JMIR Infodemiology

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

Contents

Viewpoint

Review

Original Papers

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 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" 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 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.

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 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

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]

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

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

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 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, 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].

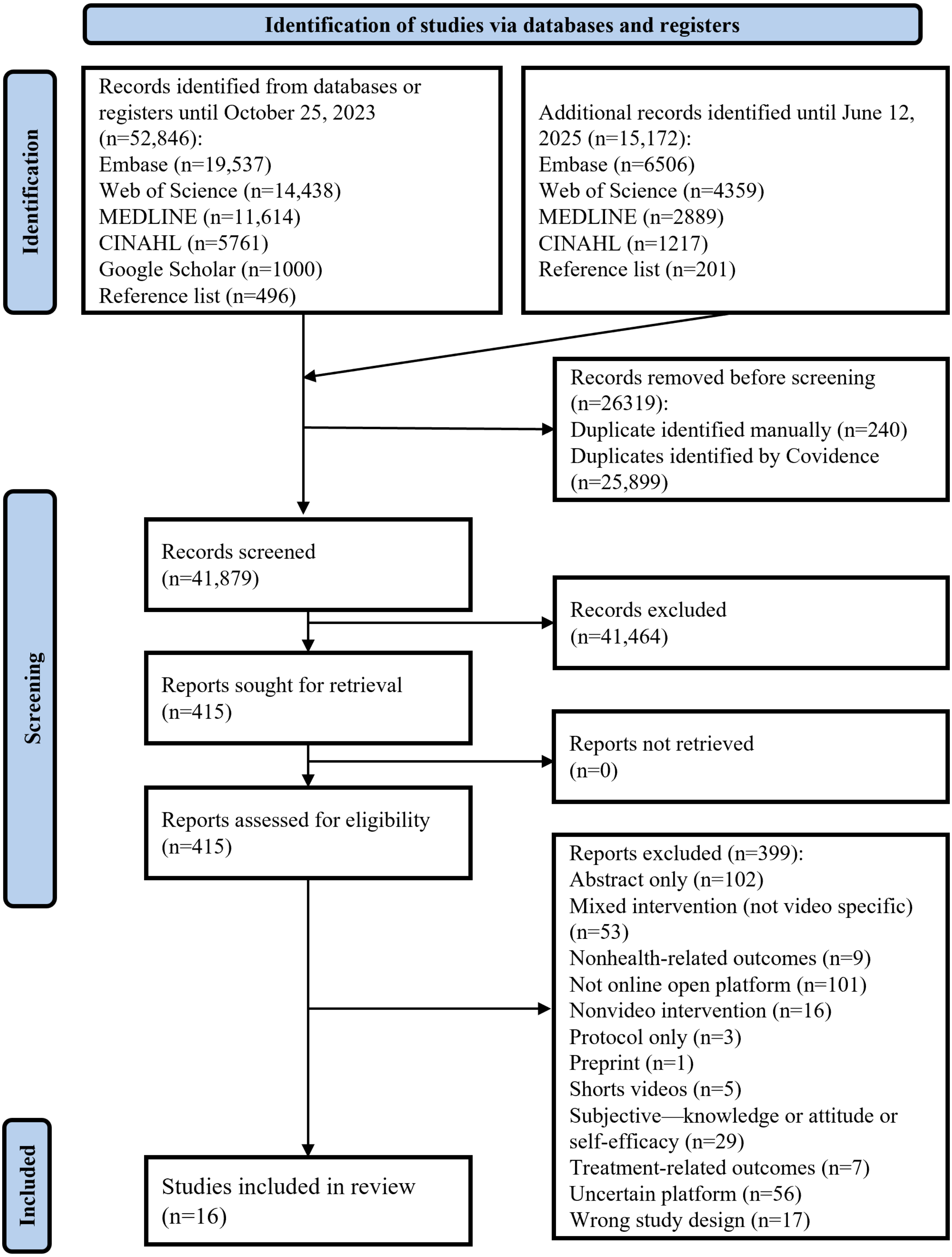
For all interventional studies, an albatross plot was constructed to allow the *P* values to be interpreted in the context of study sample size. The contour lines of the albatross plots were formed by hypothetical effect sizes [41]. *P* values were calculated from the SMD values using the Wald test [42]. 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.

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], 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 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].

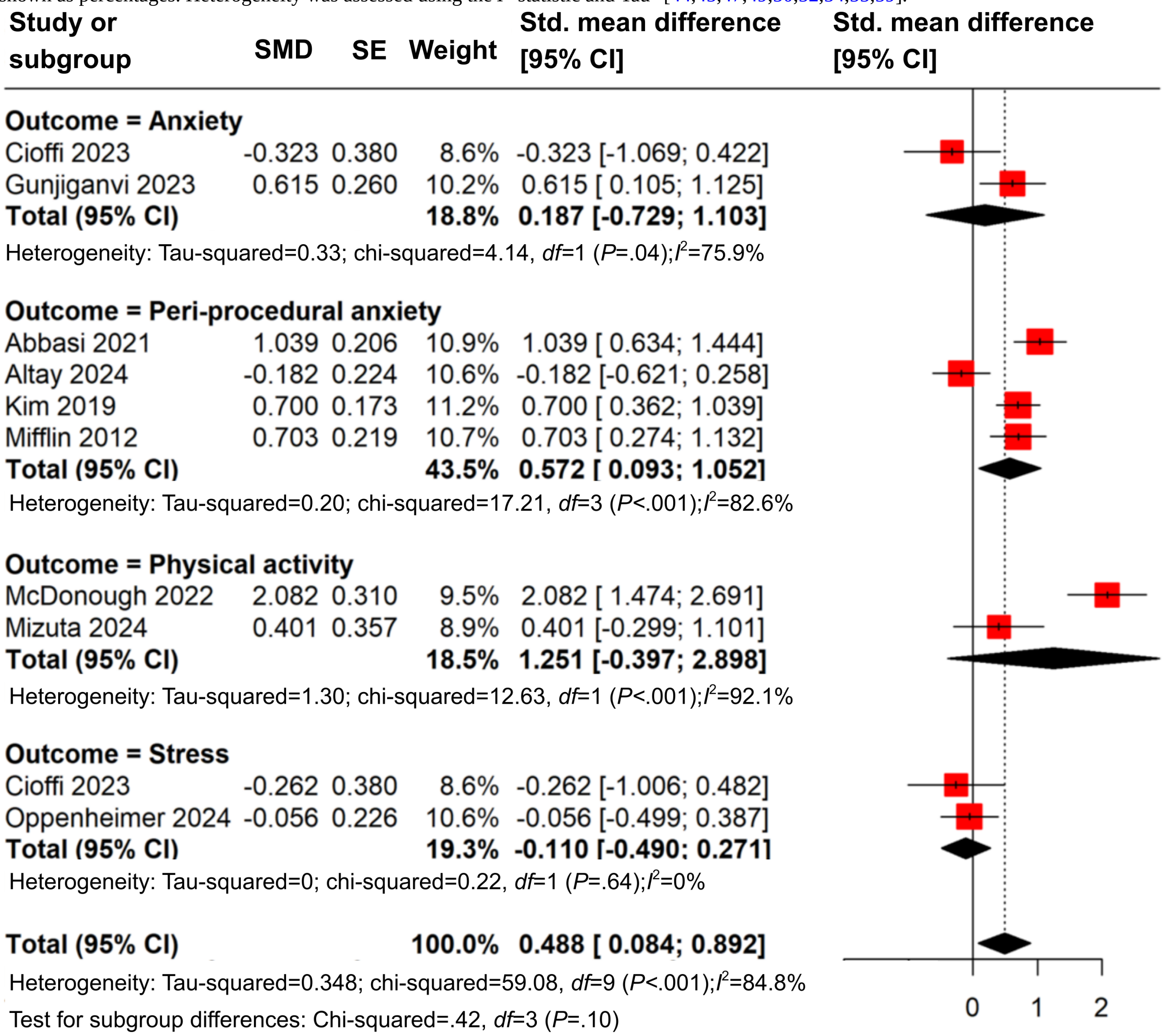
Of the two cross-sectional studies, 1 study demonstrated a positive correlation between those who regularly watch food videos and individuals' BMI, while the other reported effectiveness in YouTube-delivered mind-body interventions in improving sleep quality [57,58].

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].

Quantitative Synthesis

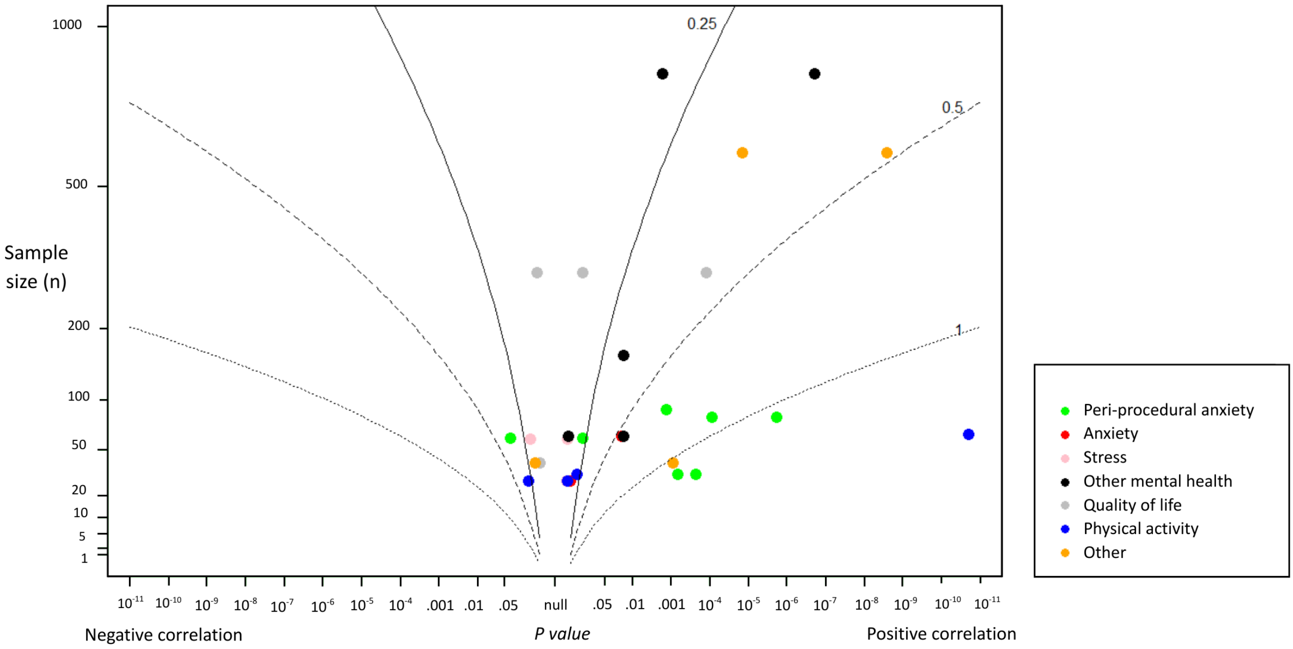
The primary outcomes and measures of the included studies were categorized based on outcome types, and the numerical values, calculated standard mean difference, and 2-tailed *P* values in the context of the studies' risk of bias are shown in Multimedia Appendix 4. While the included studies displayed 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].

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.

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 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]

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]

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]

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]

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

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

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

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.

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	

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 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).

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).

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)		

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)		

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)		

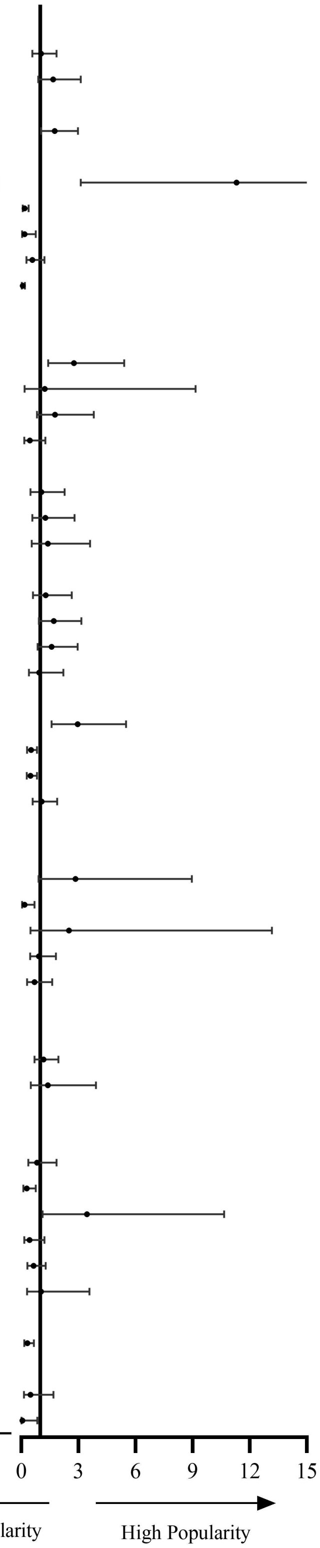
The Factors Associated With the Attractiveness of Alcohol Advertisements

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).

Figure 2. Multivariate logistic regression of popularity of alcohol-related videos. OR: odds ratio.

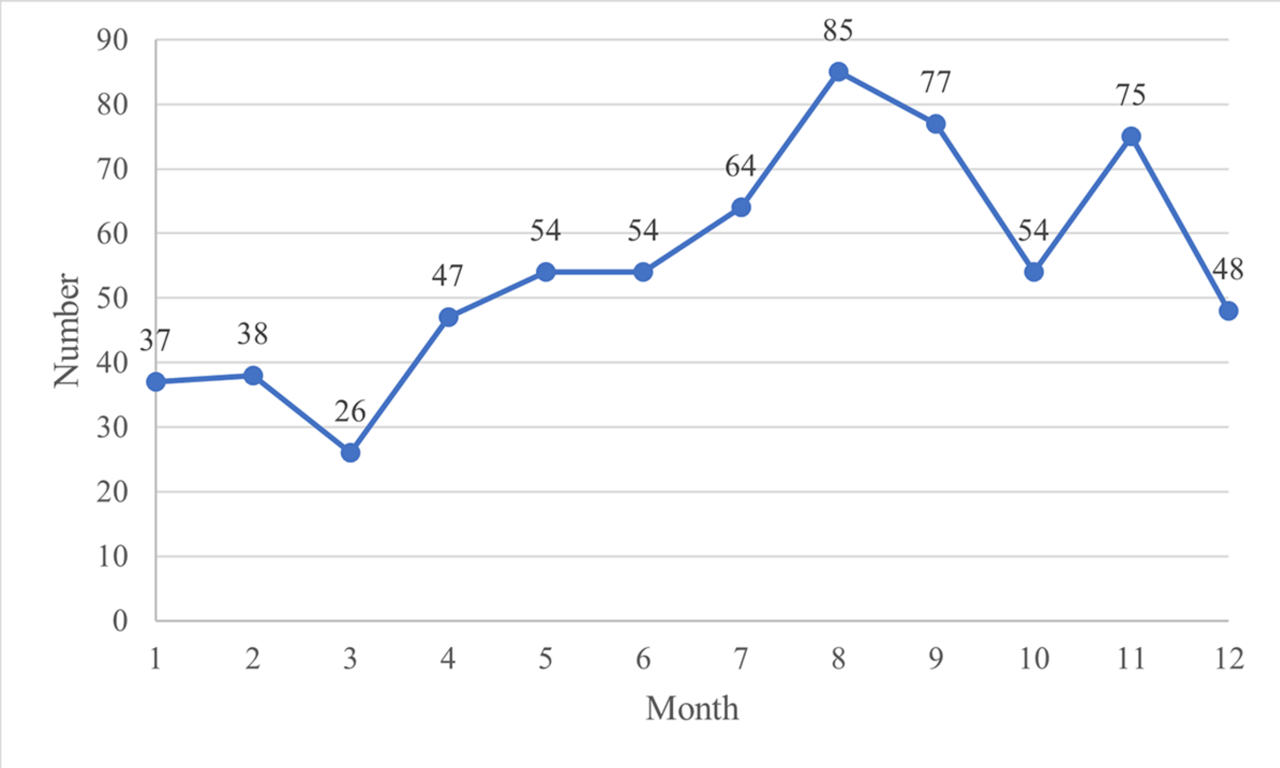
Dimension	Variable	Category	P	OR (95%CI)
Basic information				
	Duration	0s to 30s (Ref.)		
		31 to 60s	.91	1.035 (0.581-1.846)
		More than 60s	.11	1.670 (0.895-3.116)
	Brand category	Traditional (Ref.)		
		New	.04	1.759 (1.041-2.972)
	Alcohol category	Chinese liquor(Ref.)		
		Beer	<.001	11.310 (3.130-40.873)
		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).

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

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 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]

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/2024111213885f4bdbc1f915258ac52d.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]

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

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

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).

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 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 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

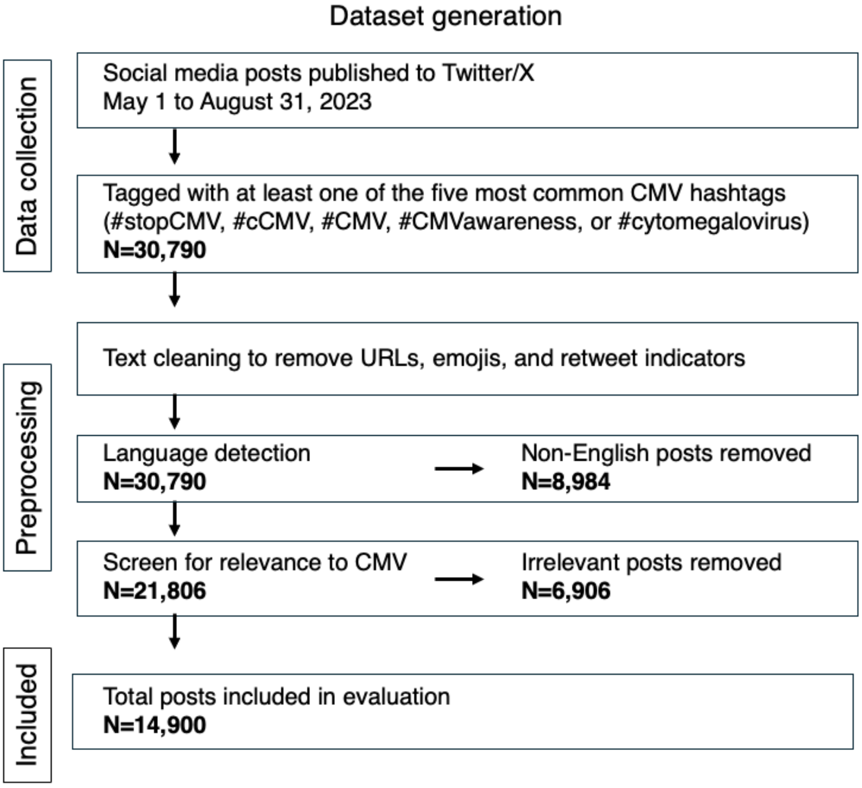
This study used only publicly available text from social media posts on Twitter/X. The research team did not interact with post authors or collect private information. All data were processed to remove or avoid inclusion of identifiable information and are reported only in aggregate. This study did not 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).

Figure 1. Flowchart of the dataset collection, preprocessing, and inclusion steps. cCMV: congenital cytomegalovirus; CMV: cytomegalovirus.



