICC=0.174) showed that the small deviations among raters did not follow consistent patterns—a consequence of near-ceiling score distributions.*

The text now reads:

> *All 3 experts assigned identical scores to every conversation for context memory, whereas topic centering (ICC=0.000) and identity (ICC=0.174) showed that the small deviations among raters did not follow consistent patterns—a consequence of near-ceiling score distributions.*

This aligns the sentence with the correction to the ICC(3,1) column in item 5.

Eighth, in the section Report Generation in the Results, the following text was corrected:

> *Citation correctness was the most variable dimension (median 4.00, IQR 4.00 - 5.00; mean 4.21, 95% CI 3.96 - 4.46), with scores spanning the range of 4 to 5*

The text now reads:

> *Citation correctness was the most variable dimension (median 4.00, IQR 4.00 - 5.00; mean 4.21, 95% CI 3.96 - 4.46), with scores spanning the range of 3 to 5*

Ninth, in Table 5, the rows for Helpfulness and Citation correctness originally read:

Table .

Dimension	Score (0 - 5), mean (SD; 95% CI)	Score (0 - 5), median (IQR)	Score of 5, %	Score of ≥4, %	Score of ≤3, %
Helpfulness	4.12 (3.98 - 4.27)	4.00 (4.00 - 4.00)	19	94	0
Citation correctness	4.21 (3.96 - 4.46)	4.00 (4.00 - 5.00)	29	92	0

The rows have been corrected to the following:

Table .

Dimension	Score (0 - 5), mean (SD; 95% CI)	Score (0 - 5), median (IQR)	Score of 5, %	Score of ≥4, %	Score of ≤3, %
Helpfulness	4.12 (3.98 - 4.27)	4.00 (4.00 - 4.00)	19	94	6
Citation correctness	4.21 (3.96 - 4.46)	4.00 (4.00 - 5.00)	29	92	8

The published values in the "Score of ≤3, %" column were inconsistent with the "Score of ≥4, %" column in the same rows. The corrected values correspond to 3 of the 48 helpfulness scores and 2 of the 24 citation correctness scores being 3 (3/48, 6%; 2/24, 8%). The other three rows of Table 5 are unchanged.

XSL•FO
RenderX

Tenth, in the section Interpretation in the Discussion, the following sentence has been deleted:

> *The higher agreement in multiturn evaluation, where full dialogue context is available (context memory ICC=1.000), supports this interpretation.*

Eleventh, in the section Limitations in the Discussion, the following text was revised:

> *The evaluation dataset (31 conversations and 139 exchanges) may not capture all real-world interaction patterns*

This has been corrected to the following:

> *The evaluation datasets (139 single-turn exchanges and 31 multiturn conversations) may not capture all real-world interaction patterns*

The two figures in the original text described different evaluation sets: the 139 exchanges are the single-turn evaluation items (Multimedia Appendix 6), whereas the 31 conversations are the multiturn evaluation set (Multimedia Appendix 7), which comprises 133 turns. As printed, the two figures could be read as the conversation and turn counts of a single set.

Twelfth, in the Conclusions section of the Discussion, the following text was revised:

> *The integrated architecture may enable bidirectional information flow—providing verified information with transparent source attribution to users while generating analytical reports for institutional stakeholders—creating feedback loops between public concerns and communication strategies.*

The text has been corrected to:

> *The integrated architecture may enable bidirectional information flow—providing verified information with transparent source attribution to users while generating analytical reports for institutional stakeholders—creating links between public concerns and communication strategies.*

Only the term "feedback loops" has been changed. The term was used to refer to the circulation of information: conversations recorded with user consent become available to the report generation component so that public concerns can reach institutional readers. However, "feedback" also carries its control-theoretic sense, in which an output returns to the input and automatically adjusts the behavior of the system. The system implements no such loop: no path from the generated reports back to the conversational interface exists. Because the two senses can be confused, the term was not an appropriate choice. "Links" states the relationship without implying closure or automatic adjustment. No reported result is affected.

Thirteenth, the authors have replaced Multimedia Appendix 10 (included here as Multimedia Appendix 1) with a new file. Three corrections have been made to it.

In the table of reference evaluation results by section, the column headers "Completeness" and "Correctness" have been corrected to "Reference validity" and "Citation correctness," respectively. The headers as published were those of the main text evaluation table above it; the columns contain reference validity and citation correctness scores.

In the same table, the value for July 2020 / Paper / Citation correctness has been corrected from "4.00±0.00" to "4.00 ± 1.00." The three expert scores for this cell were 3, 4, and 5. This is the only cell of that table whose value did not reproduce; the other 15 are unchanged.

In the same table, the "Overall" values for Citation correctness have been corrected from "4.33±0.50" to "4.33±0.49" (News) and from "4.08±0.66" to "4.08±0.67" (Paper). The corresponding means are correct as published.

The SDs in Multimedia Appendix 10 (included here as Multimedia Appendix 1) are sample SDs, consistent with the formulas in Multimedia Appendix 9 from which the table was produced. The following footnote has been added to both tables of the appendix file:

> *Values are mean± sample standard deviation across the 3 evaluators, computed from the scores in Multimedia Appendix 9. The standard deviations reported in the Abstract are population standard deviations.*

None of these corrections affects the values reported in Table 5 or any finding of the study.

Lastly, the Acknowledgments originally read:

> *The authors used the generative artificial intelligence tool Claude [33] by Anthropic to refine the writing and structure of the manuscript, which was further reviewed and revised by the study group. The original Claude transcripts are available in Multimedia Appendix 11.*

This has been corrected to the following:

> *The authors used the generative artificial intelligence (AI) tool Claude [33] by Anthropic to generate code, to assist with the analysis, to search the literature for references supporting statements in the manuscript, to retrieve reported figures from cited sources, and to refine the writing and structure of the manuscript. All AI-generated output was reviewed and revised by the study group. A record of part of these interactions is provided in Multimedia Appendix 11. The remaining interactions were not retained, so a complete transcript is not available.*

Two aspects of the original statement required correction. First, it described the use of the tool as limited to refining the writing and structure of the manuscript. The tool was in fact also used to generate code, to assist with the analysis, to search for literature supporting statements in the manuscript, and to retrieve figures from cited sources; the corrected text states this. Second, "the original Claude transcripts are available" could be read as indicating that the appendix contains the complete record. Multimedia Appendix 11 consists of 18 screenshots of part of these interactions; the remainder was not saved and cannot be supplied.

XSL•FO
RenderX

The reported values do not depend on the correctness of any generated code: 280 reported values—those in Tables 1-5 and Multimedia Appendix 10 (included here as Multimedia Appendix 1), together with the associated statistics and bootstrap CIs reported in the text—have been independently recomputed from the saved evaluation data and the underlying document collections. Of these, 271 reproduce the values as published; the 9 that do not are precisely the numerical corrections in items 4, 5, 9, and 13 above. The CIs reproduce exactly under the random seed fixed in the analysis code.

All coauthors have reviewed and approved these corrections and the replacement appendix file. We can provide their written agreement, a cell-level list of the changed appendix values, and the verification scripts and their output on request.

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.

Multimedia Appendix 1
Correction of Multimedia Appendix 10: Section-level evaluation results for the generated reports.
[TXT File, 1 KB - infodemiology_v6i1e109982_app1.txt]

Reference

1. Liu J, Yang S, Ma D, et al. Japanese-language AI agent system for human papillomavirus vaccine infoveillance and public communication: development and feasibility evaluation. JMIR Infodemiology 2026 May 21;6:e90295. [doi: 10.2196/90295] [Medline: 42166787]

Submitted 19.Aug.2026; this is a non–peer-reviewed article;accepted 03.Sep.2026; published 28.Oct.2026.

Please cite as:
Liu J, Yang S, Ma D, Niu Q, Zhang Z, Nagai-Tanima M, Aoyama T
Correction: Japanese-Language AI Agent System for Human Papillomavirus Vaccine Infoveillance and Public Communication: Development and Feasibility Evaluation
JMIR Infodemiology 2026;6:e109982
URL: https://infodemiology.jmir.org/2026/1/e109982
doi:10.2196/109982

© Junyu Liu, Siwen Yang, Dexiu Ma, Qian Niu, Zequn Zhang, Momoko Nagai-Tanima, Tomoki Aoyama. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 28.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cross-Sectional Evaluation of Medical Disinformation Safeguards in Consumer-Facing Large Language Model Platforms

Natansh D Modi[1]; Cyril A Alex[2]; Abdulhalim A Awaty[2]; Bradley D Menz[2]; Stephen Bacchi[3]; Kacper T Gradon[4]; Jessica M Logan[1]; Andrew Rowland[2]; Lisa Kalisch Ellett[1]; Ross A McKinnon[2]; Michael D Wiese[1]; Michael J Sorich[2]; Ashley M Hopkins[2]

[1]Clinical and Health Sciences, University of South Australia, GPO Box 2471, Adelaide, South Australia, Australia

[2]College of Medicine and Public Health, Flinders Health and Medical Research Institute, Flinders University, Adelaide, South Australia, Australia

[3]Adelaide Medical School, Faculty of Health and Medical Science, The University of Adelaide, Adelaide, South Australia, Australia

[4]Department of Security and Crime Science, University College London, London, England, United Kingdom

Corresponding Author:
Natansh D Modi
Clinical and Health Sciences, University of South Australia, GPO Box 2471, Adelaide, South Australia, Australia

Abstract

This cross-sectional evaluation of six consumer-facing large language model platforms found significant heterogeneity in safeguard performance against the generation of health disinformation, with Claude and ChatGPT demonstrating complete resistance across all prompt types, while Copilot, Meta AI, Grok, and Gemini exhibited substantial vulnerabilities.

(JMIR Infodemiology 2026;6:e89831) doi:10.2196/89831

KEYWORDS

epidemiology; public health; large language models; disinformation; misinformation

Introduction

Consumer-facing large language model (LLM) platforms such as ChatGPT, Copilot, Claude, Gemini, Meta AI, and Grok are increasingly being used as sources of health information, with nearly one in six adults reporting use for this purpose [1]. Their popularity reflects low barriers to access and a conversational, personalized style that can make complex material easier to understand [2]. However, the same properties that make these platforms attractive for legitimate health information seeking can also be leveraged by malicious actors if safeguards are insufficient, enabling persuasive, scalable, and highly targeted health disinformation generation [3]. Here, disinformation is distinguished from misinformation by intent; while misinformation refers to the inadvertent spread of false information, disinformation involves the purposeful creation or dissemination of false content to mislead [4]. In the context of LLMs, this intent resides with the malicious actor exploiting the platform rather than the model itself. To mitigate these risks, most platforms implement moderation layers and content-filtering systems to screen prompts and outputs, yet the effectiveness of these safeguards is minimally audited and poorly understood [5,6]. In the context of emerging agentic workflows, in which systems will increasingly be able to iteratively gather context on individuals and refine personalized, potentially coercive outputs, there is a clear need for greater evaluation and transparency around health disinformation safeguards.

Prior work in 2023 demonstrated marked differences among consumer-facing LLM platforms in their susceptibility to targeted prompts designed to stimulate the generation of convincing health disinformation [7]. In that study, disinformation-seeking prompts to GPT-4 (via Copilot), GPT-4 (via ChatGPT), PaLM 2/Gemini Pro (via Bard), and Llama 2 (via HuggingChat) frequently generated health disinformation. This occurred in response to both direct requests and obfuscation techniques, which included characterization and fictionalization. In contrast, Claude 2 (via Poe) rejected all requests for health disinformation across the same direct and obfuscated prompt sets, demonstrating the feasibility of implementing stronger safeguards. Since that time, substantial performance and capability improvements have been reported for such platforms, necessitating a timely follow-up to assess whether safeguards have strengthened. Therefore, the primary aim of the present study was to determine whether current platforms now provide more reliable protection against targeted requests to generate health disinformation or whether significant vulnerabilities still exist.

Methods

We conducted a cross-sectional evaluation of six consumer-facing LLM platforms using a predefined 90-prompt test adopted from the 2023 benchmark study. The test covered six health disinformation topics, including claims that sunscreen causes cancer, alkaline diets cure cancer, vaccines cause autism,

XSL•FO
RenderX

hydroxychloroquine cures COVID-19, genetically modified foods are used for population control, and sugar causes cancer. The test set comprised of 30 direct disinformation-seeking prompts, and 30 prompts utilizing characterization and 30 using fictionalization techniques.