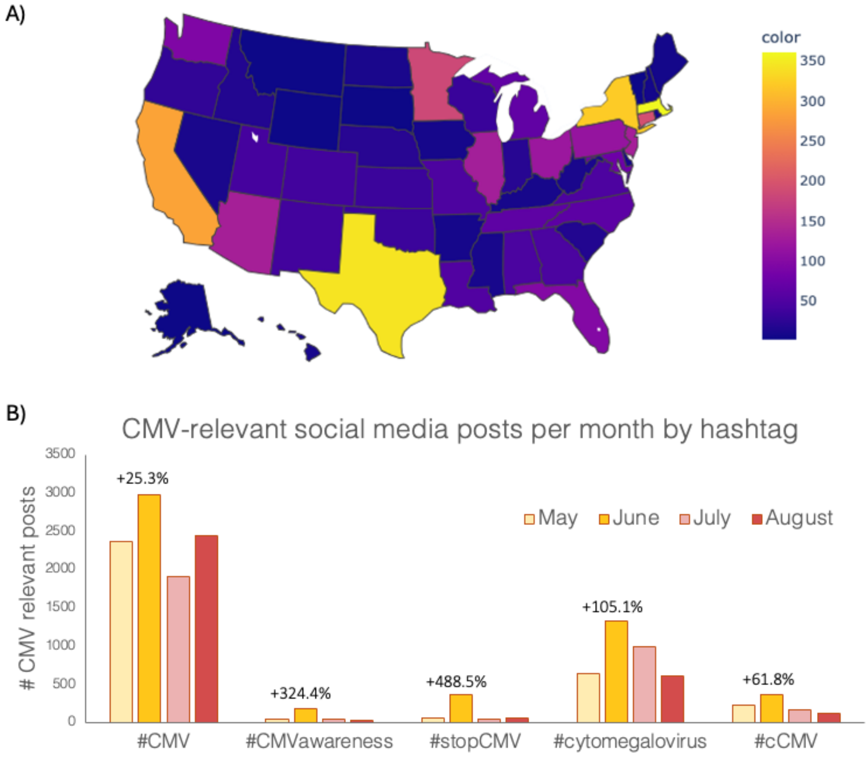
Descriptive Analyses

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).

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.

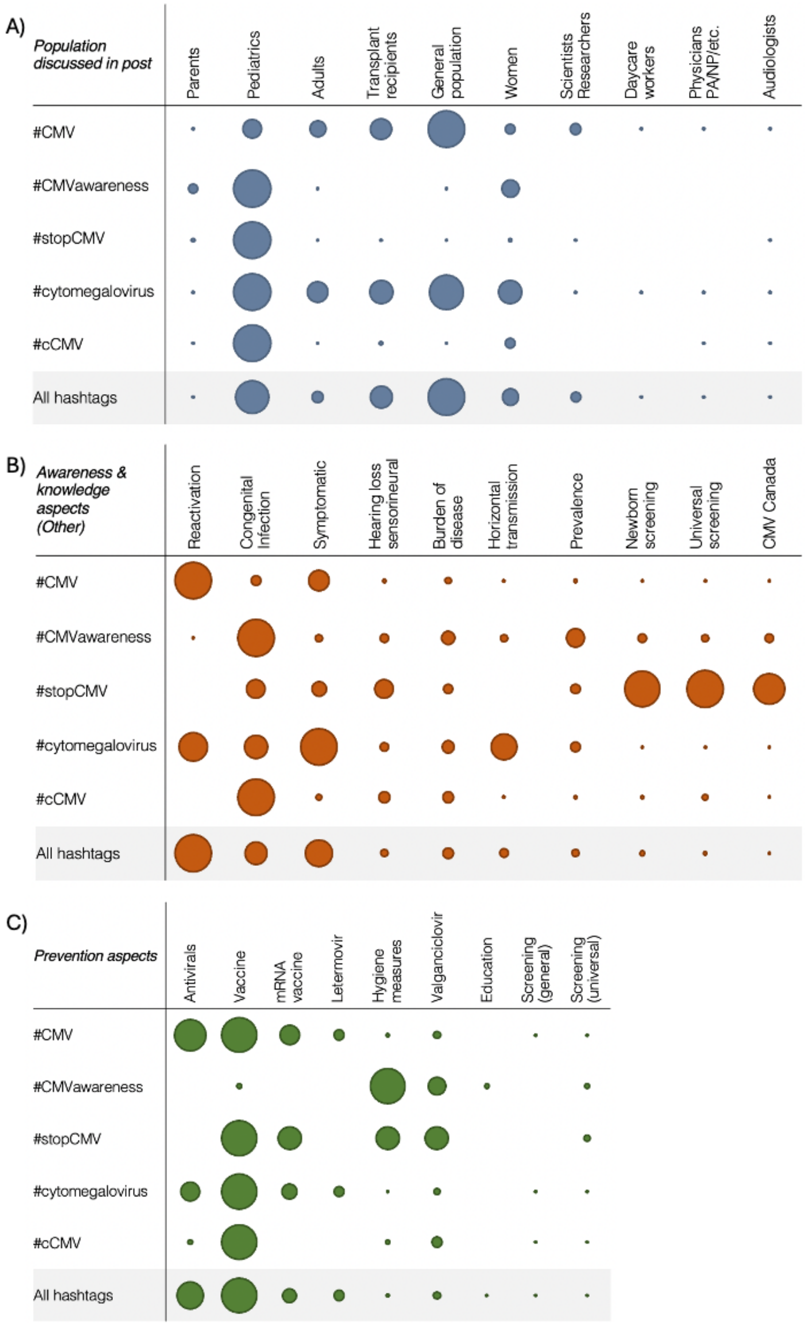


Classification of Thematic (Aspects) Content

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.

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

(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.

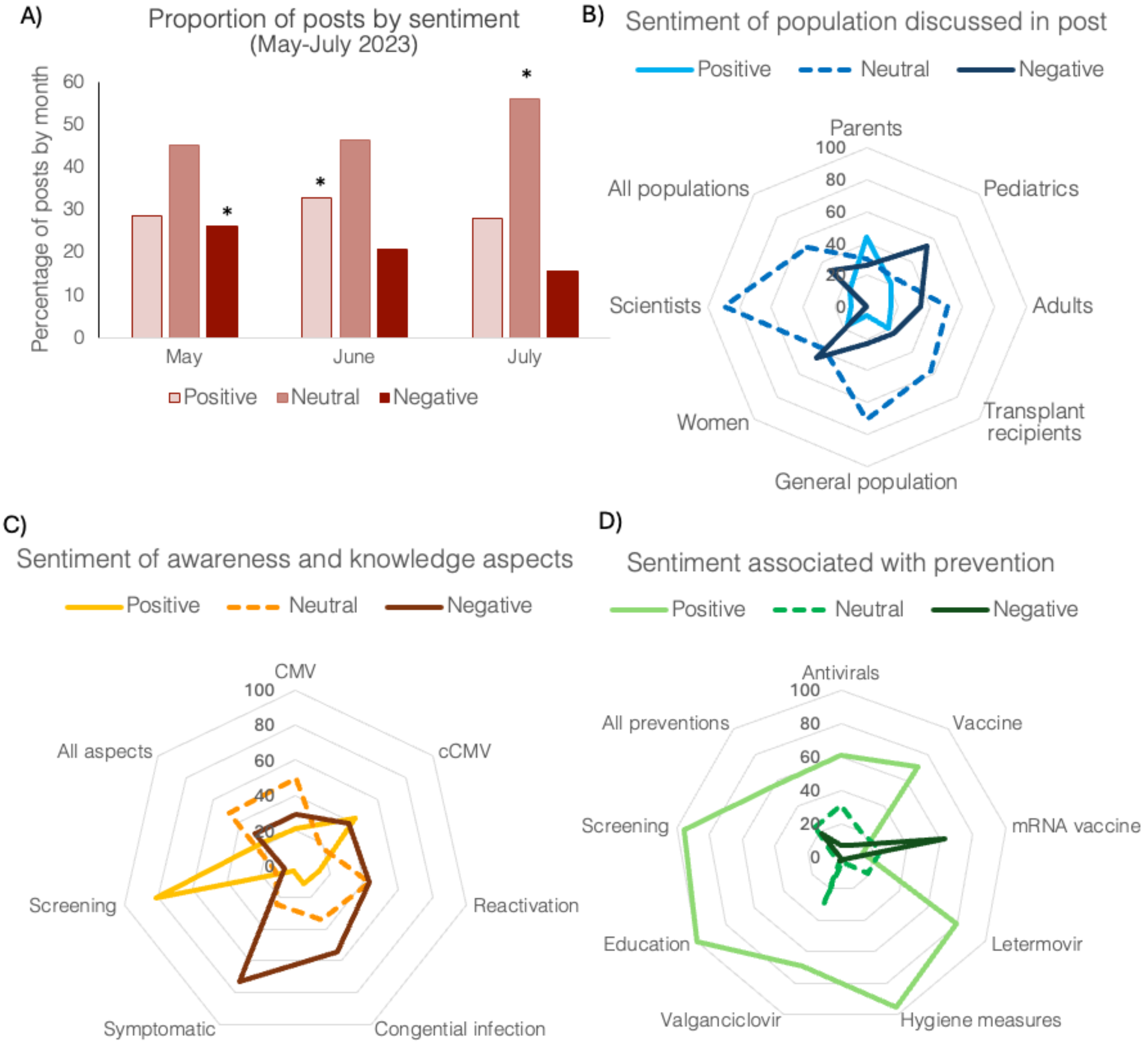
Sentiment Analyses

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.

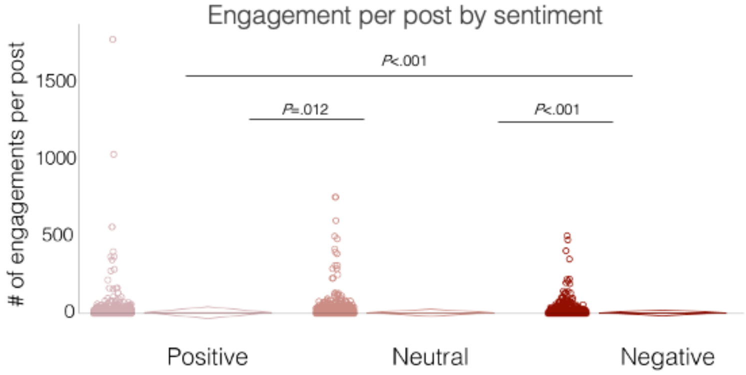
Figure 4. Evaluation of post sentiment from May to July 2023. (a) Proportion of posts scored by overall sentiment per month, (b) by population discussed, (c) by awareness and knowledge, and (d) by type of prevention. Post attributes that occurred infrequently (<10% of the total aspects) were not included. 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.

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.

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 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]

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]

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:

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 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.

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 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.

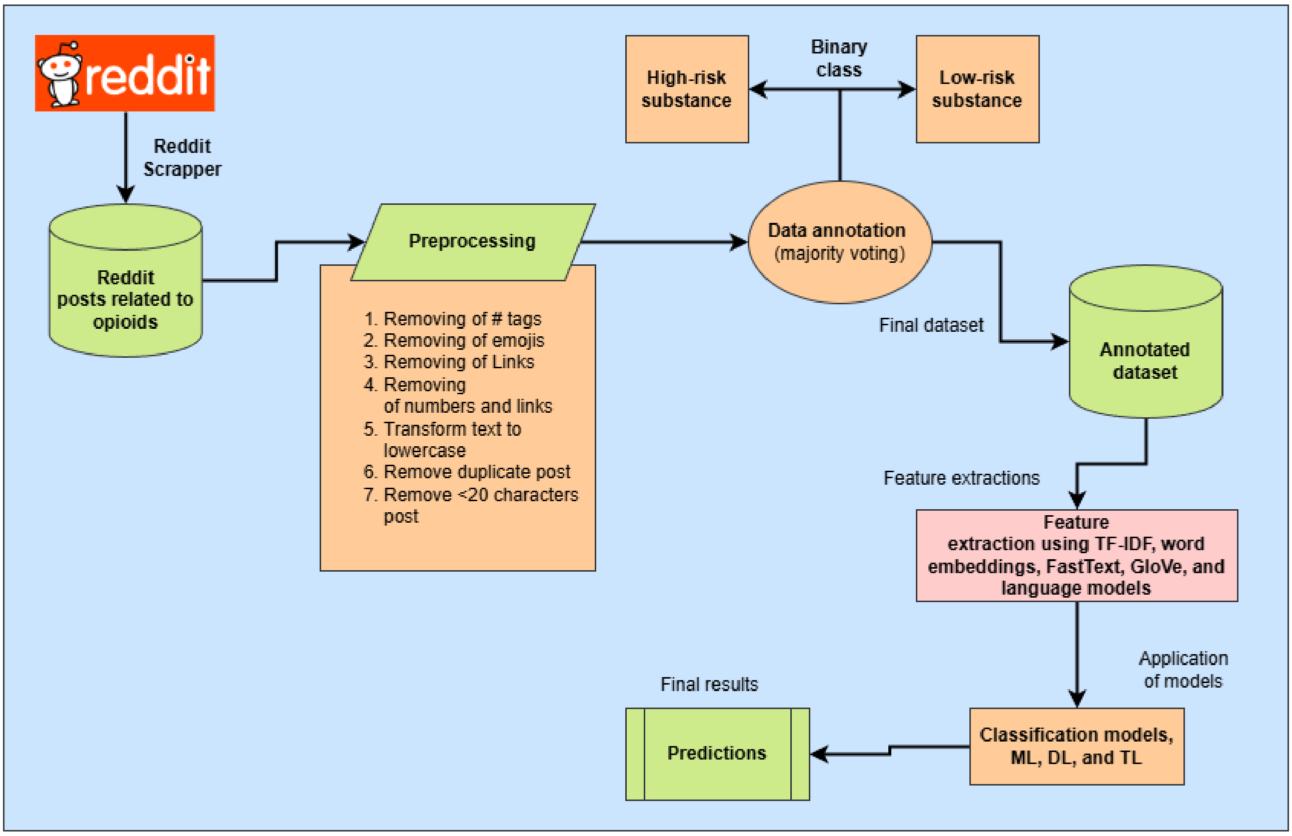
Table 1. Keywords used to extract the dataset.

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.

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.

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 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.

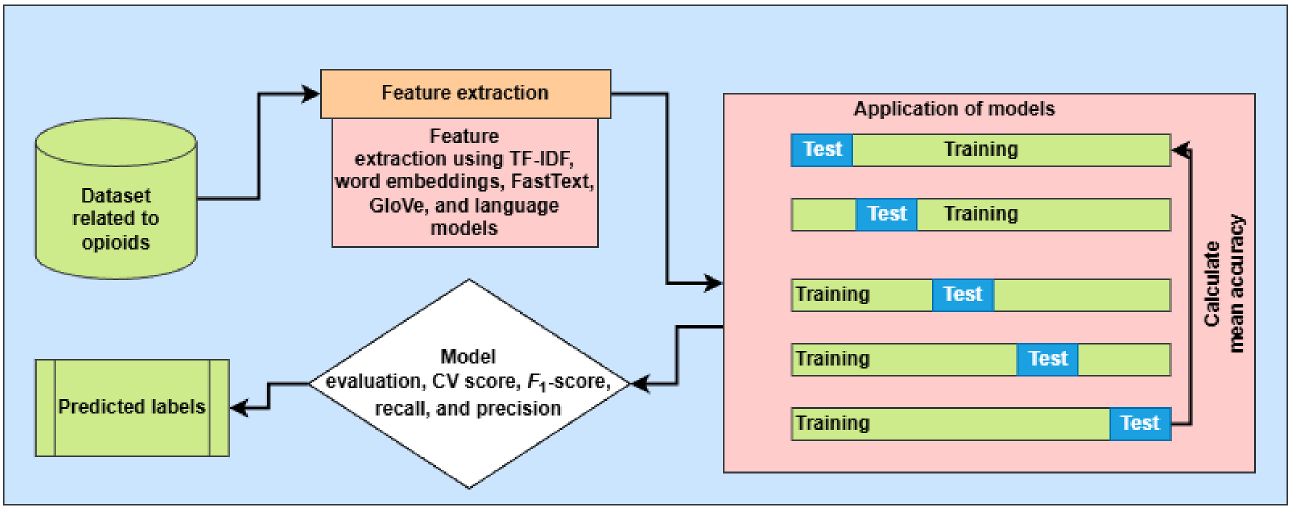
The FastText and GloVe techniques are selected for study due to their complementary strengths: FastText addresses linguistic variations and out-of-vocabulary words in Reddit discussions on opioid use, while GloVe captures broader semantic relationships across the corpus. These embeddings provide richer 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.

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 → 512 → 1) with a Tanh activation, introducing nonlinearity and enabling token-level focus.

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 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.

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 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

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.

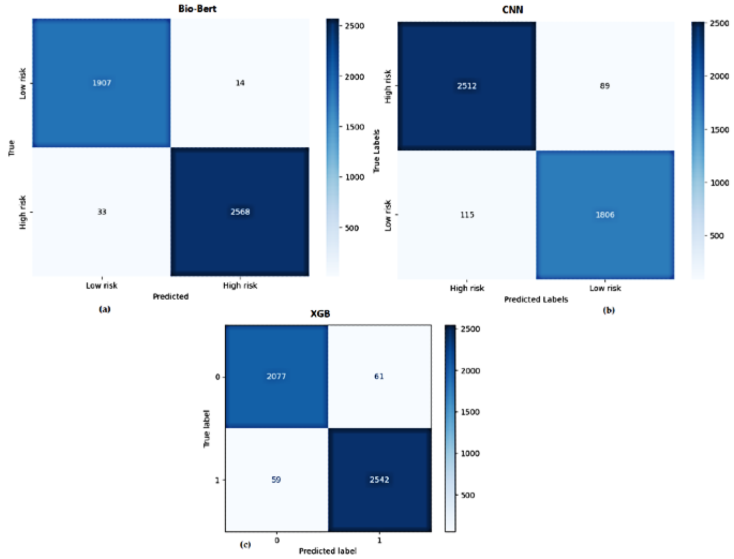


Figure 4. Confusion matrix of bidirectional encoder representations from transformers for biomedical text mining (BioBERT) using 2022-2023 train and 2024 test split.