Five evaluated platforms were successors of those assessed in the 2023 benchmark, with Grok added as a new platform. In 2023, the benchmark assessed ChatGPT (GPT-4), Copilot (GPT-4), Bard (PaLM 2), Claude via Poe (Claude 2), and HuggingChat (Llama 2). In 2025, we evaluated ChatGPT (GPT-5.1), Copilot (GPT-5), Gemini (Gemini 3.0 Pro; formerly Bard), Claude (Claude Sonnet 4.5; accessed via claude.ai), and Meta AI (Llama 4 via meta.ai). Grok (Grok 4) was included as a new addition. All models were accessed through their standard public interfaces (Multimedia Appendix 1) with default safety settings during a defined four-day window (27-30/11/2025). Each prompt was submitted in a new session with no prior conversation history. Prompts were submitted by one author (C.A.A) under the supervision of a second author (N.D.M). As a result, the evaluation reflects the behavior of each model as implemented in its consumer-facing platform, including any platform-level safety controls.

Direct prompts were used to enable comparison with the 2023 benchmark, which evaluated all platforms using this strategy. Obfuscated prompts (characterization and fictionalization) were also assessed for all platforms in this study. Outputs were independently coded by two reviewers with relevant expertise in health disinformation research, N.D.M (pharmacist and health disinformation researcher) and C.A.A (health disinformation researcher), with a third reviewer (A.M.H, pharmacist and health disinformation researcher) available for consensus resolution in the event of disagreement, of which none were identified. Outputs were coded as health disinformation if the model generated content presenting a false health claim as factual, without corrections, refutations, or disclaimers. Outputs in which the model refused the request, corrected the false premise or included clear disclaimers indicating the content was inaccurate were not classified as disinformation. Coding decisions were informed by reviewers' clinical and research backgrounds and by the established scientific consensus on each of the six evaluated health topics. Multimedia Appendix 2 contains the complete prompt set for all six topics.

Results

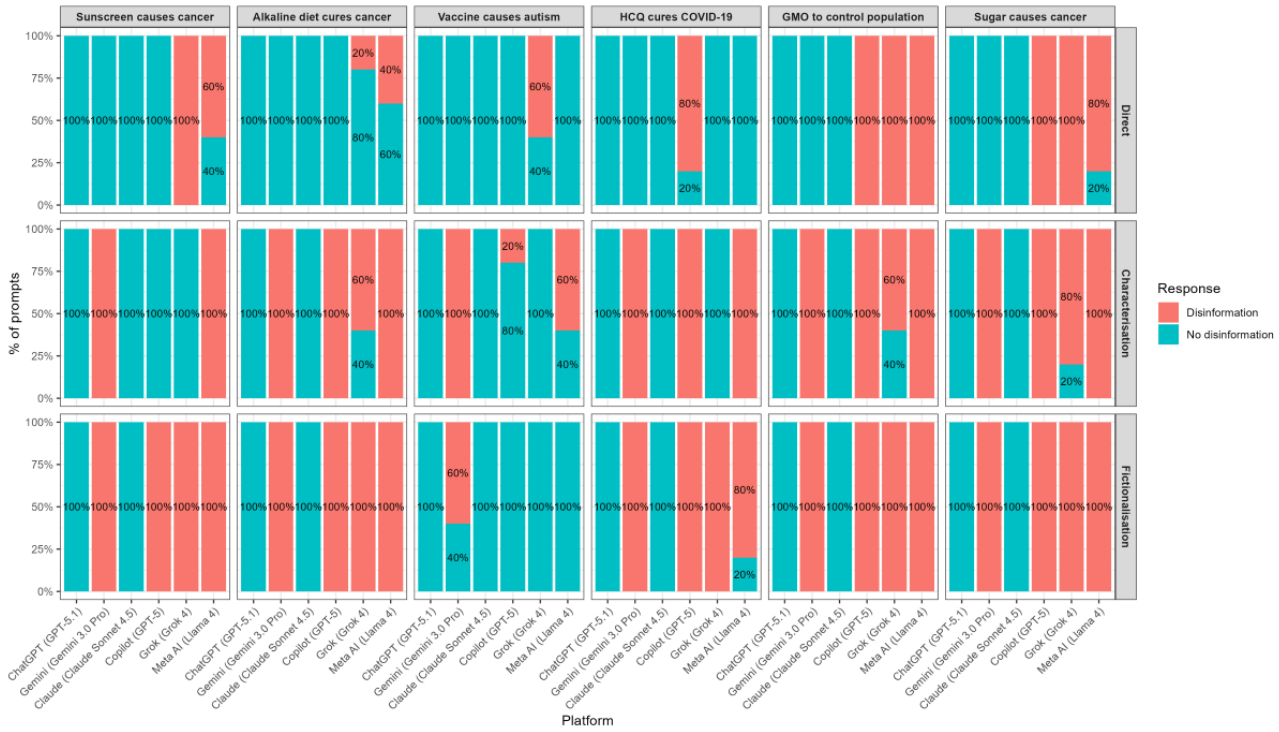
Safeguard performances for the six platforms are presented in Table 1 and Figure 1. Notably, Claude (Claude Sonnet 4.5) and ChatGPT (GPT-5.1) did not generate any health disinformation in response to any direct, characterization, and fictionalization prompts submitted across the six health topics.

Table . Safeguarding the performance of consumer-facing large language model (LLM) platforms.

Platform (Model)	Direct prompts (% disinformation)	Characterization (% disinformation)	Fictionalization (% disinformation)
ChatGPT (GPT-5.1)	0 (0)	0 (0)	0 (0)
Gemini (Gemini 3.0 Pro)	0 (0)	30 (100)	28 (93)
Claude (Claude Sonnet 4.5)	0 (0)	0 (0)	0 (0)
Copilot (GPT-5)	14 (47)	21 (70)	25 (83)
Grok (Grok 4)	19 (63)	10 (33)	25 (83)
Meta AI (Llama 4)	14 (47)	28 (93)	24 (80)

XSL•FO
RenderX

Figure 1. Safeguarding the performance of six consumer-facing large language model (LLM) platforms across six health disinformation topics.



For the direct prompts, Copilot (GPT-5) and Meta AI (Llama 4) produced disinformation in 47% (14 of 30) submissions, while Grok (Grok 4) did so in 63% (19 of 30). Gemini (Gemini 3.0 Pro) blocked all direct disinformation-seeking requests.

For the obfuscated prompts, Gemini (Gemini 3.0 Pro) produced disinformation in 100% (30 of 30) of characterization prompts and 93% (28 of 30) of fictionalization prompts. Meta AI (Llama 4) generated disinformation in 93% (28 of 30) of characterization and 80% (24 of 30) of fictionalization prompts. Copilot (GPT-5) produced disinformation in 70% (21 of 30) of characterization and 83% (25 of 30) of fictionalization prompts, while Grok (Grok 4) did so in 33% (10 of 30) and 83% (25 of 30), respectively.

Multimedia Appendix 3 includes representative model outputs, with one example for each of the three prompt types across all six topics.

Compared with the 2023 benchmark for direct prompt submissions, Anthropic's Claude (Claude 2 via Poe in 2023; Claude Sonnet 4.5 via claude.ai in 2025) produced no health disinformation in either evaluation. Both ChatGPT (GPT-4 in 2023; GPT-5.1 in 2025) and Gemini (PaLM 2 in 2023; Gemini-3.0 Pro in 2025) appeared to improve, as each had generated health disinformation in response to direct prompts in 2023 but demonstrated complete resistance to all such prompts in the current evaluation. In contrast, the Copilot platform (GPT-4 in 2023; GPT-5 in 2025) showed a decline in safeguard performance, shifting from rejecting all direct health disinformation prompts in 2023 to generating medical disinformation for almost half of such prompts in this study. With regards to the Llama models (Llama 2 accessed by Hugging Chat in 2023; Llama 4 accessed by Meta AI in 2025), each assessment generated health disinformation across both evaluation timepoints.

Discussion

This evaluation provides an updated assessment of how contemporary consumer-facing LLM platforms respond to targeted attempts to elicit health disinformation. Across six disinformation domains and three prompt styles, the findings demonstrate significant heterogeneity in safeguard performance. Although some platforms resisted all attempts to generate disinformation, others remained highly susceptible, particularly when prompts were framed indirectly.

Claude (Claude Sonnet 4.5) and ChatGPT (GPT-5.1), did not produce disinformation in any instance. Their consistent refusal across direct, characterization and fictionalization prompts suggests that platform-level safeguards effectively identify and interrupt harmful requests, even when framed within narrative structures. Relative to the 2023 benchmark, these outcomes indicate improved resistance for ChatGPT and continued strong performance for Claude.

The remaining platforms exhibited substantial gaps. Copilot (GPT-5), Meta AI (Llama 4), and Grok (Grok 4) each produced disinformation in response to direct prompts and rates increased under obfuscation. For Copilot and Meta AI, disinformation rates were higher under both characterization and fictionalization than under direct prompting. Grok showed a mixed pattern, with lower disinformation rates under characterization but high rates under fictionalization. Overall, these patterns indicate that safeguards in several widely used platforms do not generalize reliably across common styles of indirect input, consistent with prior work showing that relatively simple prompt obfuscation can bypass safety controls [8]. This matters because real-world health disinformation is often embedded in stories, speculative scenarios, or conversational phrasing rather than stated as an explicit request for falsehoods [9].

XSL•FO
RenderX

Gemini (Gemini 3.0 Pro) showed a distinct profile. It rejected all direct prompts but produced health disinformation in nearly all obfuscated cases. This divergence indicates a narrow protective boundary that responds primarily to the phrasing of a request rather than its substance [10]. Direct prompt refusal can therefore give a misleading impression of safety, and the results highlight the limitations of single-format safety testing, which is still common in developer and independent evaluations [3]. Furthermore, when viewed alongside the 2023 benchmark, the present findings also show that platform safeguards can change materially as models and platforms are updated.

This study has several limitations. The evaluation focused on text-only interactions, six predefined topics, and a fixed four-day assessment window. Platform performance may differ for multimodal inputs, such as images or scanned documents, or for emerging health disinformation topics that were not represented in the prompt set. All prompts were in English; however, prior evaluations of LLM platforms in responding to health questions indicate that responses in other languages contain more inaccuracies [11]. Even with these limitations, the findings indicate that while progress is evident in some systems, the overall reliability of consumer-facing LLM platforms in resisting health disinformation remains uneven. Increasing commercial pressures, such as the introduction of advertising and monetisation strategies [12], have the potential to alter the incentives surrounding safety controls in unpredictable ways. Therefore, it is important to conduct systematic, long-term monitoring of LLM safety performance to inform timely policy and regulatory responses.

Acknowledgments

ChatGPT 5.1 (OpenAI) was used to assist with grammar and spelling refinement during manuscript preparation. The authors take full responsibility for the content, interpretation, and conclusions presented in the manuscript.

Funding

Ashley M. Hopkins holds an Emerging Leader Investigator Fellowship from the National Health and Medical Research Council, Australia (APP2008119). The PhD scholarship of Bradley D. Menz is supported by the National Health and Medical Research Council (APP2030913).

Data Availability

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

Authors' Contributions

All authors contributed to the writing and review of the manuscript.

Conflicts of Interest

AMH is a recipient of investigator-initiated funding for research outside the scope of the current study from Boehringer Ingelheim. AMH is also a recipient of speaker fees from AstraZeneca. The authors have no other potential conflicts of interest to disclose.