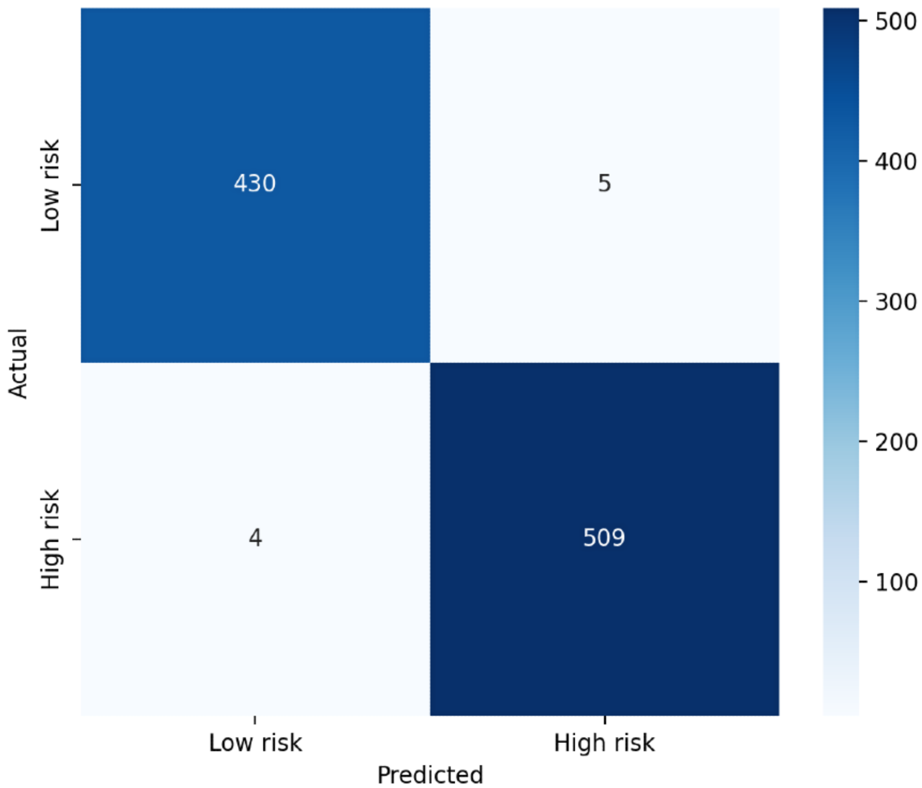
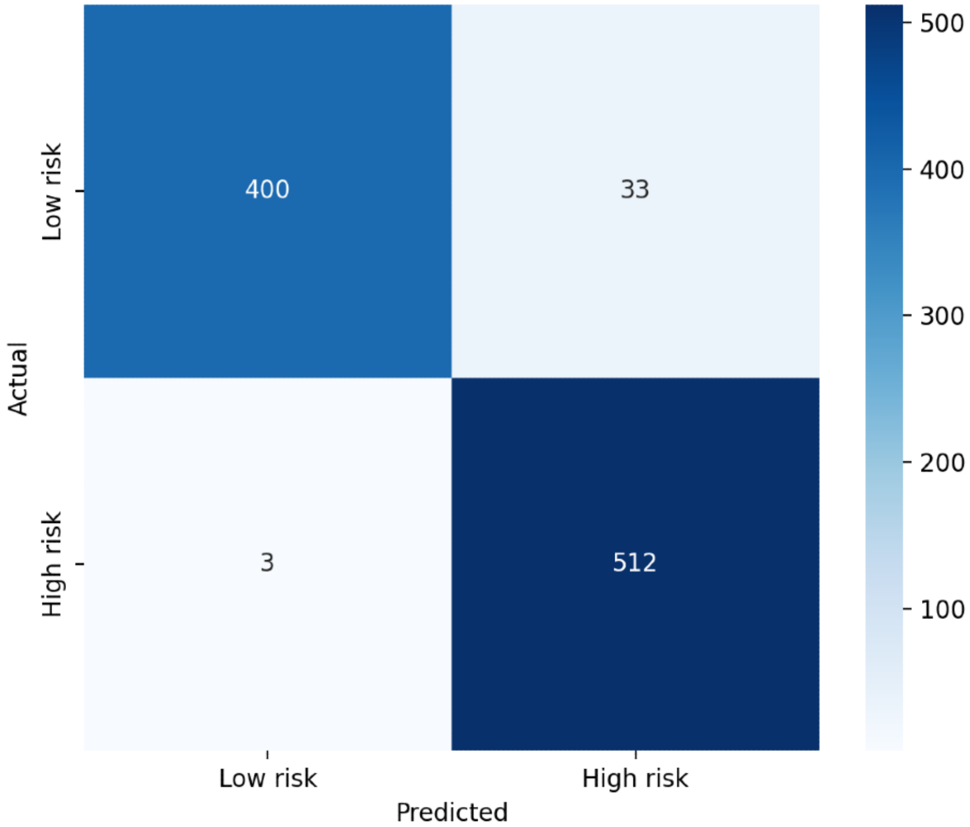


Figure 5. Confusion matrix of bidirectional encoder representations from transformers for biomedical text mining (BioBERT) using the traditional self-attention mechanism.



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 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.

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]

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]

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:

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

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 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?

- 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.

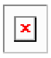
In the second step, the researchers identified 296 of the 438 posts as containing social support. These 296 posts were then classified into 1 of 4 types of social support—informational, emotional, instrumental, and network support—based on prior literature [12,22]. In addition, each post was coded according 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.

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

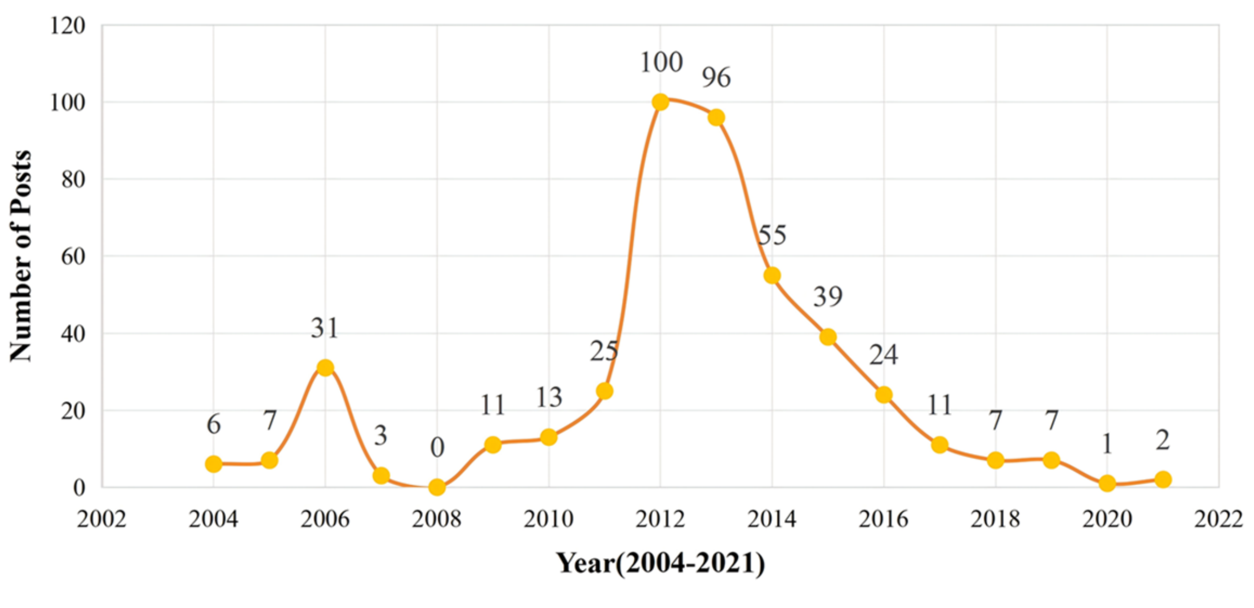
This study analyzes publicly available, anonymized data obtained from an online platform. No private, personally identifiable, or sensitive information was accessed or collected. All data were deidentified before analysis. The research did not involve direct interaction with human participants. In accordance 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.

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).

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.

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.

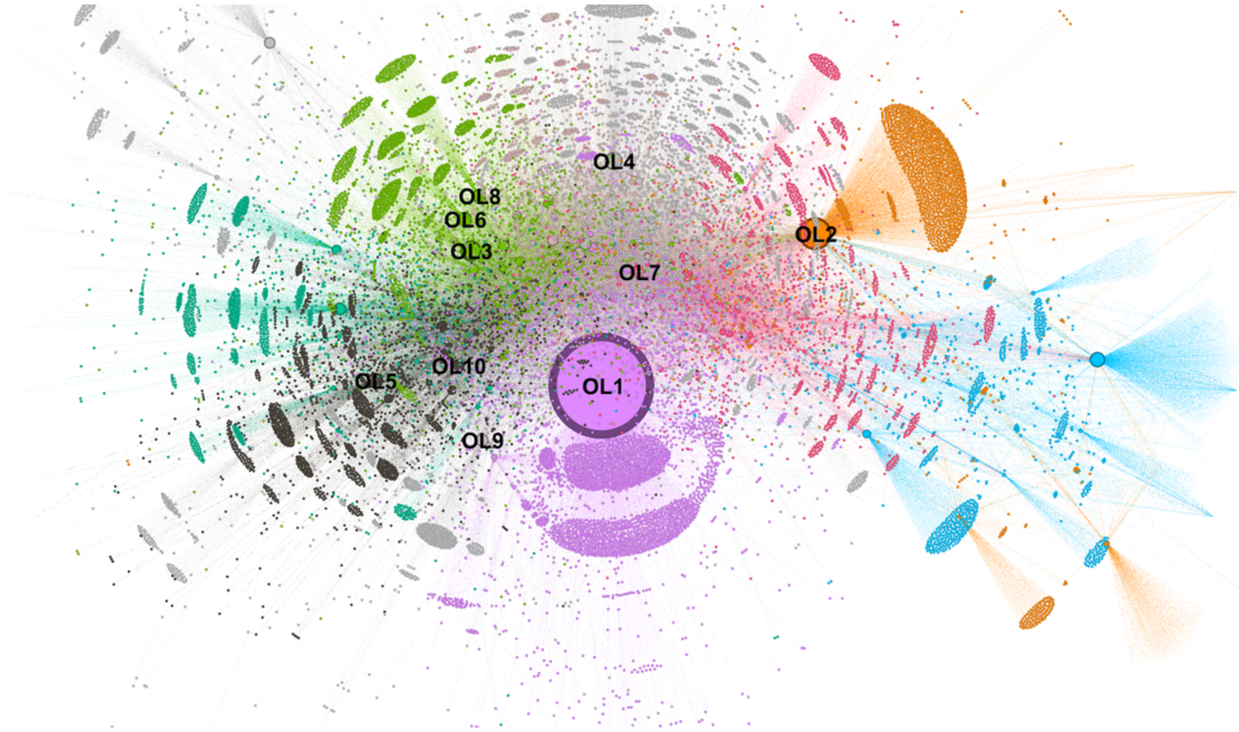


Table 3 presents the top 10 OLs identified using the Borda ranking method, which aggregates scores across 4 centrality metrics and a user-tenure metric.

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.

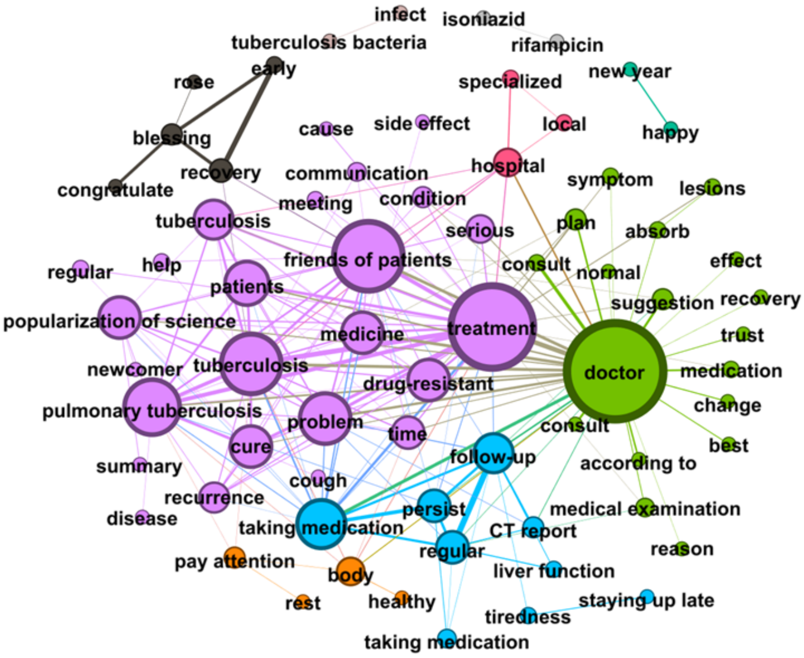
The first-ranked OL served as the administrator of the TB forum at the time of data collection and is a former patient with TB who had completed treatment and fully recovered. The second-ranked OL, a medical doctor specializing in pulmonary TB, is the founder and initial administrator of the TB forum.

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.

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.

Figure 4. Semantic networks of opinion leader 2 (OL2) posts and associated user replies.

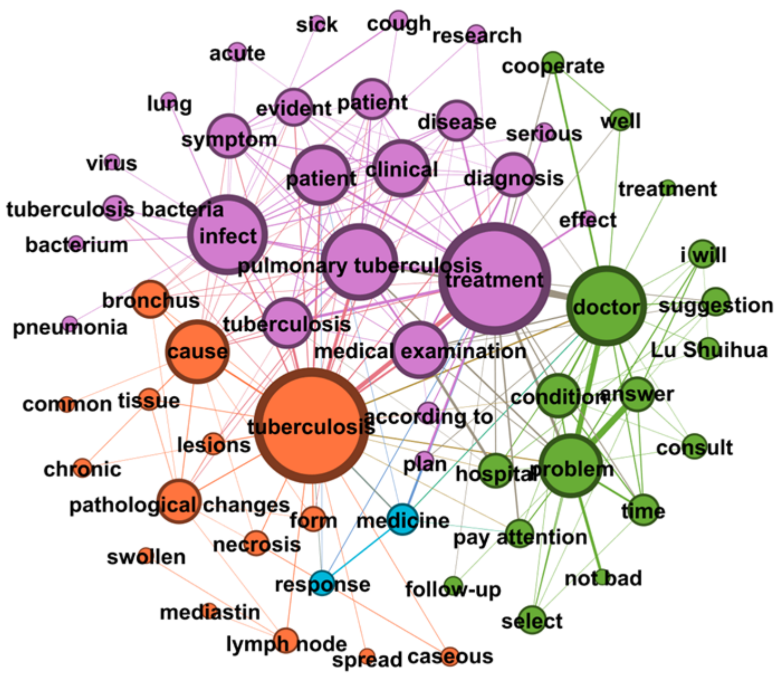


Figure 5. Semantic networks of opinion leader 3 (OL3) posts and associated user replies.

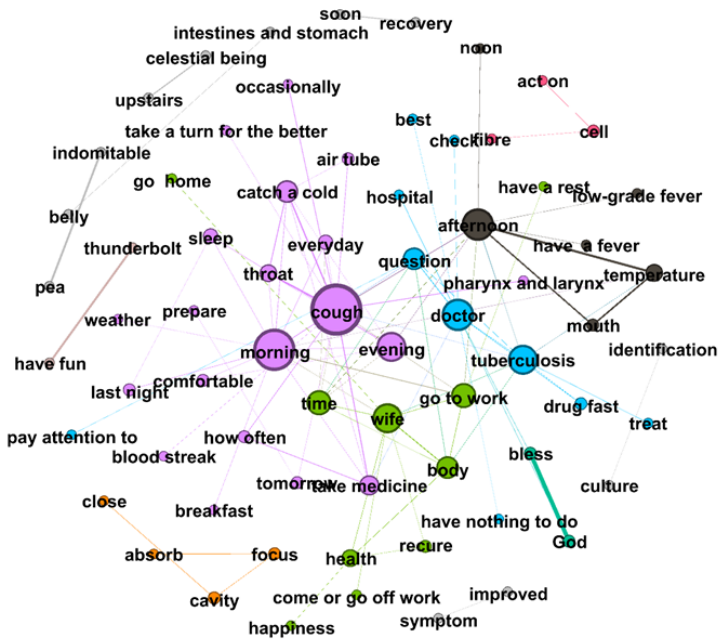


Figure 6. Semantic networks of opinion leader 4 (OL4) posts and associated user replies.

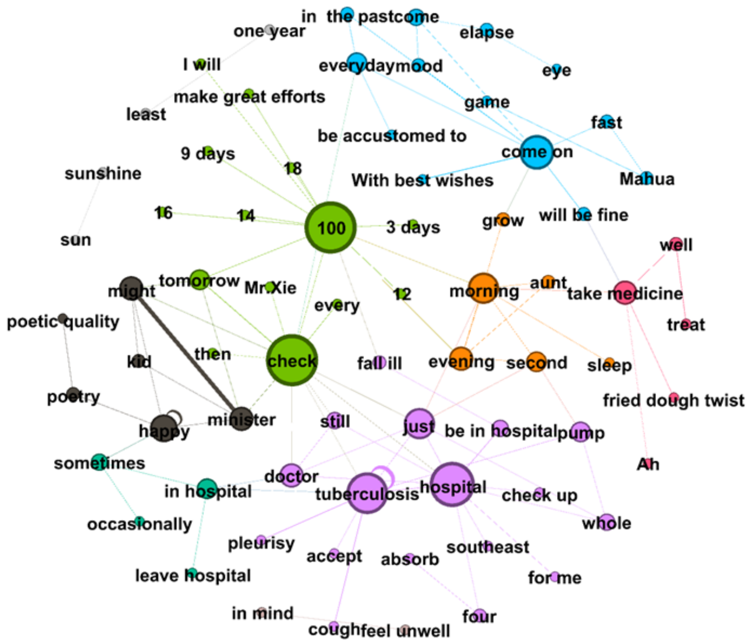


Figure 7. Semantic networks of opinion leader 5 (OL5) posts and associated user replies.

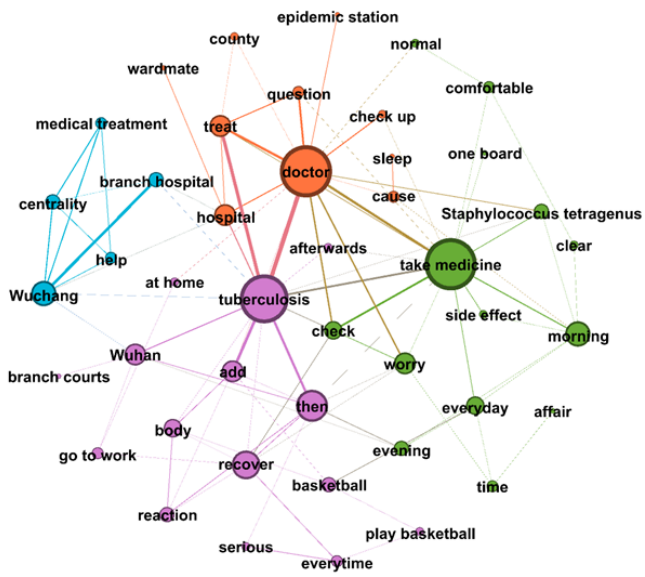


Figure 8. Semantic networks of opinion leader 6 (OL6) posts and associated user replies.

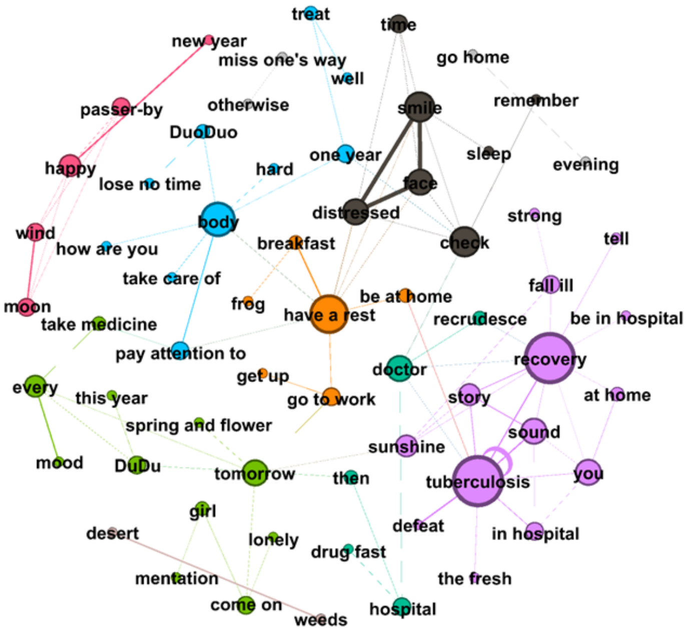


Figure 9. Semantic networks of opinion leader 7 (OL7) posts and associated user replies.