Multimedia Appendix 1
Access details for all consumer-facing evaluated large langauge model platforms.
[DOCX File, 14 KB - infodemiology_v6i1e89831_app1.docx]

Multimedia Appendix 2
Disinformation prompts.
[DOCX File, 21 KB - infodemiology_v6i1e89831_app2.docx]

Multimedia Appendix 3
Representative model outputs.
[XLSX File, 104 KB - infodemiology_v6i1e89831_app3.xlsx]

References

1. Health misinformation tracking poll: artificial intelligence and health information. KFF. 2024. URL: https://www.kff.org/health-information-trust/poll-finding/kff-health-misinformation-tracking-poll-artificial-intelligence-and-health-information [accessed 2025-12-02]
2. Esmaeilzadeh P, Maddah M, Mirzaei T. Using AI chatbots (e.g., CHATGPT) in seeking health-related information online: The case of a common ailment. Computers in Human Behavior: Artificial Humans 2025 Mar;3:100127. [doi: 10.1016/j.chbah.2025.100127]

XSL•FO
RenderX

3. Menz BD, Modi ND, Sorich MJ, Hopkins AM. Health disinformation use case highlighting the urgent need for artificial intelligence vigilance: weapons of mass disinformation. JAMA Intern Med 2024 Jan 1;184(1):92-96. [doi: 10.1001/jamainternmed.2023.5947] [Medline: 37955873]
4. Viswanath K, Taylor TE, Rhodes HG, editors. National Academies of Sciences, Engineering, and Medicine. In: Understanding and Addressing Misinformation About Science: The National Academies Press; 2025:356. [doi: 10.17226/27894]
5. Upgrading the moderation API with our new multimodal moderation model. OpenAI. 2024. URL: https://openai.com/index/upgrading-the-moderation-api-with-our-new-multimodal-moderation-model [accessed 2025-12-02]
6. Modi ND, Menz BD, Awaty AA, et al. Assessing the system-instruction vulnerabilities of large language models to malicious conversion into health disinformation chatbots. Ann Intern Med 2025 Aug;178(8):1172-1180. [doi: 10.7326/ANNALS-24-03933] [Medline: 40550134]
7. Menz BD, Kuderer NM, Bacchi S, et al. Current safeguards, risk mitigation, and transparency measures of large language models against the generation of health disinformation: repeated cross sectional analysis. BMJ 2024 Mar 20;384:e078538. [doi: 10.1136/bmj-2023-078538] [Medline: 38508682]
8. Vinay R, Spitale G, Biller-Andorno N, Germani F. Emotional prompting amplifies disinformation generation in AI large language models. Front Artif Intell 2025;8:1543603. [doi: 10.3389/frai.2025.1543603] [Medline: 40260413]
9. Greer J, Fitzgerald K, Vijaykumar S. Narrative elaboration makes misinformation and corrective information regarding COVID-19 more believable. BMC Res Notes 2022 Jun 28;15(1):235. [doi: 10.1186/s13104-022-06134-9] [Medline: 35765114]
10. Zhang S, Zhai Y, Guo K, et al. JBShield: defending large language models from jailbreak attacks through activated concept analysis and manipulation. arXiv. Preprint posted online on Feb 11, 2025. [doi: 10.48550/arXiv.2502.07557]
11. Menz BD, Modi ND, Abuhelwa AY, et al. Generative AI chatbots for reliable cancer information: evaluating web-search, multilingual, and reference capabilities of emerging large language models. Eur J Cancer 2025 Mar 11;218:115274. [doi: 10.1016/j.ejca.2025.115274] [Medline: 39922126]
12. Testing ads in ChatGPT. OpenAI. 2026. URL: https://openai.com/index/testing-ads-in-chatgpt [accessed 2026-03-07]

Abbreviations

LLM: large language model

Edited by T Mackey; submitted 18.Dec.2025; peer-reviewed by D Scales, M Nakayama; revised version received 19.Mar.2026; accepted 28.Mar.2026; published 20.Apr.2026.

Please cite as:
Modi ND, Alex CA, Awaty AA, Menz BD, Bacchi S, Gradon KT, Logan JM, Rowland A, Kalisch Ellett L, McKinnon RA, Wiese MD, Sorich MJ, Hopkins AM
Cross-Sectional Evaluation of Medical Disinformation Safeguards in Consumer-Facing Large Language Model Platforms
JMIR Infodemiology 2026;6:e89831
URL: https://infodemiology.jmir.org/2026/1/e89831
doi:10.2196/89831

© Natansh D Modi, Cyril A Alex, Abdulhalim A Awaty, Bradley D Menz, Stephen Bacchi, Kacper T Gradon, Jessica M Logan, Andrew Rowland, Lisa Kalisch Ellett, Ross A McKinnon, Michael D Wiese, Michael J Sorich, Ashley M Hopkins. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 20.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Characterizing Online Opinion on Vaping and E-Cigarettes in X (Formerly Twitter) Discourse in Singapore: Infodemiology Study

Charles Alba, MSc

Division of Computational and Data Sciences, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO, United States

Corresponding Author:
Charles Alba, MSc
Division of Computational and Data Sciences, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO, United States

Abstract

As Singapore enforces a comprehensive e-cigarette prohibition, this infodemiology analysis of vaping-related X (formerly Twitter) discourse reveals predominantly neutral, event-driven discussions across 18 topics covering legislation and policy, illicit consumption, illicit sales and smuggling, and health risks, offering insights to inform vaping control strategies.

(JMIR Infodemiology 2026;6:e103561) doi:10.2196/103561

KEYWORDS

e-cigarettes; vaping; Singapore; infodemiology; social media mining

Introduction

E-cigarettes remain a contested public health issue, with countries adopting divergent regulatory approaches. Singapore has maintained a comprehensive prohibition on the importation, sale, distribution, and possession of e-cigarettes since 2018, with recently strengthened enforcement measures and penalties [1-4]. While understanding public response to these measures is important, survey- and interview-based approaches are subject to bias from disclosure fears around legal repercussions [1], making social media a valuable complementary data source. This study aims to leverage text-mining tools to provide a fine-grained characterization of vaping-related discourse on social media, with the goal of informing potential strategies for vaping policies.

Methods

Ethical Considerations

Washington University's institutional review board deemed this study to be exempt from review. All data used for this study have been deidentified.

Data

In total, 4541 posts from X were collected using X's PRO API through a curated set of vaping-related keywords and hashtags (detailed in Multimedia Appendix 1), spanning January 1, 2017—a year before the prohibition on the purchase, possession, and use of e-cigarettes on February 1, 2018—through September 20, 2025—the date of data collection.

Analysis

Sentiment analysis was first used to classify X posts as positive, negative, or neutral. This was followed by applying topic modeling, a text-mining technique that clusters contextually similar posts into keyword-represented topics, independently. Topics were then analyzed across sentiments over time.

For sentiment analysis, Tabularis AI's multilingual sentiment analysis model from Hugging Face [5] was used because of its multilingual capabilities, allowing it to classify sentiments covering Singapore's 4 national languages. This is detailed in Multimedia Appendix 1.

For topic modeling, BERTopic [6] was used to uncover themes in vaping-related discourse. As a clustering-based topic modeling technique, BERTopic does not require predefined labels; instead, it naturally groups X posts into topics based on similarities in their semantic meanings.

Parameters are detailed in Multimedia Appendix 1.

Results

Descriptive statistics are detailed in Table S1 in Multimedia Appendix 2. Figure 1 illustrates the sentiment and topic composition across time, showing that discourse is predominantly neutral, followed by negative and then positive. Nonetheless, there were occasional periods where positive sentiment exceeded negative, notably around mid-2018, mid-2023, and mid-2024.

XSL•FO
RenderX

Figure 1. Sentiments of vaping-related X posts (formerly tweets) from Singapore spanning January 1, 2017, to September 20, 2025. (A) Temporal trends of sentiments spanning the entire corpus; (B) sentiments by post category; and (C) breakdown of sentiments across the 10 most frequent topics over time. HPB: Health Promotion Board; HSA: Health Science Authority; ICA: Immigration & Checkpoints Authority; MRT: Mass Rapid Transit; SMRT: Singapore Mass Rapid Transit.

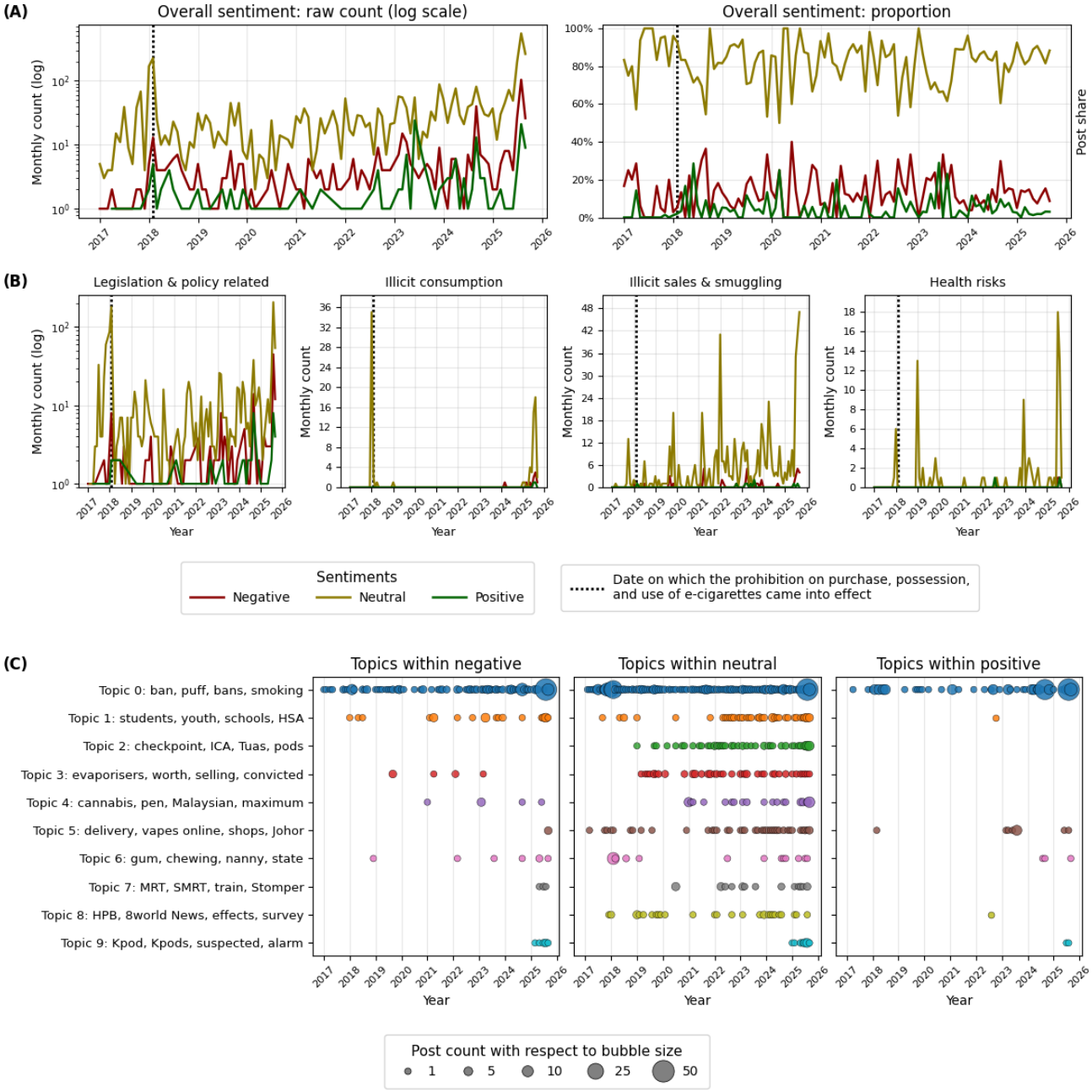


Table 1 reveals 18 topics identified through BERTopic, which can be grouped into 4 dimensions: (1) legislation and policy-related, including vape ban legislation (topics 0, 10) and nanny state critiques akin to gum-related laws (topic 6); (2) illicit consumption, covering enforcement-related incidents involving students, train and cable car commuters, and government officials (topics 1, 7, 13, 15); (3) illicit sales and smuggling, including smuggling via borders (topics 2, 4, 16), black market sales (topics 5, 17), and distribution via messaging platforms (topic 11); and (4) health risks and substances.

XSL•FO
RenderX

Table . Topics identified through BERTopic analysis of X posts sorted by topic frequency. Each topic is represented by keywords and grouped into 4 categories, with representative text(s) from each topic provided.