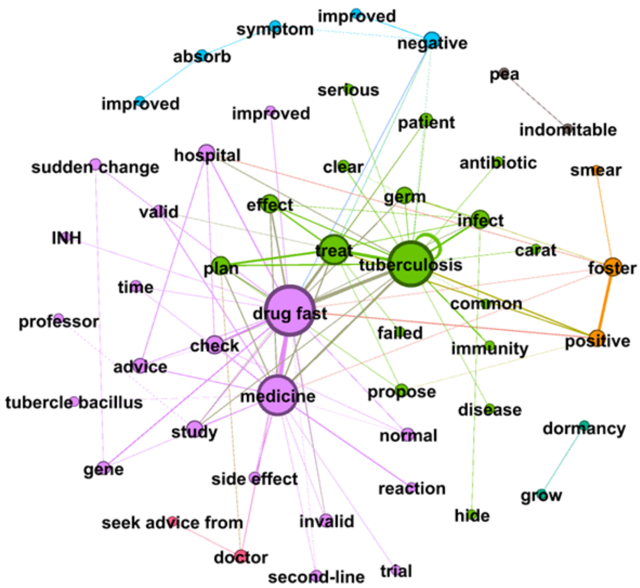


Figure 10. Semantic networks of opinion leader 8 (OL8) posts and associated user replies.

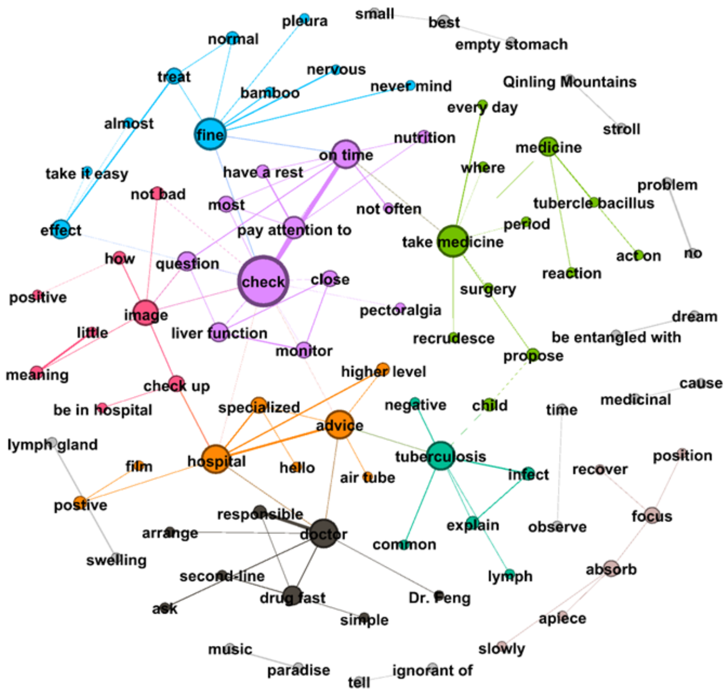
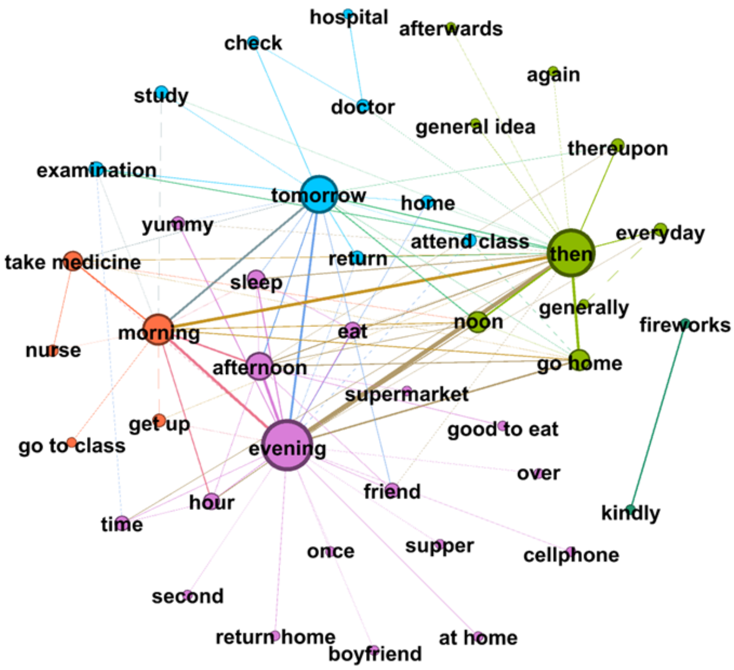


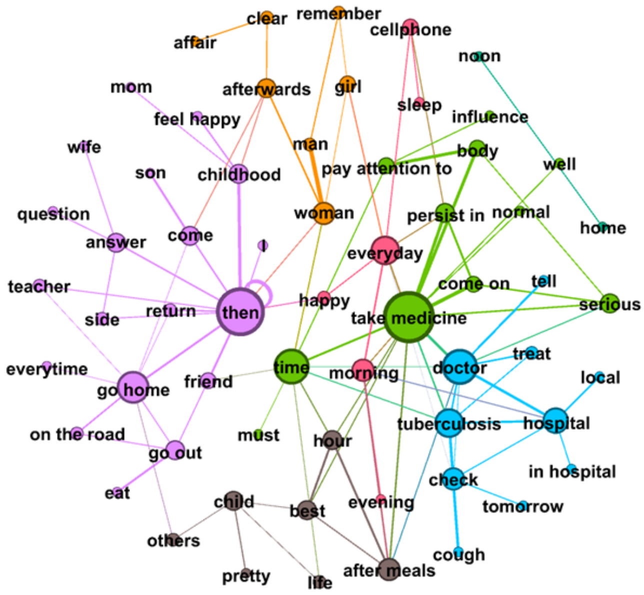
Figure 11. Semantic networks of opinion leader 9 (OL9) posts and associated user replies.



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.

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.

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 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]

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.

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

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.

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 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.

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.

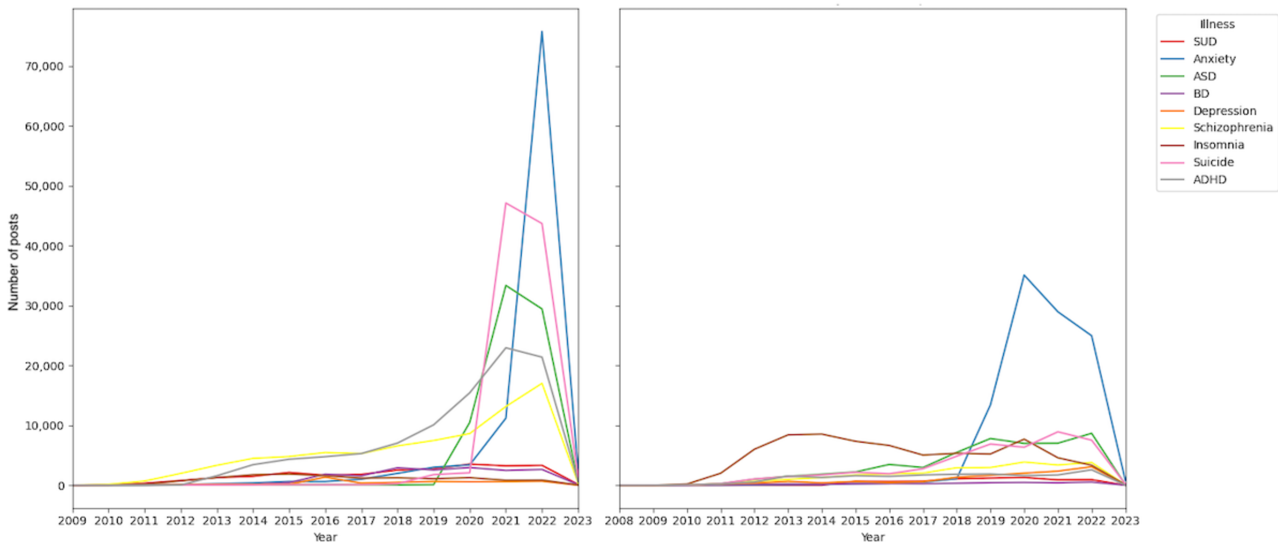
Results

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

(*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.

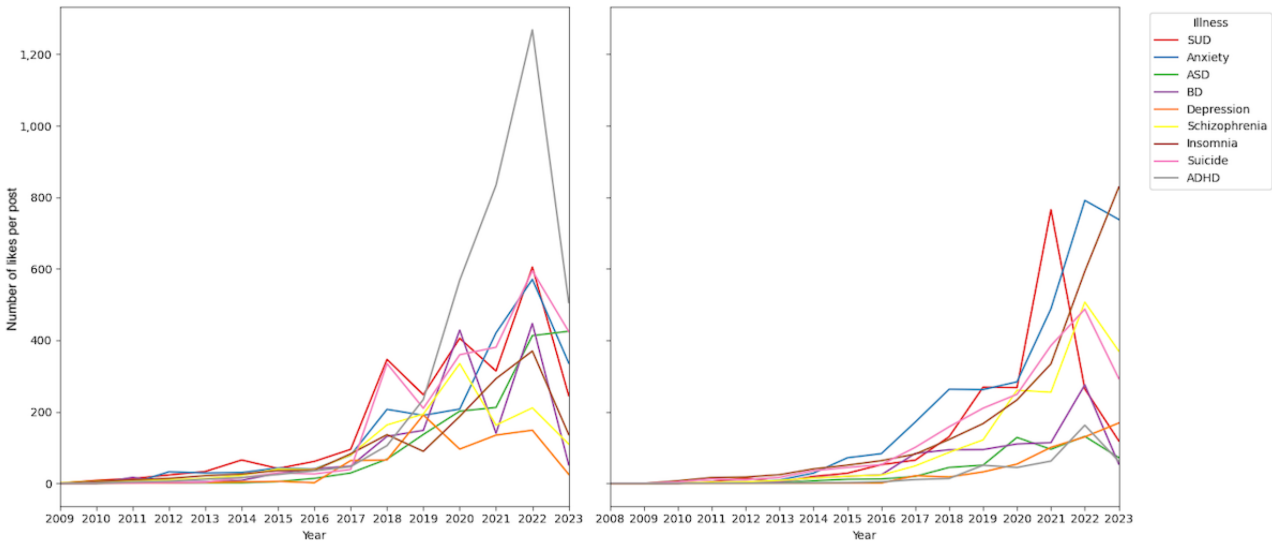
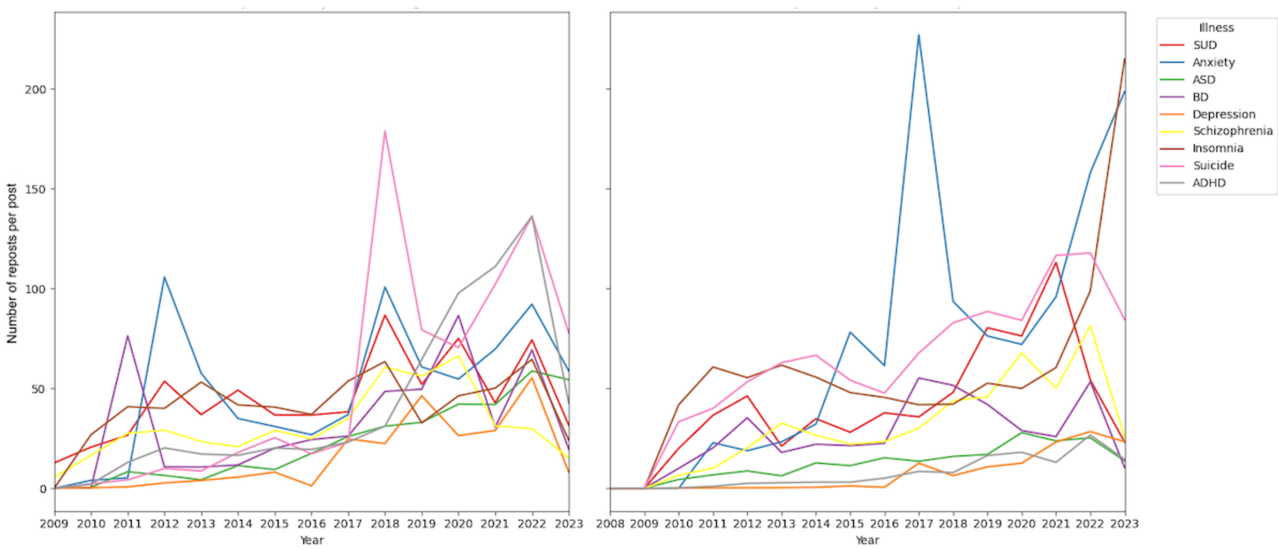


Figure 3. Number of reposts per post over time for various mental health conditions in both English (left) and Spanish (right). ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; BD: bipolar disorder; SUD: substance use disorder.



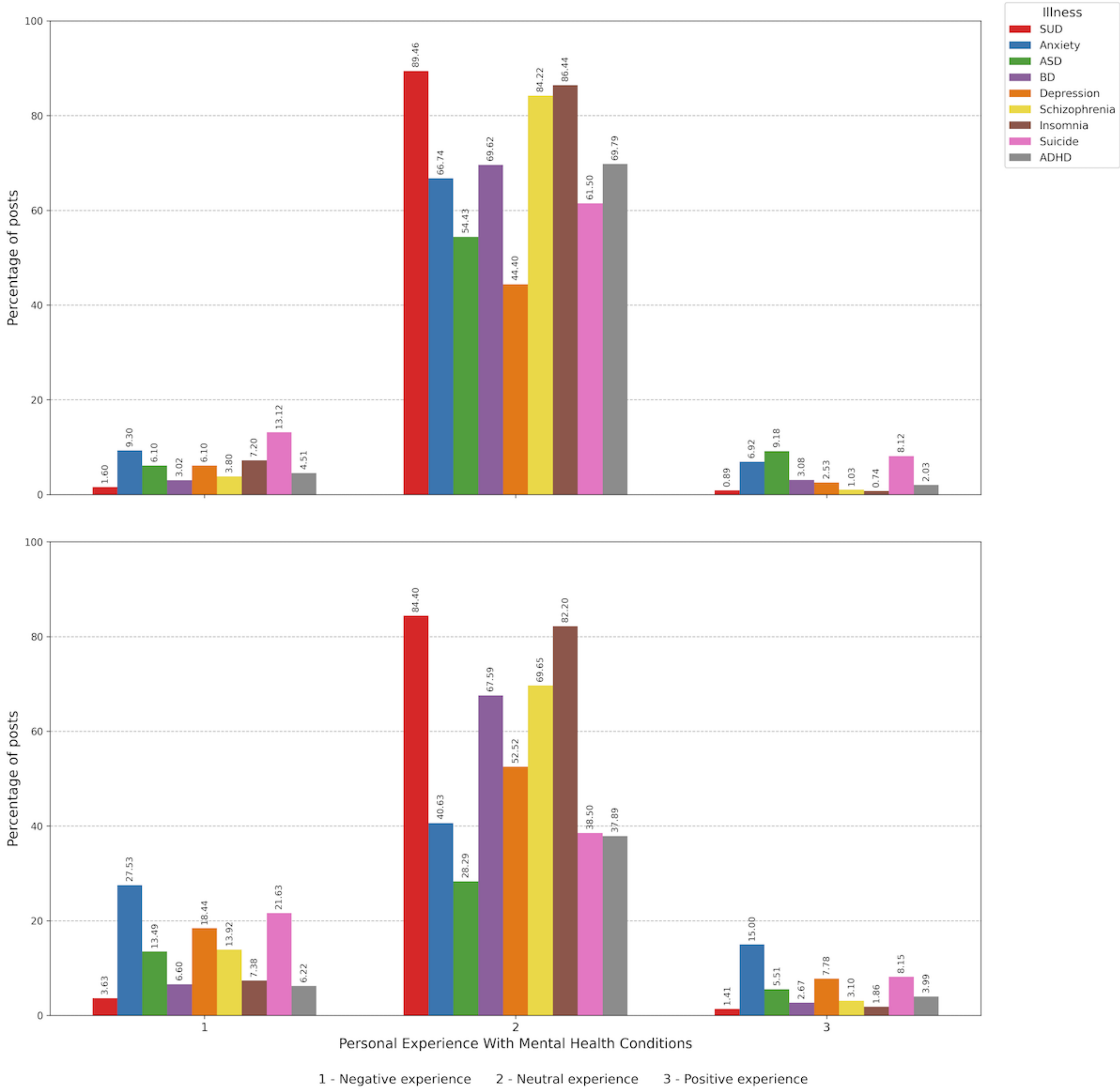
Types of Personal Experiences in Posts About ADHD and ASD Compared to Other Mental Illnesses

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.

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.

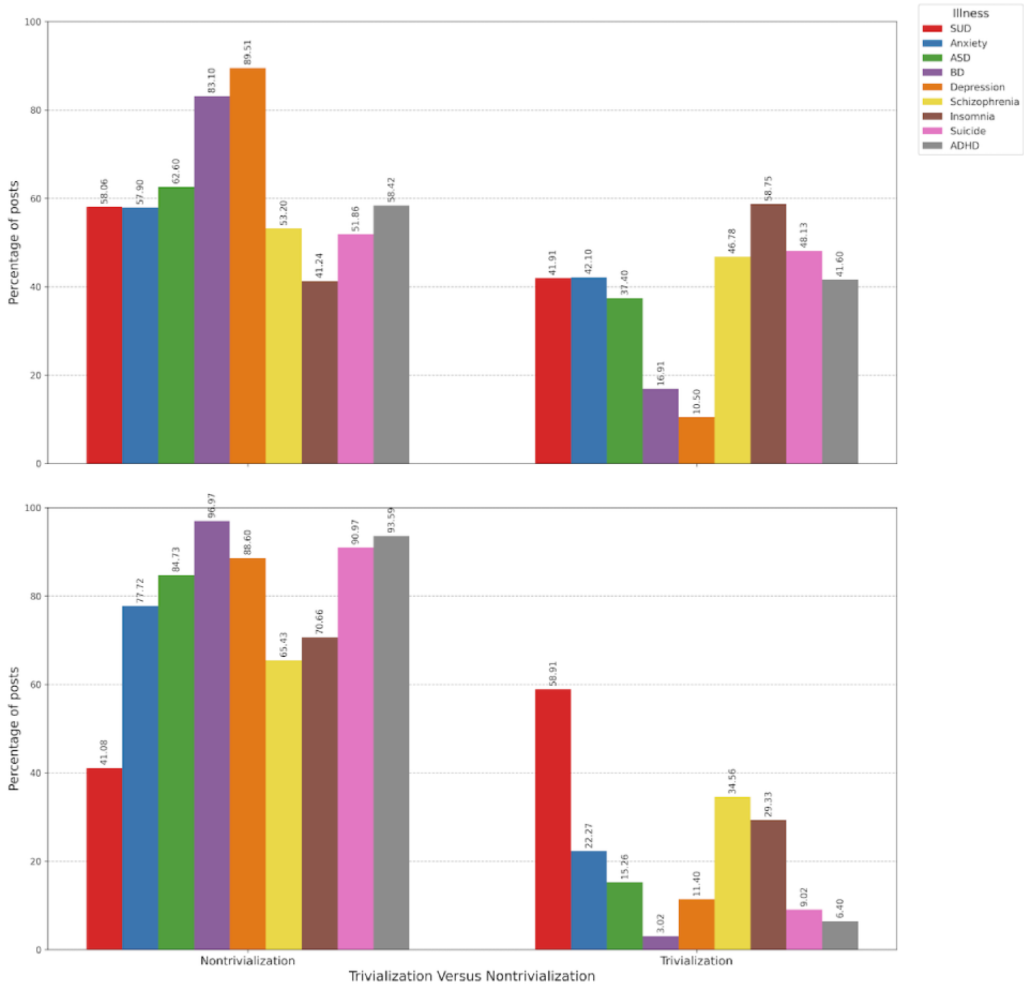


Trivialization of ADHD and ASD in Posts Compared to Other Mental Illnesses

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).