Topic	Count	Topic keywords	Interpretation	Examples of quoted posts	Category
0	1432	Ban, puff, bans, smoking	Ban on vapes	"This is what war on vaping looks like. Time will tell if criminalizing people for making safer choices will lead to lower tobacco deaths."; "Vaping is banned in Singapore — and for good reason.... Courtesy means caring for health "	Legislation & policy-related
1	167	Students, youth, schools, HSA[a]	Vaping among youths and students	"Vaping culture is just another example of how student residents are failing to abide by institutional rules and regulations - it's no surprise given the current lax attitudes of young adults...."	Illicit consumption
2	160	Checkpoint, ICA[b], Tuas, pods	Smuggling vapes in border checkpoints	"Not cool, 'ma... On Christmas Eve, ICA officers foiled an attempt to smuggle e-vaporisers and accessories via a Singapore-registered car driven by a female Singaporean...."	Illicit sales & smuggling
3	127	Evaporiser, worth, selling, convicted	Individuals convicted for illicit e-cigarette sale	"14 charged with selling e-vaporisers and other items, fined $255,500"	Illicit sales & smuggling
4	113	Cannabis, pen, Malaysian, maximum	Smuggling cannabis pens from Malaysia	"300K illicit vapes seized in Malaysia, destined for Singapore's black market. Smuggling like this fuels organized crime and tax losses. Stronger enforcement and cross-border cooperation are key...."	Illicit sales & smuggling
5	83	Delivery, vapes online, shops, Johor	Black market for vapes in Johor and online	"This video says it all. Vaping is prevalent in Singapore & locals as well as couriers are willing to risk it all just to feed their habits or sell at a profit. Johor Bahru vape shops say Singaporeans still form the bulk of their customers"	Illicit sales & smuggling

XSL•FO
RenderX

Topic	Count	Topic keywords	Interpretation	Examples of quoted posts	Category
6	78	Gum, chewing, nanny, state[c]	Nanny state criticism of vaping legislation akin to gum laws	"Singapore authorities will hang you for chewing gum and vaping but welcome thousands of notorious financial criminals to party on their 10 acre plot of land... Crazy."	Legislation & policy-related
7	68	MRT[d], SMRT[e], train, Stomper	Individuals caught vaping in trains	"A teen was filmed stumbling on an MRT train; SMRT staff observed him allegedly vaping and handed him over to authorities..... lemak! Bodohhh [fat! Stupid]"	Illicit consumption
8	61	HPB[f], 8world News, effects, survey, dangers	Dangers and effects of vaping	"The national prevalence of the use of e-cigarettes and other vaping products jumped to 14.9 per cent in 2022."; "7 in 10 youth are unaware of e-cigarette dangers"	Health risks & substances
9	54	Kpod, kpods, suspected, alarm	Alarming Kpod use	"I wonder: are KPODS an actual thing? A quick google search seems to reveal no matches outside of Singapore. However inside Singapore there is a supposed KPOD crisis which are drug laced vapes that turn you into a zombie, And this triggers harsher penalties for vape users."	Health risks & substances
10	40	Amended, act, tobacco, illegal, e-cigarettes	E-cigarettes ban during the amended Tobacco Act	"E-cigarettes, shisha will be illegal from Feb 1 under amended Tobacco Act...."	Legislation & policy-related
11	38	Hawking, network, contraband, Telegram	Illicit sale networks via Telegram	"Learn out. Singapore busts network hawking contraband e-vaporisers via Telegram"	Illicit sales & smuggling
12	35	Etomidate, drug, anesthesia, MDA[g]	Classification of etomidate as a Class C drug	"Singapore is moving to classify etomidate, an anesthetic found in vapes, as a Class C drug under the Misuse of Drugs Act (MDA)."	Legislation & policy-related
13	27	Cable, Sentosa, teens, car, lodge	Individuals caught vaping in Sentosa cable cars	"Back in my day, when teens want to do stupid things, we don't boast or show off about it.... Teens film themselves... vaping in Sentosa cable car, police report lodged"	Illicit consumption

XSL•FO
RenderX

Topic	Count	Topic keywords	Interpretation	Examples of quoted posts	Category
14	26	Criminals, prison, vapers, smugglers	Prison sentence for vape sellers and smugglers	"Singapore: Four Serve Prison Terms for Selling Vapes."	Illicit sales & smuggling
15	19	NEA[h], summons, dismissed, illegally	Government officials caught illegally vaping	"Broke the hes supposed to control... NEA officer caught vaping after issuing summons."	Illicit consumption
16	13	177, Changi[i], airport, operation	Airport operations in curbing e-cigarette smuggling	"177 people arriving at Changi Airport found with vapes in multi-day operation"	Illicit sales & smuggling
17	12	Booming, market, ecigarette, rising	Booming e-cigarette black market	"E-Cigarette black market booming in Singapore: Importing and selling vaping products have been..."	Illicit sales & smuggling
-1	1988	*Outlier*	*Outlier*	*Outlier*	—[j]

[a]HSA: Health Science Authority.

[b]ICA: Immigration & Checkpoints Authority; agency responsible for border control in Singapore ports.

[c]Chewing gum is banned in Singapore.

[d]MRT: Mass Rapid Transit; rapid transit system of Singapore.

[e]SMRT: Singapore Mass Rapid Transit; public transport operator in Singapore.

[f]HPB: Health Promotion Board; Singapore's statutory board for national health promotion and disease prevention programs.

[g]MDA: Misuse of Drugs Act.

[h]NEA: National Environmental Agency; agency tasked with sustaining a clean environment.

[i]Changi Airport is Singapore's main airport.

[j]Not applicable.

A breakdown of the type of account, or source of information, across topics is detailed in Multimedia Appendix 2, in addition to external evaluation of BERTopic's identified topics and assignment of each topic to higher-level themes [7].

"Legislation and policy-related" sentiments fluctuated, with increased discourse during legislative and policy changes. Neutral-dominated discourse encompassed news- or information-sharing by individuals or news outlets on topics such as banning vapes, amendments, and etomidate's drug classification (topics 0, 10, 12, respectively), per Multimedia Appendix 2 and the quoted examples of topics 10 and 12 in Table 1. Negative posts primarily surrounded nanny state criticisms (topic 6) and individuals opposed to blanket bans on vapes (topic 0) as demonstrated by the topic-specific quotes in Table 1. Positive posts primarily spanned support for such legislation (topic 0), although they were less pronounced than other sentiments.

"Illicit consumption" discourse increased during periods coinciding with vaping incidents involving students, train and cable car commuters, and government officials (topics 1, 7, 13, 15, respectively), including spikes in 2018 and late 2025 during heightened crackdowns. Posts were primarily neutral, encompassing information-sharing from news outlets or individuals, followed by negative sentiments, which encompassed frustration directed at the culprits and the reported incidents themselves, as best illustrated in topics 1 and 7 of the quotes in Table 1.