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.

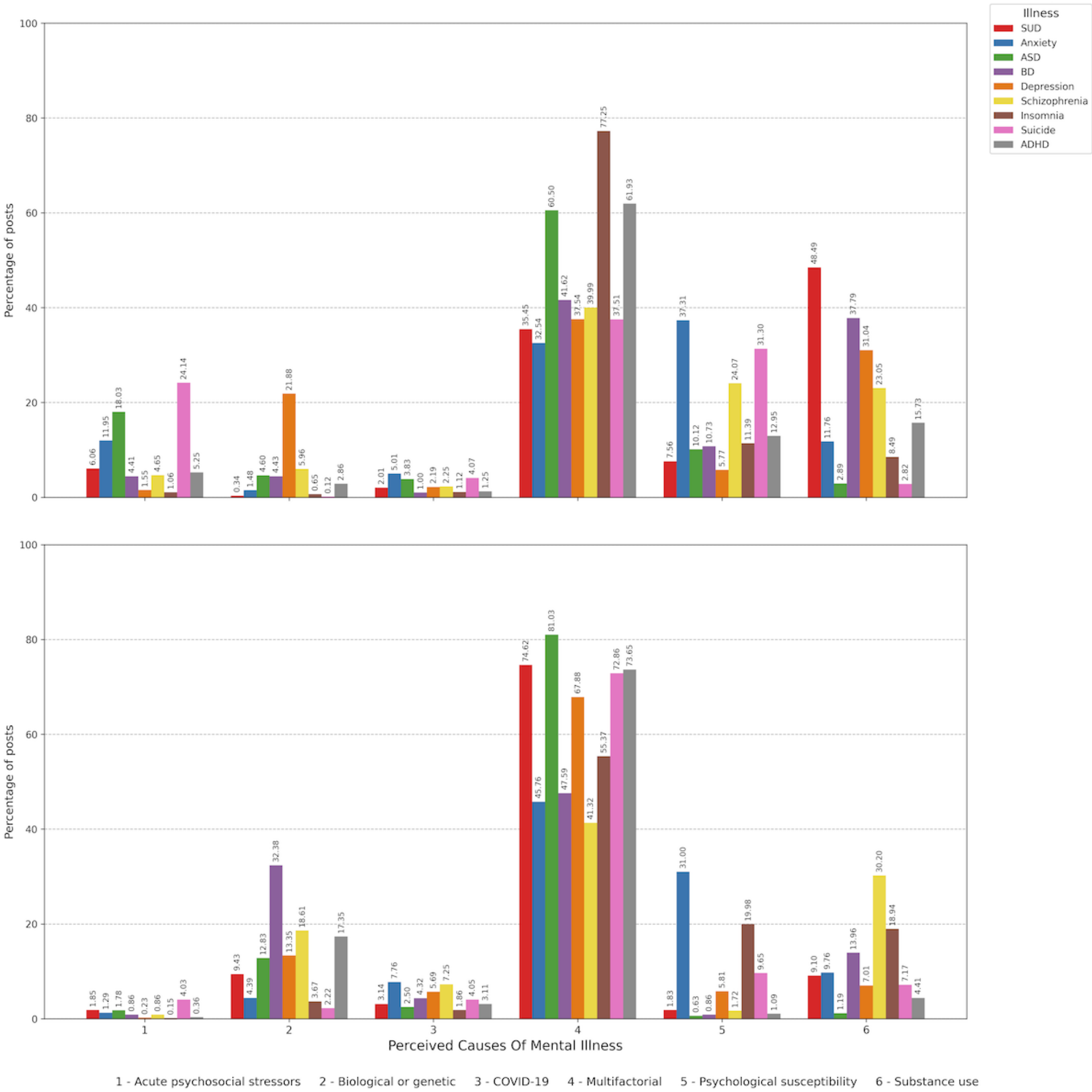


User-Perceived Causes of Mental Illness Analysis

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 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.

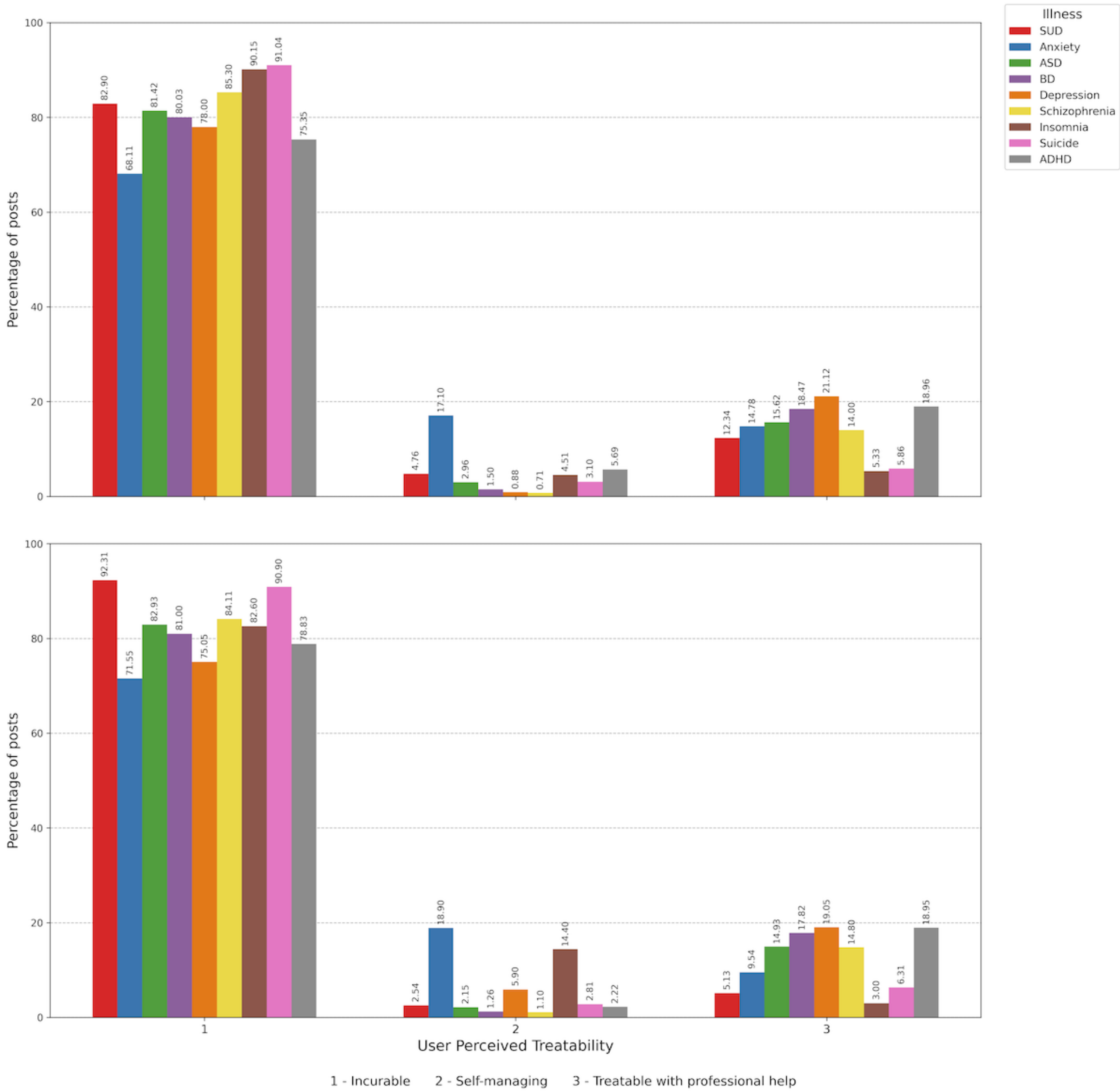


Percentage of User-Perceived Treatability of Mental Illnesses in Posts About ADHD and ASD vs Other Mental Illnesses

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).

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 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 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]

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]

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]

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:

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 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 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 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.

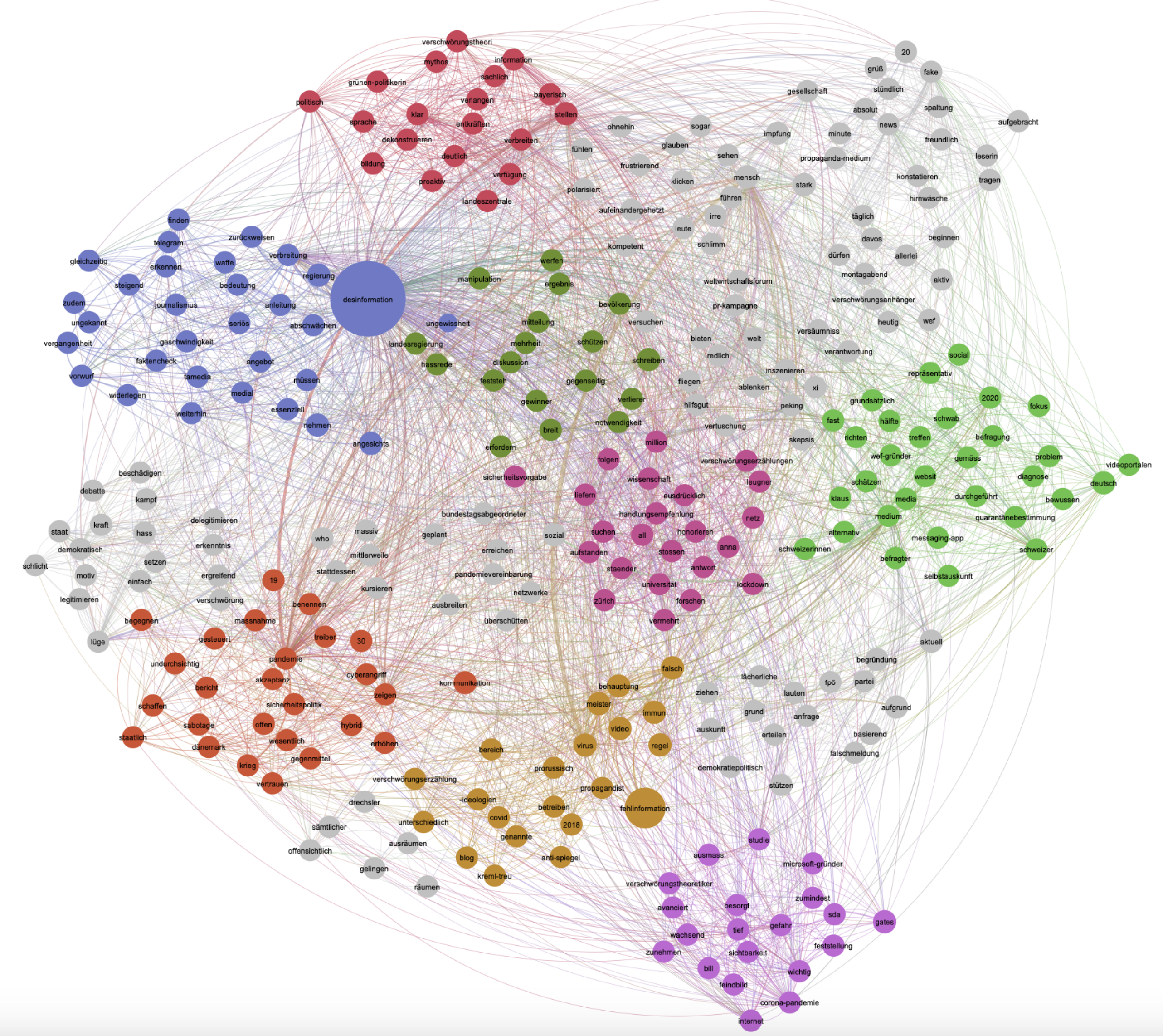
> *Der immer wieder erhobene Vorwurf, Medien würden im Zusammenhang mit der Corona-Pandemie «Panikmache» betreiben, wird durch die Studie entkräftet. Während der ersten Welle, als die Situation neu und vieles noch unbekannt war, berichteten Journalistinnen und Journalisten in 16 Prozent der untersuchten 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]

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”*).

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.*

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" *(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*

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

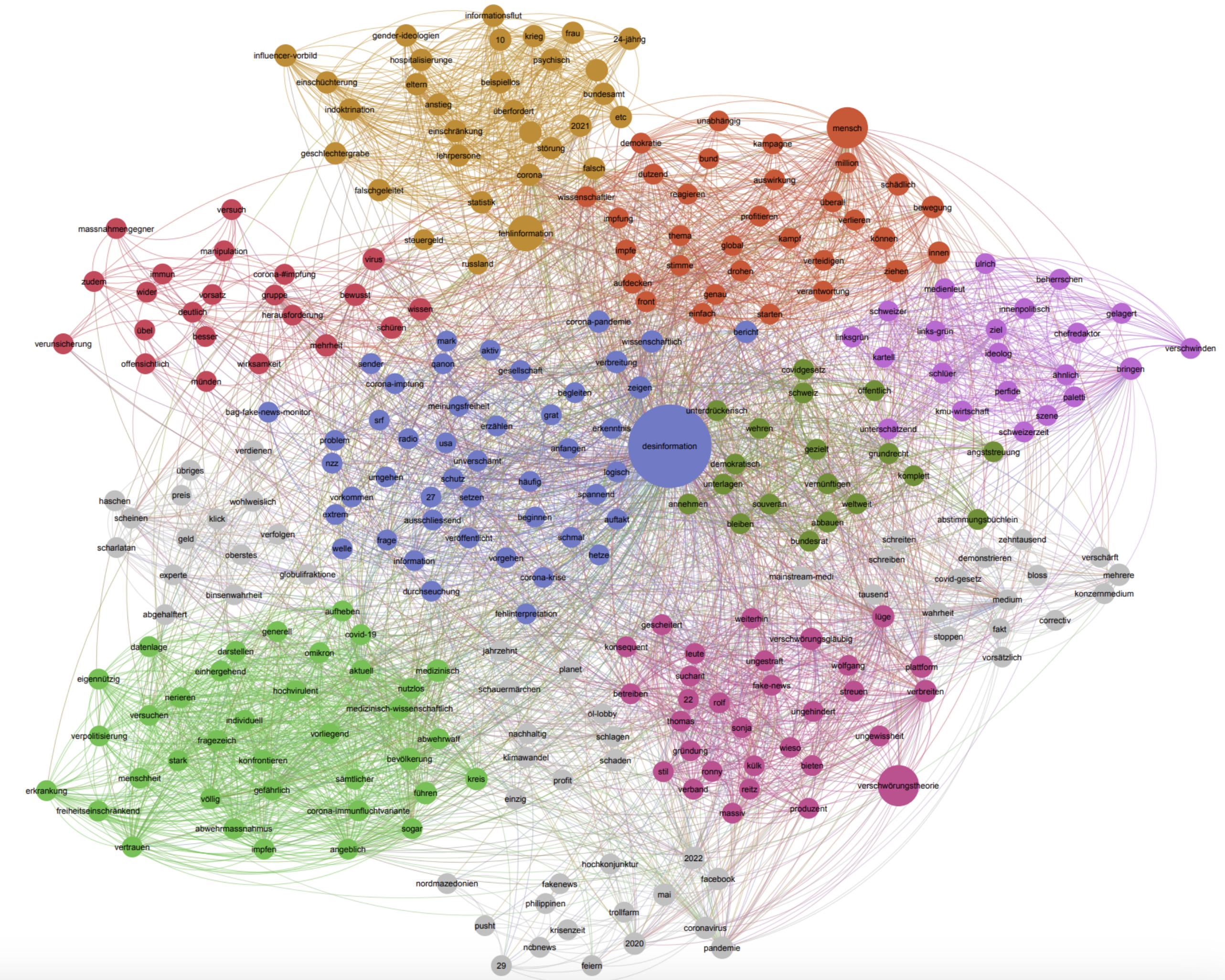
> *or clicks. My acquaintance cannot distinguish between a MIT or ETH study and a survey by a Füdli institute, she cannot classify scientific research, like many in our society. […] And it is the task of the state to communicate this information unequivocally. Also in schools. And sorry, most of them 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]

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”*).

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*

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]

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 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]

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

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

Quality, Reliability, and Dissemination of In Vitro Fertilization–Related Videos on Chinese Social Media: Cross-Sectional Analysis of 300 Short Videos

Dapeng Chu[1*], PhD; Xueyan Bai[1*], MD; Feng Guo[2*], BA

[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

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 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 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 coefficient was used to assess relationships between variables. Statistical significance was defined as P<.05.

Ethical Considerations

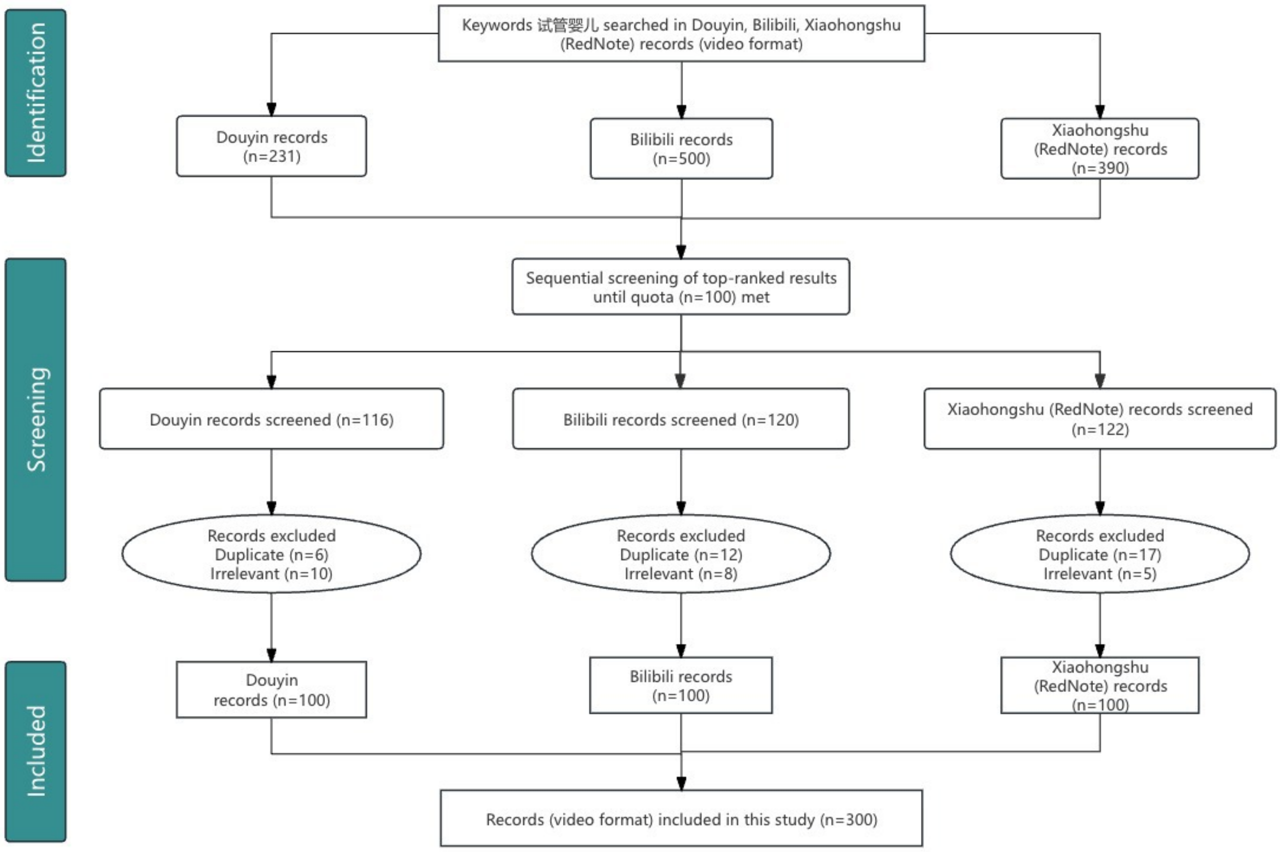
This study analyzed publicly available, user-generated IVF-related videos and their metadata from Chinese social media platforms. The research involved no interaction or intervention with individuals and did not collect, store, or process identifiable personal data. In accordance with institutional and 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).

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.

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.

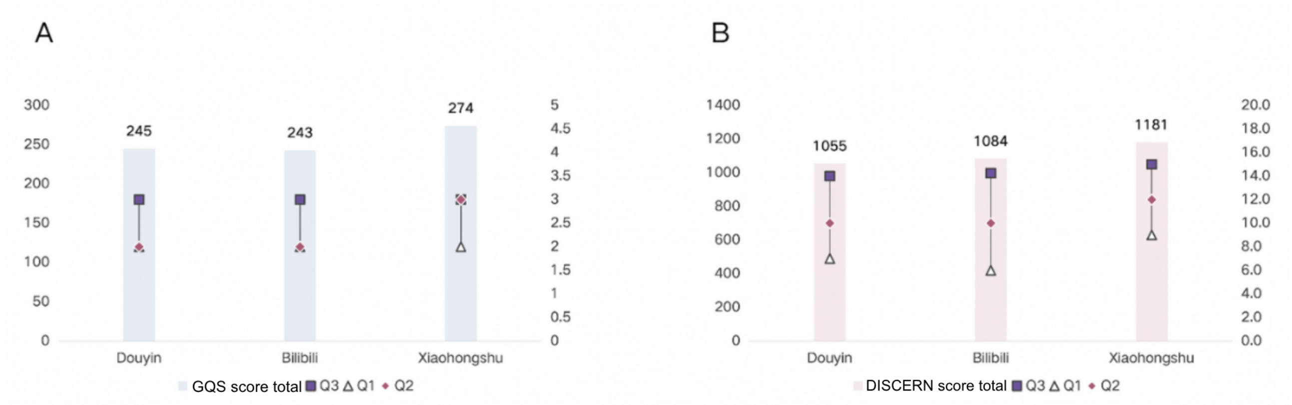
Video attributes also differed significantly. Bilibili hosted older (median 788.5, IQR 293.3-1300.8 d) and longer videos (median 325, IQR 169-692 s), whereas Douyin featured more recent, shorter-form content (median 41.5, IQR 25.8-59 s; *P*<.001 for both). Douyin creators had the largest median follower 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).

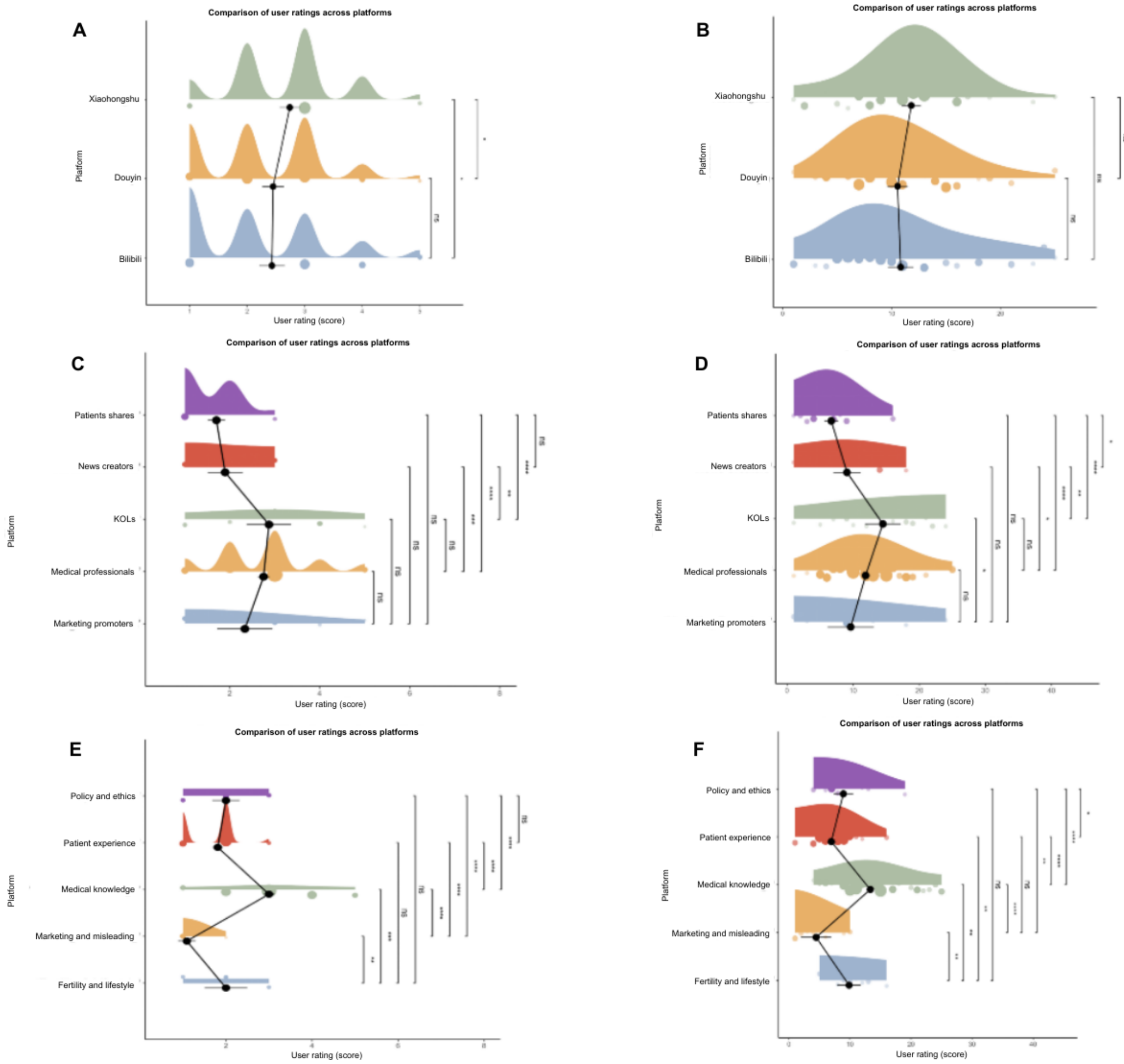
Platform-level analysis showed that Xiaohongshu videos achieved a statistically significant higher quality, with a median GQS of 3.0 (IQR 2.0-3.0; *P*=.04) and a median mDISCERN score of 12.0 (IQR 9.0-15.0; *P*=.02) compared to Douyin and Bilibili (Figure 2).

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).

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.

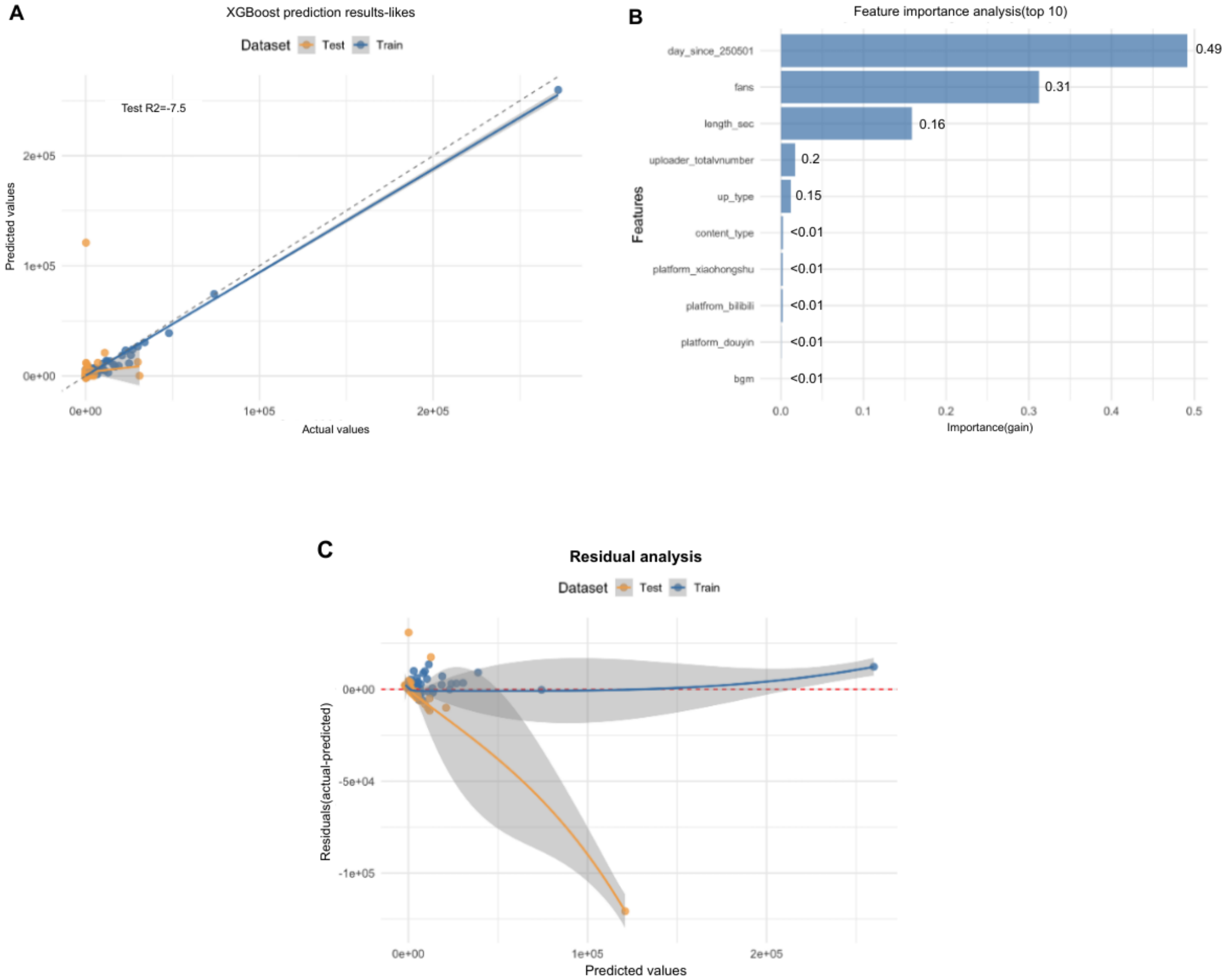


Predictors of Video Dissemination

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.

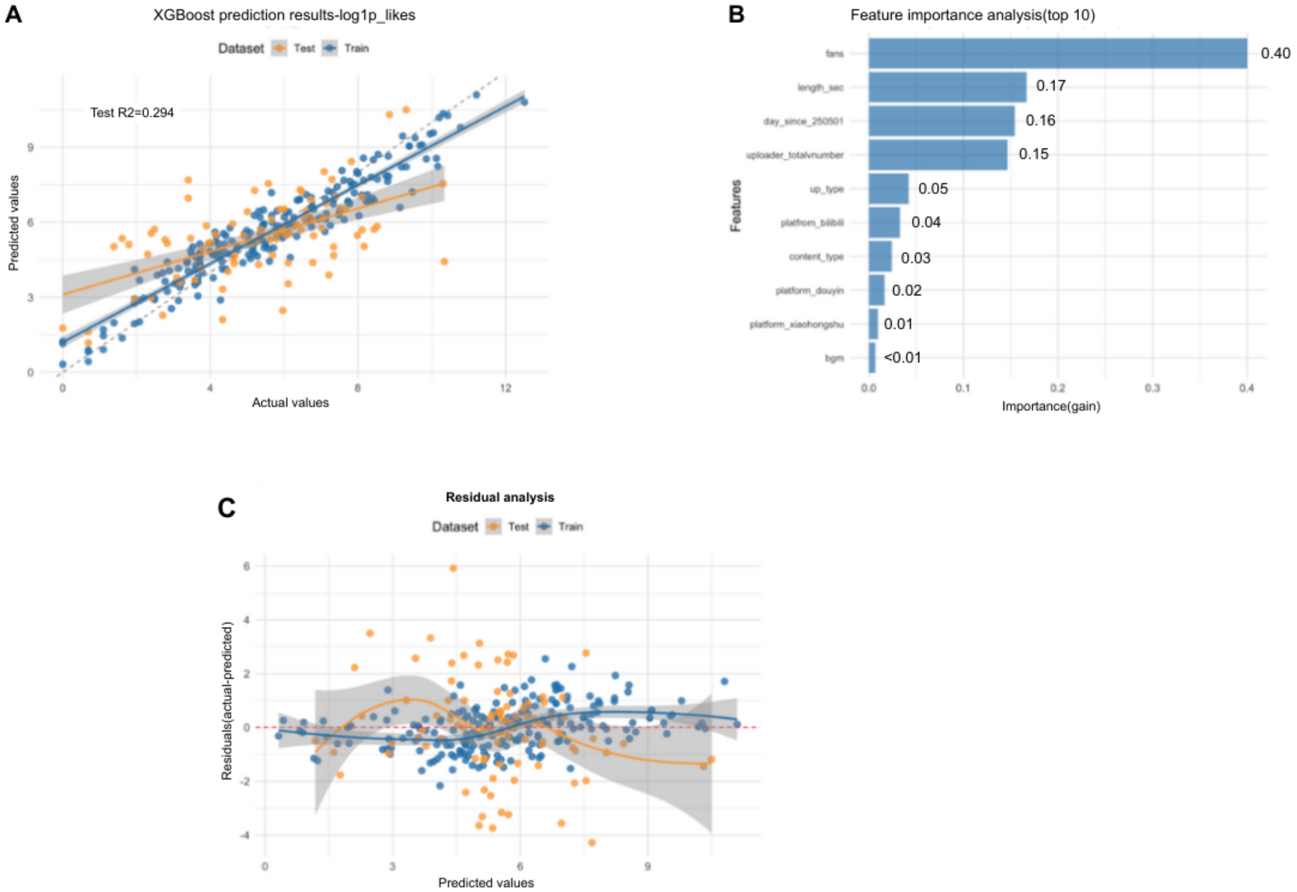
Figure 4. Results of the XGBoost model interpreting the factors influencing video likes: (A) the goodness of fit of the XGBoost model (R^2=−7.5); (B) the importance of each variable on the number of video likes; and (C) the Shapley Additive Explanation summary plot showing the impact of important variables 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).

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 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 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]

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]

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:
Chu D, Bai X, Guo F
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

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 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.

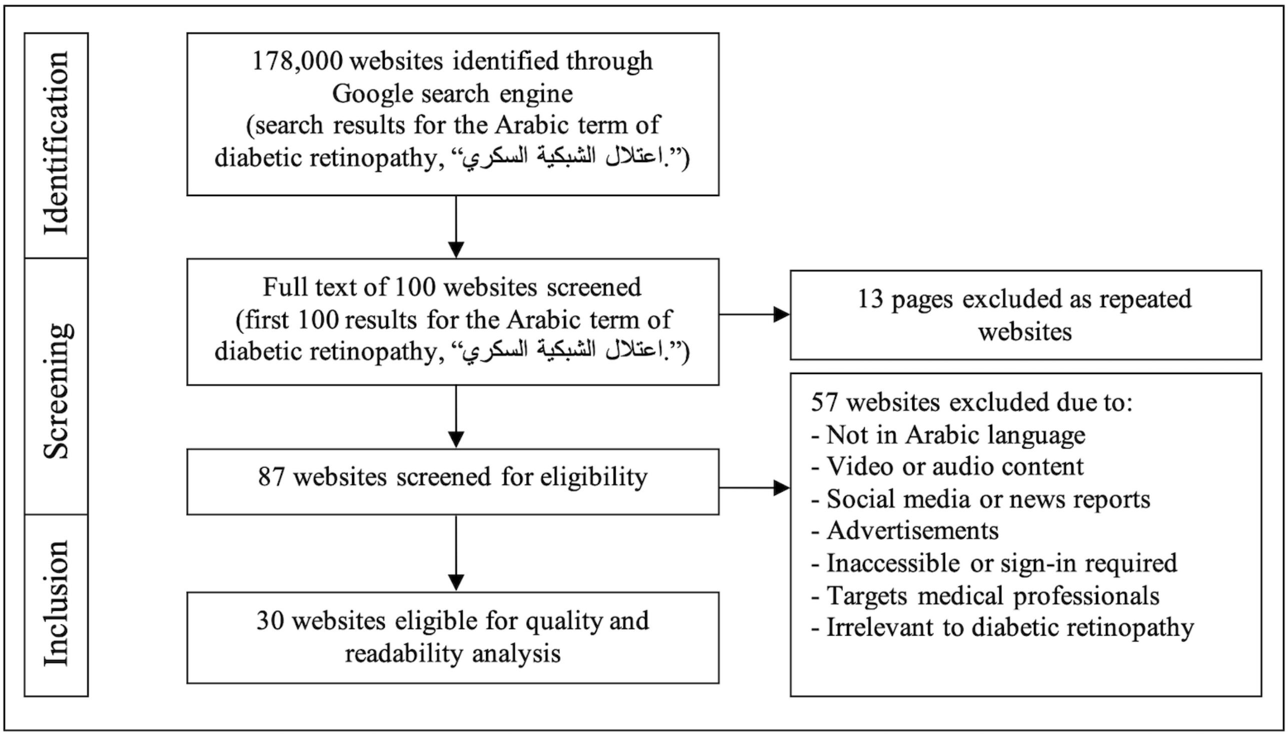


Table 1 exhibits the list of websites eligible for evaluation. Some websites, such as Mayo Clinic, have identical URLs in different languages but automatically localize content to Arabic. Websites were classified into four main categories: hospitals and medical centers, health portals (websites dedicated to health information), commercial (websites selling a service or product), and nonprofit organizations.

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)."

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).

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).

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.

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.

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.*

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.

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]

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]

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

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.