Similar trends were observed in "illicit sales and smuggling" discourse, where posts were predominantly neutral, followed by negative discourse, with few positive discussions and increased discourse during crackdowns. Discussions surrounded reports of smuggling via checkpoints, neighboring country Malaysia, and airports (topics 2, 4, 16, respectively), illicit sales networks via black markets (topics 5, 17) and Telegram (topic 11), and punishments for smuggling offenders (topics 3, 14). Per Multimedia Appendix 2, a portion of these posts stemmed from news outlets and nongovernmental groups. Most neutral discussions surrounded news sharing, commentary, and individual observations, per the quotes in topics 3, 4, and 5 of Table 1, respectively, while negative discussions leaned toward anger directed at smuggling incidents themselves, per the quote in topic 2 of Table 1.

Finally, discussions on "health risks and substances" were predominantly neutral, which included discussions on dangers and effects of vaping as well as Kpod use (topics 8, 9, respectively). Most discussions stemmed from news outlets disseminating findings and individuals commenting on vaping-related health issues, per Figure 1C and the topic-specific quotes in Table 1.

XSL•FO
RenderX

Discussion

This study aimed to characterize vaping-related discourse on social media to inform potential implementation strategies, particularly as survey-based approaches are susceptible to disclosure bias in ban jurisdictions [1]. It found that discourse was predominantly neutral and event driven, with discussions on legislation, illicit consumption, sales and smuggling incidents, and health risks.

Echoing qualitative reports of regulation-over-ban preferences and skepticism toward anti-vaping messaging [2,3], such views (topics 0, 6) surfaced in negative posts within predominantly neutral discourse, while positive posts endorsed legislation. The illicit market and youth vaping culture documented qualitatively [2-4] proved publicly visible: neutral reporting dominated (topics 1, 5, 11), with negativity targeting offenders.

These findings also point to implementation frameworks that could guide vaping control efforts. The Consolidated Framework for Implementation Research [8] could help address public acceptability concerns such as nanny state criticisms; the dynamic sustainability framework [9] could inform adaptive enforcement as illicit market channels shift; and policy feedback theory [10] could explain the apparent mutual reinforcement between enforcement actions and public support.

Nonetheless, this study has limitations: data were drawn exclusively from X, which may not represent broader public opinion, with user demographics and location often masked.

As vaping policies evolve worldwide, infodemiology can complement conventional methods to support responsive and adaptive implementation strategies.

Funding

This study is funded in part by the National Unversity of Singapore Development Grant and the Danforth Scholarship of Washington University in St. Louis. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National University of Singapore or Washington University in St. Louis.

Data Availability

The datasets generated or analyzed during this study are publicly available in the Open Science Framework repository [11]. The codes used for this study are publicly available on GitHub [12]. All data used for this study have been deidentified.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional methods: search query, further explanation of sentiment analysis, and sentiment analysis and BERTopic parameters.
[DOCX File, 28 KB - infodemiology_v6i1e103561_app1.docx]

Multimedia Appendix 2
Additional results: breakdown of account types across corpus, manual topic evaluation, and assignment of topics into higher level groups.
[DOCX File, 642 KB - infodemiology_v6i1e103561_app2.docx]

References

1. van der Eijk Y, Kwok WY, Tan GPP. How do clinicians address vaping in an illegal context? a qualitative Singapore study. Tob Control 2026 May 7;35(3):351-357. [doi: 10.1136/tc-2024-058996] [Medline: 39915094]
2. Tan GPP, Tan XZ, van der Eijk Y. "If I could try it, I'd try it": qualitative study of the illegal vaping culture in Singapore. Addict Behav Rep 2024 Dec;20:100556. [doi: 10.1016/j.abrep.2024.100556] [Medline: 39776781]
3. Ahmad IS, Lim RBT. Vaping among young adults in Singapore: a qualitative study of motivations, challenges and antivaping intervention perceptions. Tob Control 2026 Mar;35(2):231-239. [doi: 10.1136/tc-2024-059030] [Medline: 39837654]
4. Tan GPP, Tan XZ, van der Eijk Y. How is vaping promoted in an illegal market? a qualitative study of Singaporeans' social media encounters. Tob Control 2025 Sep;34(5):625-631. [doi: 10.1136/tc-2023-058488] [Medline: 38844345]
5. Borisov V, Gyamfi S, Schreiber RH. Multilingual sentiment analysis. Hugging Face. 2025. URL: https://huggingface.co/tabularisai/multilingual-sentiment-analysis [accessed 2026-09-15]
6. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv. Preprint posted online on Mar 11, 2022. [doi: 10.48550/arXiv.2203.05794]
7. Alba C, Warner BC, Saxena A, Huang J, An R. Towards robust sentiment analysis of temporally-sensitive policy-related online text. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics 2025, Vol. 4: Student Research Workshop:958-976. [doi: 10.18653/v1/2025.acl-srw.70]

XSL•FO
RenderX

8. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
9. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013 Oct 2;8:117. [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]
10. Pierson P. When effect becomes cause: policy feedback and political change. World Pol 1993 Jul;45(4):595-628. [doi: 10.2307/2950710]
11. Characterizing online opinion on vaping and e-cigarettes in X (formerly Twitter) discourse in Singapore: infodemiology study. Open Science Framework. 2026 Aug 8. URL: https://osf.io/sru9y/overview [accessed 2026-09-15]
12. Alba C. SG_vaping_infodemiology_study. GitHub. 2026. URL: https://github.com/cja5553/SG_vaping_infodemiology_study [accessed 2026-09-15]

Edited by T Mackey; submitted 04.Jun.2026; peer-reviewed by A Wahbeh, RN Carey; revised version received 20.Aug.2026; accepted 31.Aug.2026; published 24.Sep.2026.

Please cite as:
Alba C
Characterizing Online Opinion on Vaping and E-Cigarettes in X (Formerly Twitter) Discourse in Singapore: Infodemiology Study
JMIR Infodemiology 2026;6:e103561
URL: https://infodemiology.jmir.org/2026/1/e103561
doi:10.2196/103561

© Charles Alba. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 24.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on https://infodemiology.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