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 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

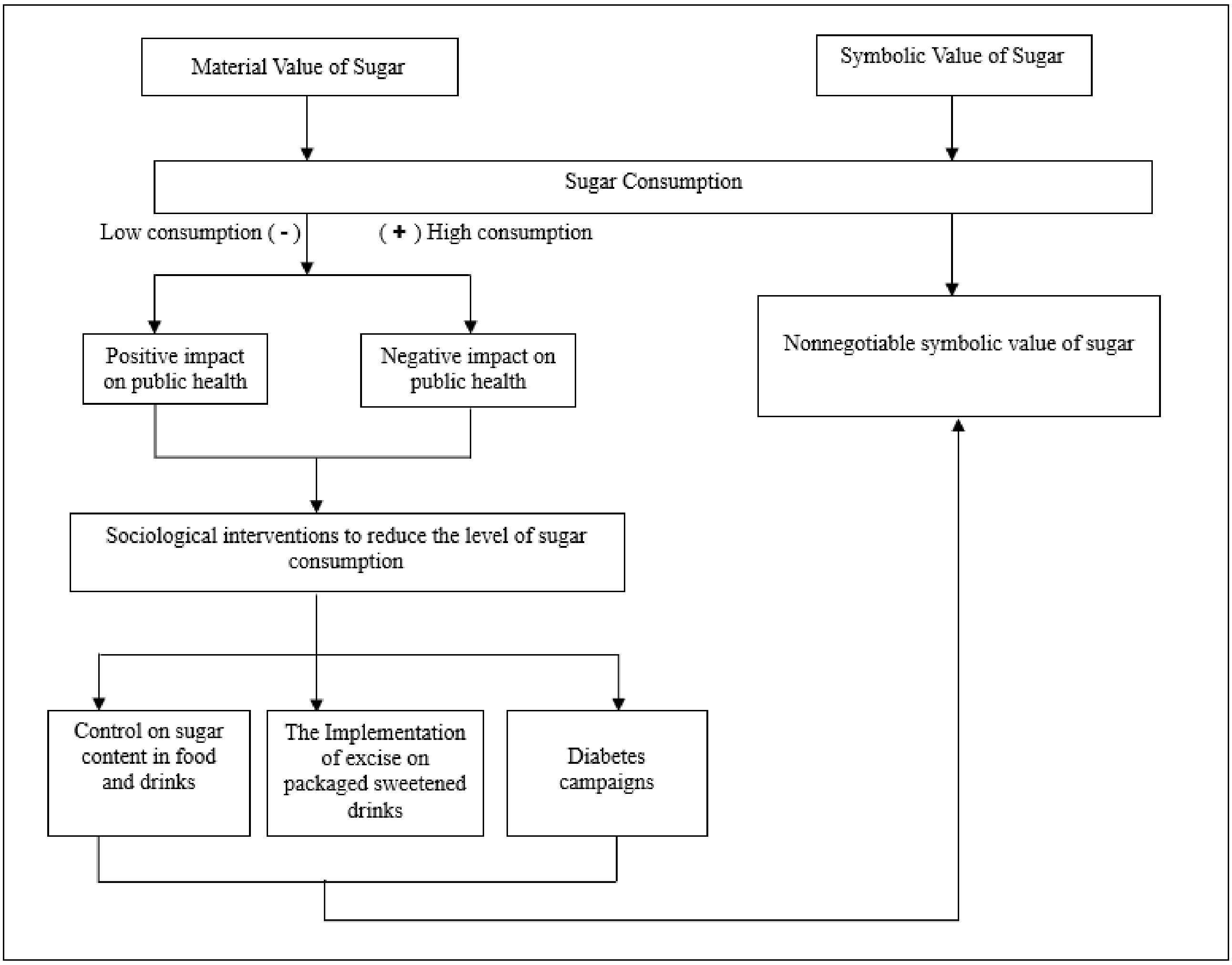
This study examines the nonnegotiable symbolic value in sugar consumption. Sugar plays a pivotal role in cultural communication as well as in economic and political contexts. Eating patterns are inseparable from cultural context, while food culture shapes individual identity [3].

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.

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.

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.

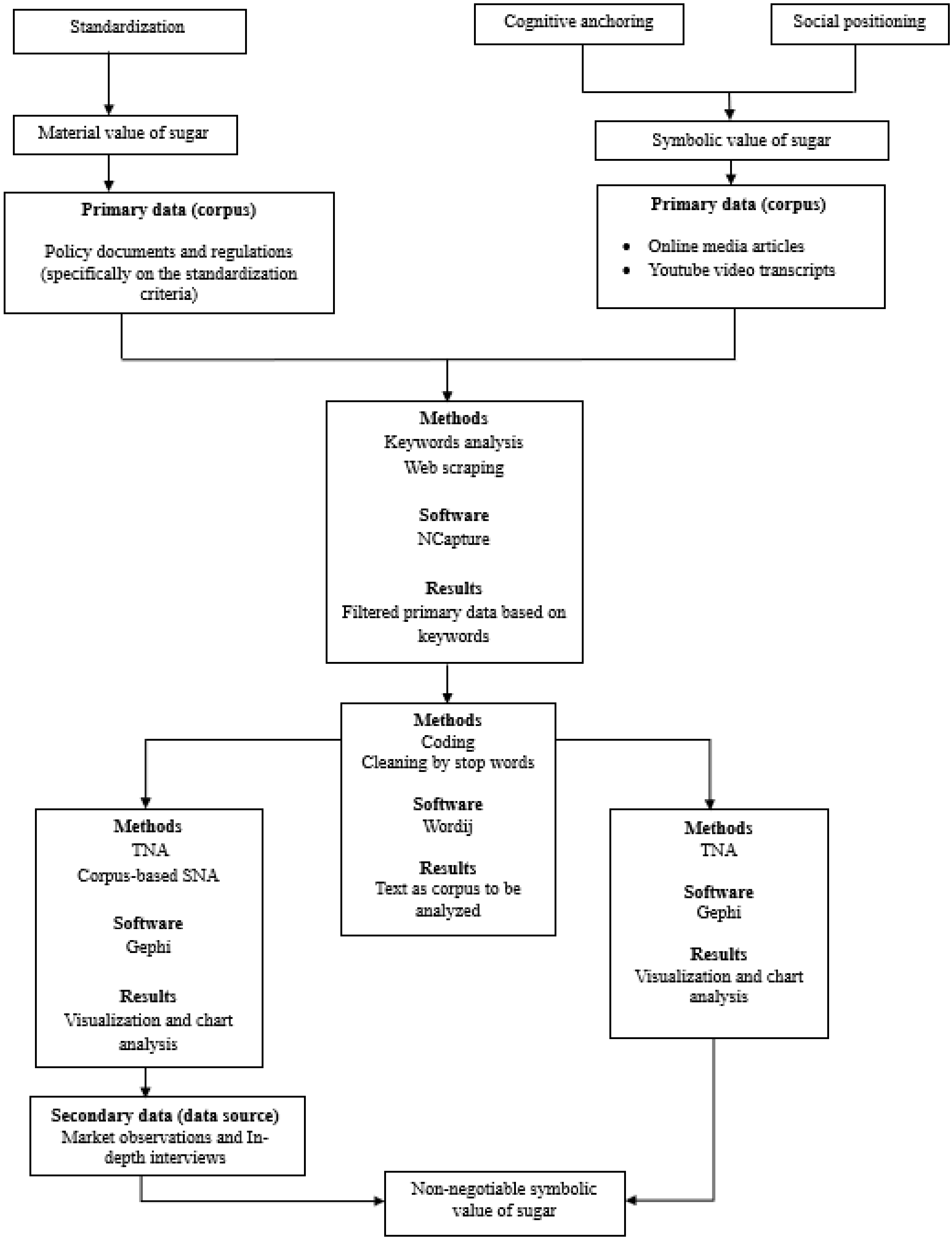
[a]SNI: Indonesian National Standard.

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.

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 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

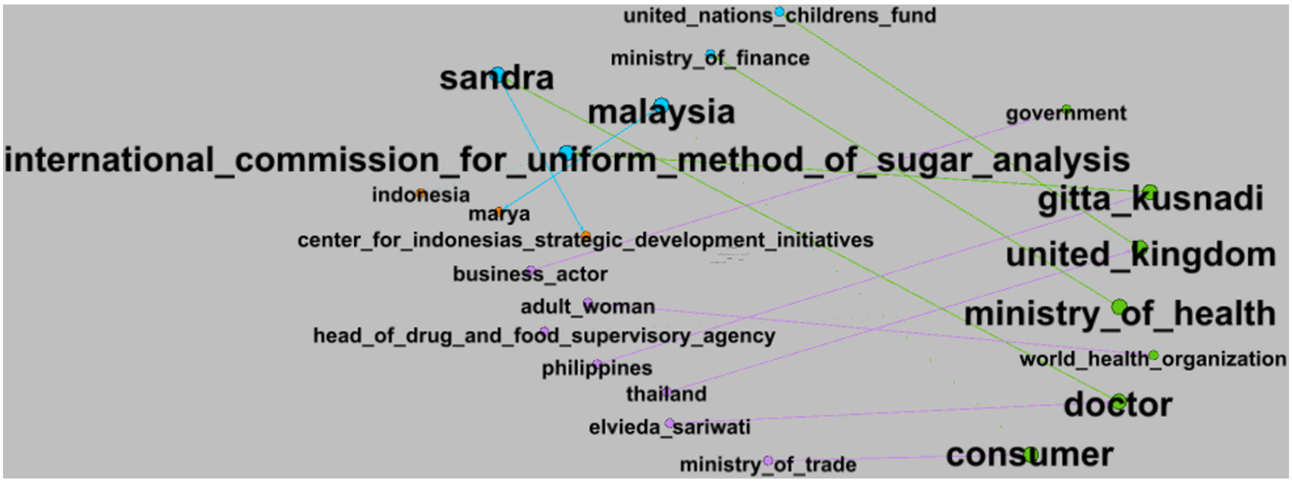
The TNA visualization of the sugar and artificial sweetener market, focusing on standardization, reveals 27 nodes and 58 edges with 6 clusters (Figure 3). These clusters include: purple (38/626, 6.07%), green (33/626, 5.27%), yellow (26/626, 4.15%), orange (21/626, 3.35%), blue (20/626, 3.19%), and pink (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.

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.

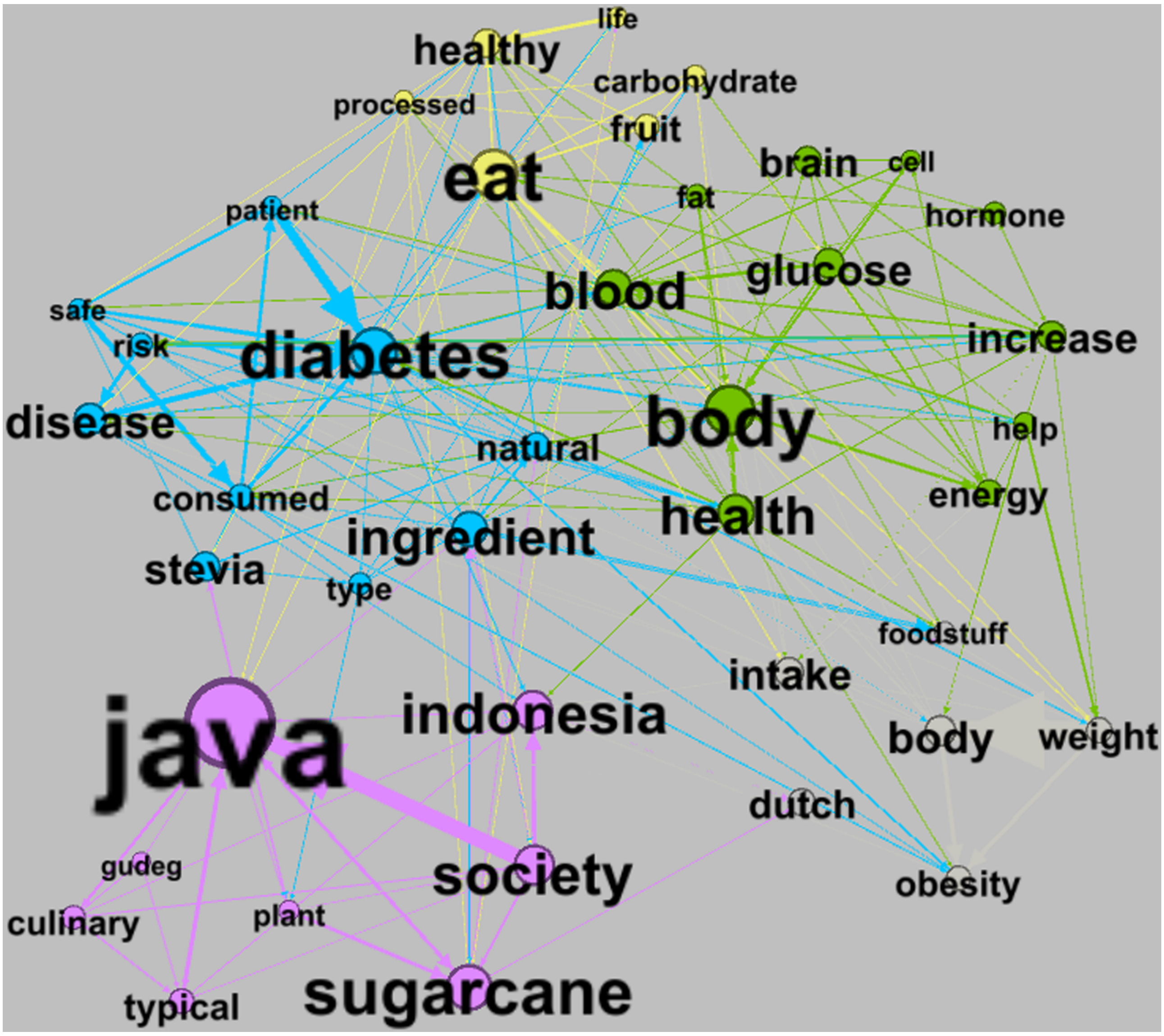
TNA Visualizations of the Nonnegotiable Symbolic Value of Sugar in Processed Foods and Beverages Containing Refined Sugar

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 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.

Figure 6. Textual network analysis (TNA) visualization of sugar based on social positioning.

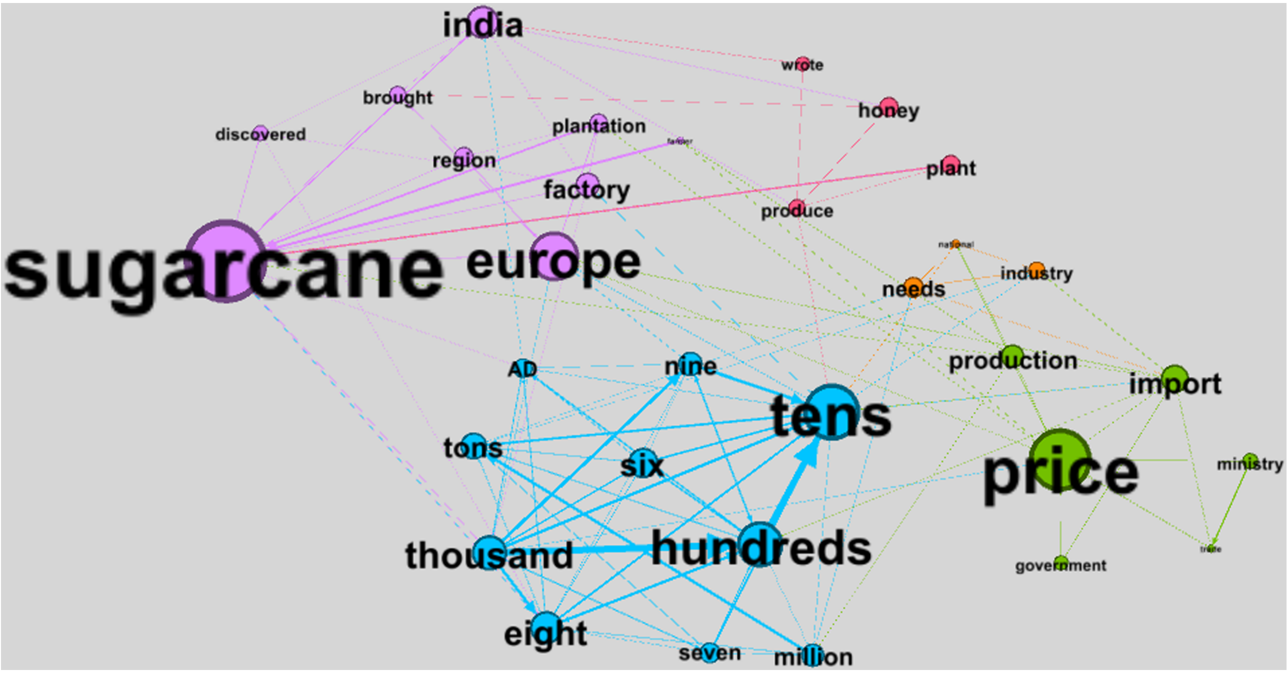


Figure 6 presents a TNA visualization of the discourse structure surrounding sugarcane and the sugar market through interconnected thematic clusters. The purple cluster centers on “sugarcane,” “Europe,” “India,” “plantation,” and “factory,” representing the historical, geographical, and production bases 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.

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.

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.

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.

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.

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.

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 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]

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]

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

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 (*B*s −0.49 to −0.69; .003≤*P*≤.04) and text-centered videos (*B*s −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 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:

- 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 queries: “depression symptoms,” “symptoms of depression,” “depression signs,” and “signs of depression.”

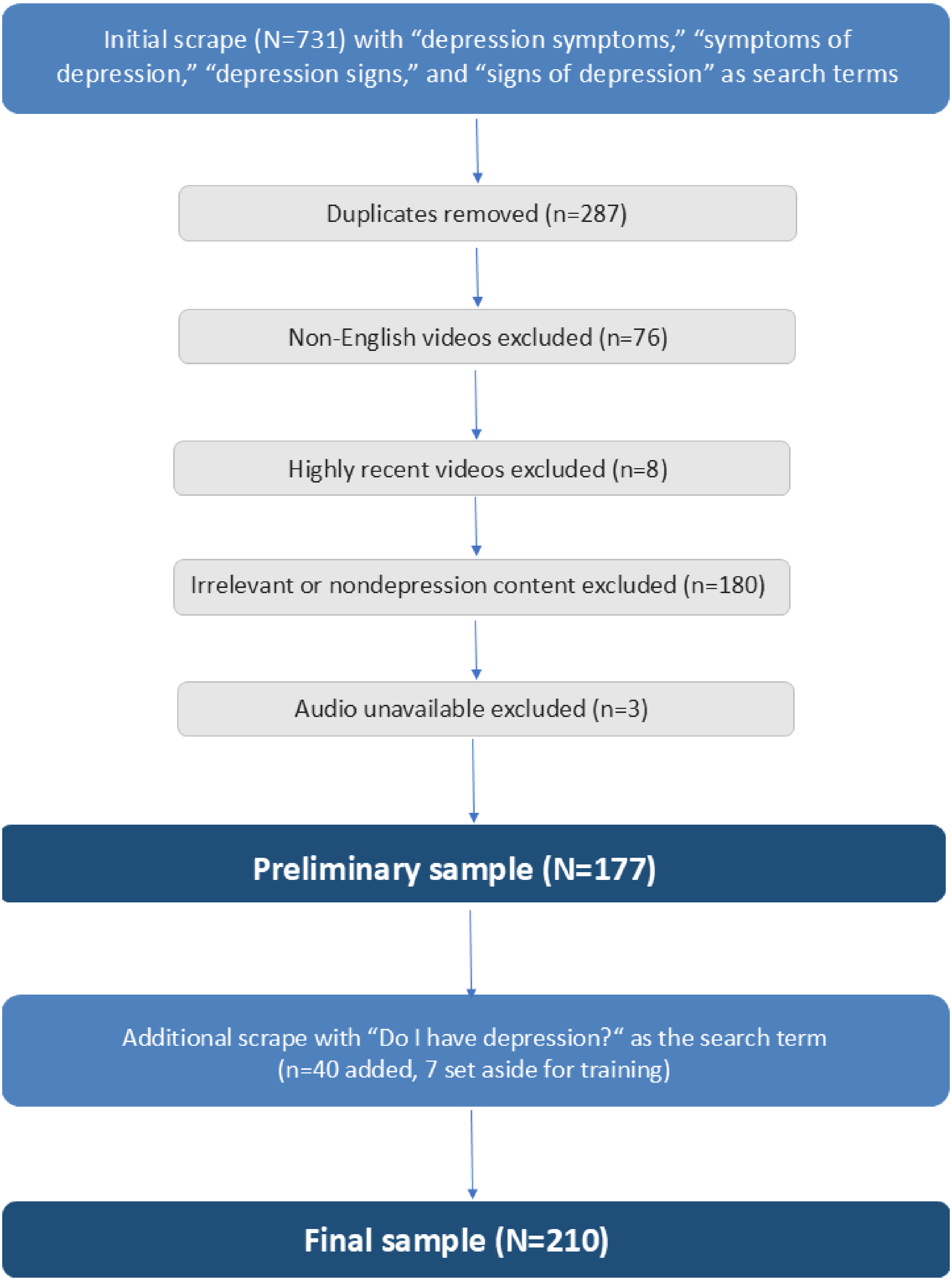
To minimize preexisting algorithmic biases—such as personalized content curation based on prior user behavior, interactions, or viewing history—a new TikTok account was created for data collection. This ensured that the search results were not influenced by an established user profile and instead reflected 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.

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.



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 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 (α=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.

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 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.

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 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 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]

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

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

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 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

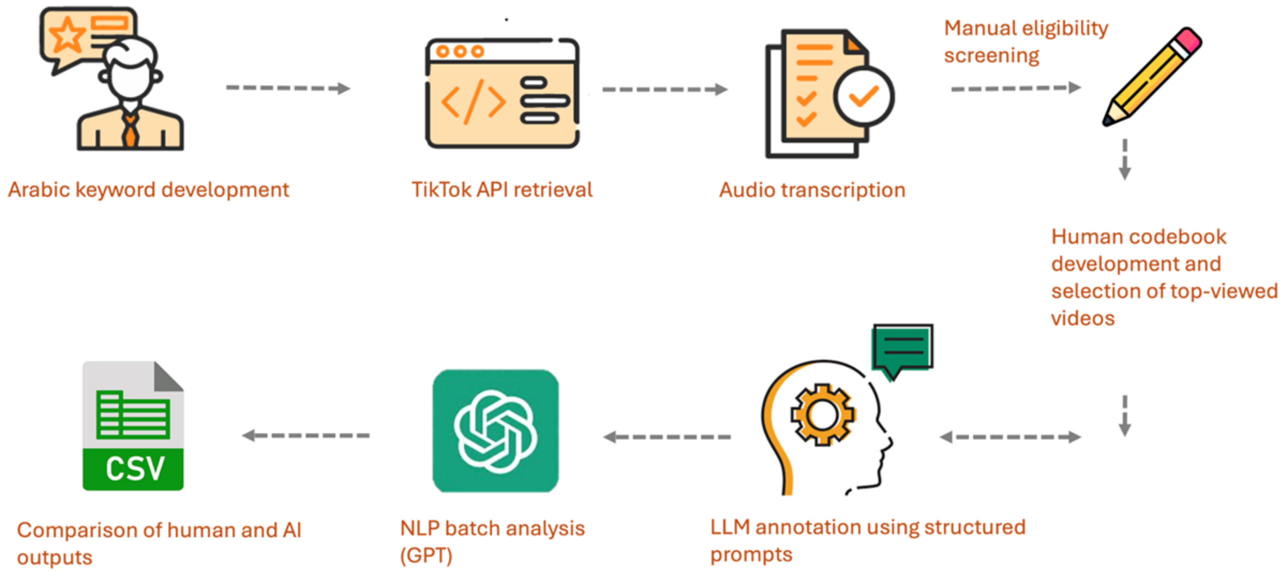
Advancements in artificial intelligence (AI), particularly large language models (LLMs) such as GPT, offer new opportunities for analyzing digital health communication in understudied languages [25,26]. GPT-4 has demonstrated high accuracy in detecting sentiment, misinformation, and medical accuracy 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.

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.



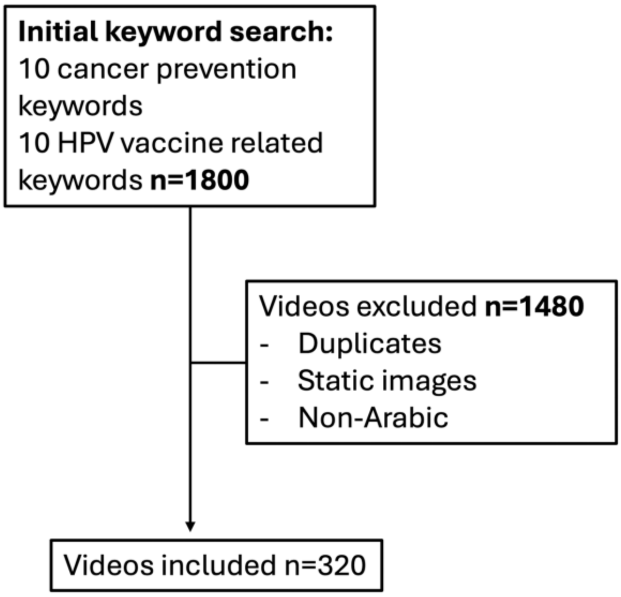
Using the TikTok application programming interface (API), we retrieved Arabic-language TikTok videos related to cancer prevention and the HPV vaccine from 2021 to 2024. This time frame was selected based on data from the Arab Youth Survey, which indicated an increasing trend in TikTok market penetration 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.

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).

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).

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.”

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).

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	

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%

(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 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 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]

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]

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:

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].

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 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

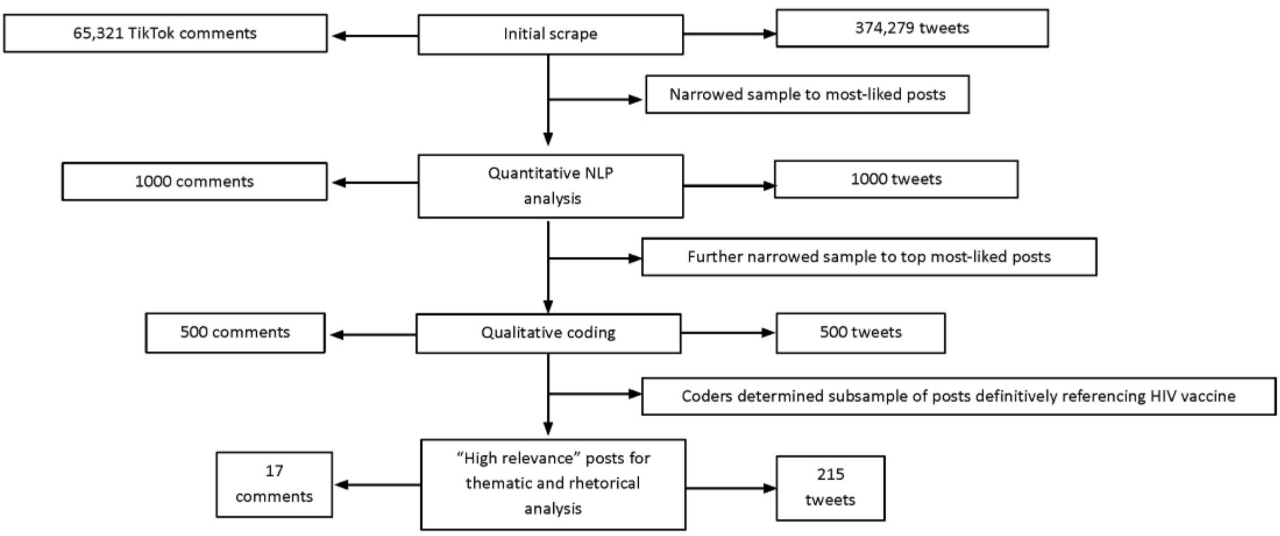
The study was approved by the ethics committees at the University of the Witwatersrand (reference: 230904), the University of Pennsylvania (854420), the University of Cape Town (666/2023), and Boston University (H-44422). The study used only secondary data and removed usernames and identifying details 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).

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.

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]

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*

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 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.

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]

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

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.

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.

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.

Data Labeling

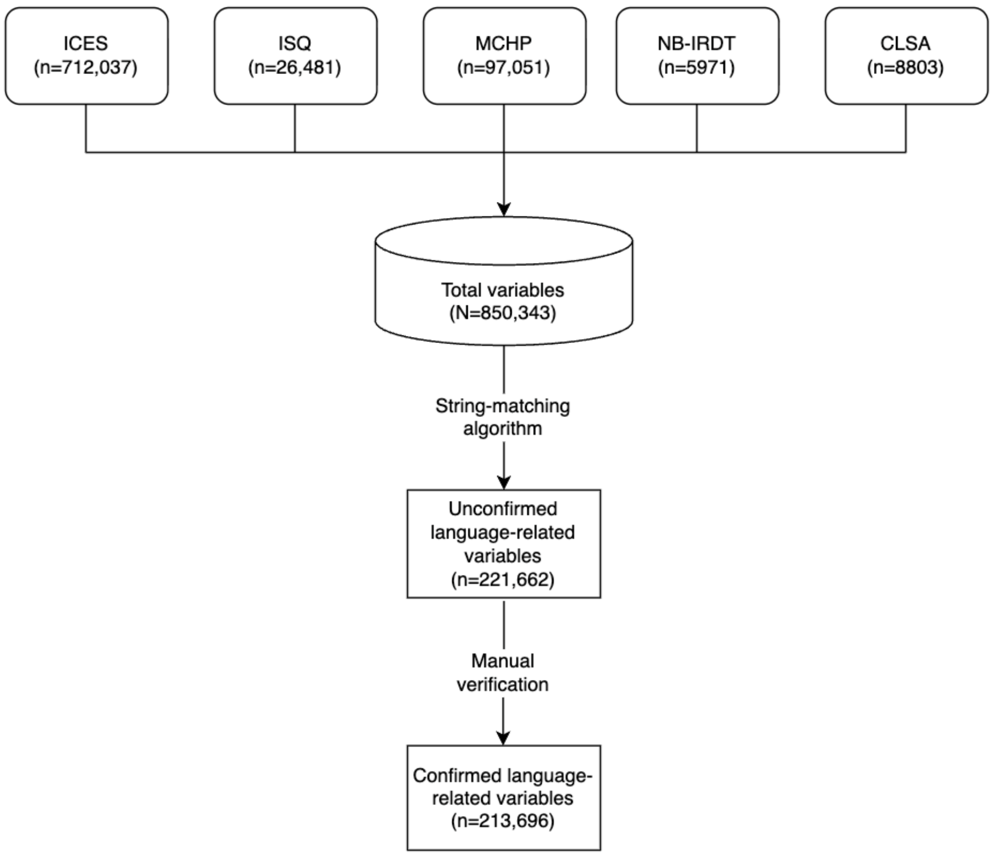
Data labeling (or tagging) is the common process of assigning one or more descriptive tags or labels to a dataset [32], which can make it easier to search and filter results while enabling other uses of the data (eg, machine learning) [33]. To provide an example of searchability in our proof-of-concept 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.



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 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 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]

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]

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

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, 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 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 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.

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

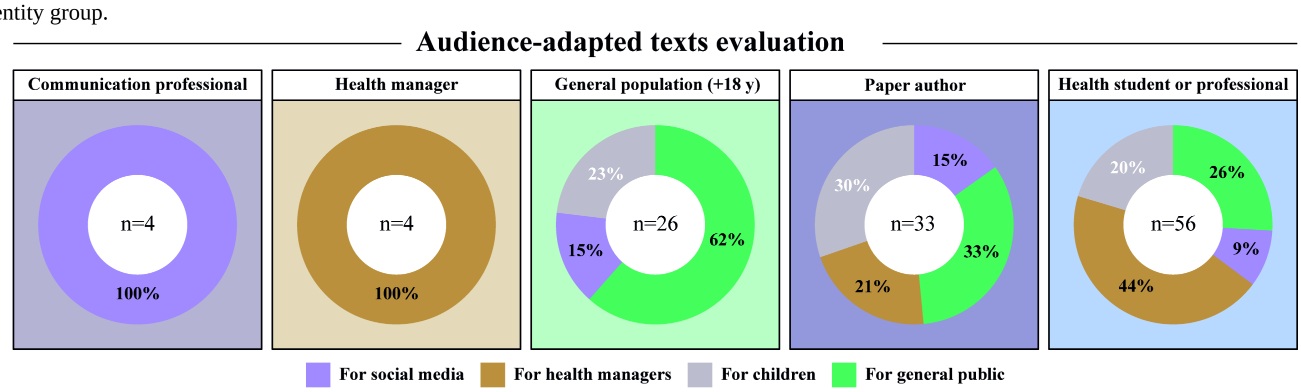
Overview

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 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)	

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).

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.

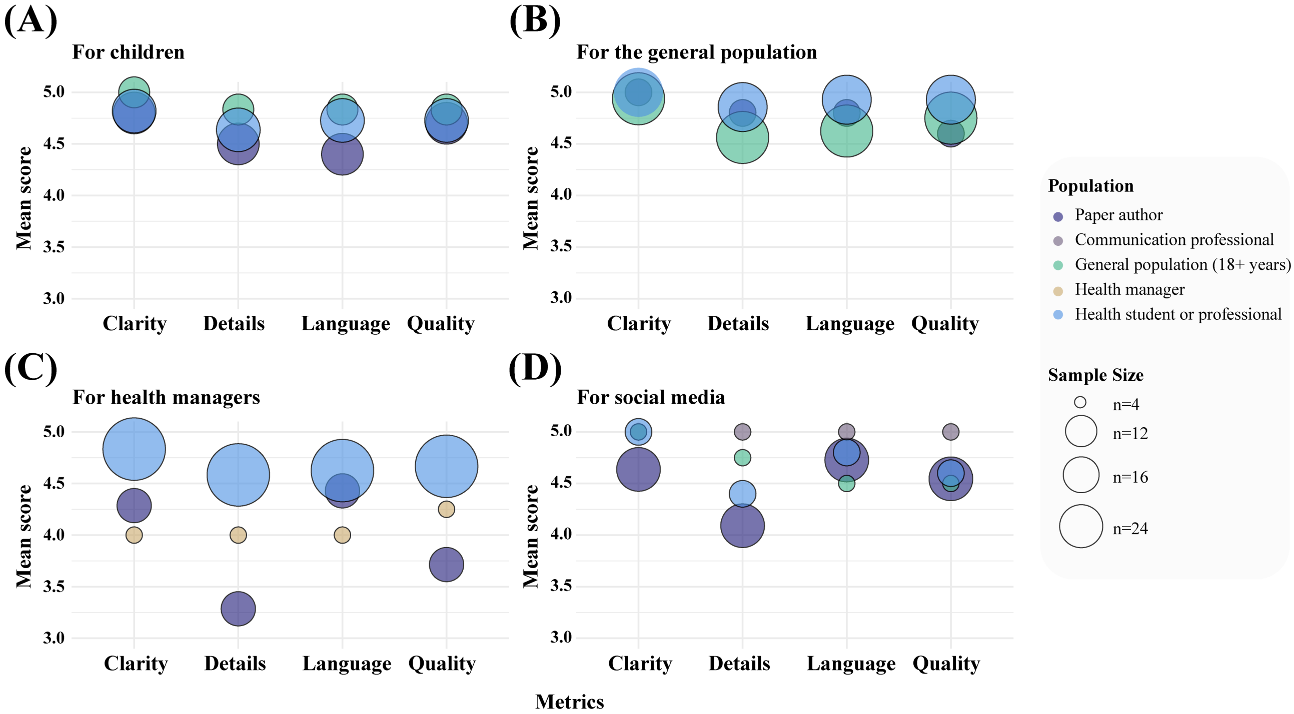


Figure 3. Proportional contributions of evaluation domains to each participant's total score, grouped by reviewer identity. A 100% stacked bar chart in which each vertical bar represents an individual participant's normalized total evaluation score (scaled to 100%), with participants organized along 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.

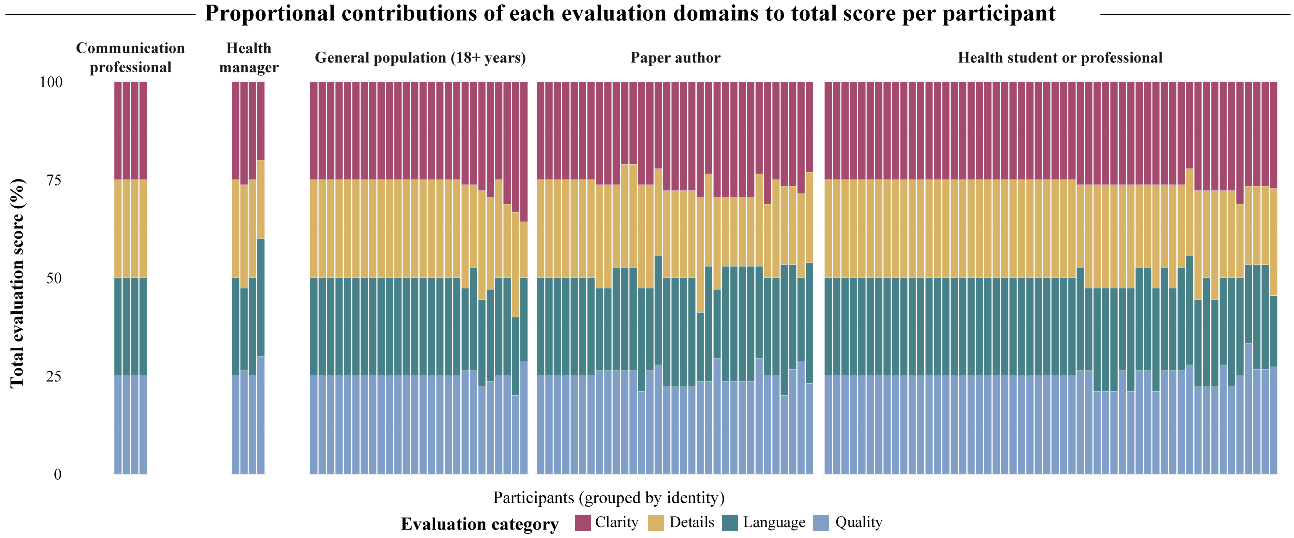


Table . Average (SD) evaluation scores by audience and stakeholder group for each adapted output. This table presents the central tendency and dispersion (mean, SD) of four evaluation domains: (1) clarity, (2) detailing, (3) language suitability, and (4) content quality, across four target‑audience contexts (“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.

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).

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

	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

	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).

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 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 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

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]

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

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