ARTICLE REVIEW

# When Viral Goes Wrong: A Systematic Review and Meta-Analysis of Mental Health Misinformation on TikTok

Christopher Tan[1*], Nadia Nathania[2], Sandra Sandra[1], Natasya Ayusandra Mahersaputri[1]

## ABSTRACT

**Introduction**: TikTok is a major source of mental health information, however, the prevalence of misleading content or misinformation across topics remains uncertain. This meta-analysis estimated the pooled prevalence of misleading or misinformation regarding mental health content on TikTok and compared prevalence across topic categories. **Methods**: We conducted a systematic review and meta-analysis from observational content analyses of publicly accessible TikTok videos. Misinformation prevalence was extracted as n/N and pooled using a random-effects meta-analysis of proportions on the logit scale, with back-transformed pooled estimates. Prespecified subgroup analyses compared attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and other mental health topics. **Results**: Eleven studies (2,978 videos) were included. The pooled prevalence of misleading or misinformation content was 31% (95% CI [18%, 48%]; $I^2$ = 96.9%), with a wide prediction interval (4%-83%). Topic-stratified prevalence was highest for ADHD-related content (51%, 95% CI [41%, 61%]; $I^2$ = 0%), followed by ASD (48%, 95% CI [23%, 75%]; $I^2$ = 91.4%), and lowest for other mental health topics (15%, 95% CI [8%, 28%]; $I^2$ = 93.2%). Sensitivity analyses yielded similar pooled estimates, although overall certainty of evidence was very low. **Conclusions**: Approximately one in three TikTok mental health videos contained misleading or misinformation content, with marked variation across topics. These findings support the need for targeted digital mental health literacy efforts and platform-aware clinical guidance.

**Keywords:** TikTok, Misinformation, Mental Health, ADHD, Autism

1. Department of Psychiatry, Faculty of Medicine, Universitas Indonesia; Dr. Cipto Mangunkusumo National Referral General Hospital, Jakarta, Indonesia
2. Faculty of Medicine, Udayana University, Denpasar, Indonesia

* ✉ email: christopherctan25@gmail.com

## INTRODUCTION

Social media has become an important part of how young people seek, interpret, and discuss mental health information, particularly when they want information that feels immediate, private, and easy to access.[1] For many young users, online spaces can complement formal care by offering peer connection, mental health education, and a lower-threshold route to exploring symptoms and help-seeking.[2] At the same time, the growing normalization of checking symptoms online and self-diagnosis in psychiatry has raised concerns that some users may rely on digital content before, or instead of, seeking professional assessment.[3]

Among current platforms, TikTok is particularly influential because it combines short-form video, high emotional salience, and algorithmic personalization in a way that makes mental health content highly visible and widely shareable.[4-6] This creates both opportunity and risk. Mental health content on TikTok may raise awareness, reduce stigma and broaden access to psychoeducation, however, the same platform features may also amplify oversimplified, inaccurate, or misleading claims.[2,7,8] More broadly, misinformation is already recognized as a substantial problem across social media, and recent work suggests that mental health content is not exempt from this pattern.[7,9]

Primary TikTok studies have already identified misleading content across several mental health topics, including ADHD, autism, and broader mental health discussions.[5,8,10,11] However, the available evidence remains fragmented by topic, sampling strategy, creator ecosystem, and misinformation definition, making it difficult to judge the overall burden of misinformation across TikTok mental health content as a whole.[5,8,10,12] This issue is especially important on TikTok because users are exposed not only to content they actively search for, but also to algorithmically recommended videos that may blur the boundaries between personal experience, psychoeducation, and diagnostic information.[4,5]

To our knowledge, no previous quantitative synthesis has pooled the prevalence of misleading or misinformation content across TikTok mental health videos while also comparing prevalence across major topic areas. This gap matters because topic-specific studies may not provide a clear sense of how common misleading content is overall, nor whether certain areas, such as ADHD or autism, may be especially vulnerable to misinformation. Therefore, the present systematic review and meta-analysis aimed to estimate the pooled prevalence of misleading or misinformation content in TikTok mental health videos and to compare prevalence across attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), and other mental health topics.

## METHODS

### *Study Design*

We conducted and reported this systematic review and meta-analysis in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines and the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) statement.[13,14] The study protocol was prospectively registered with PROSPERO prior to data extraction (CRD420261281996). No language restriction was applied.

### *Data Sources and Search Strategy*

A systematic literature search was conducted on 8 January 2026 across five electronic databases: PubMed, Google Scholar, Europe PMC, ProQuest, and Wiley Online Library. The search strategy combined terms related to the platform of interest ("TikTok" OR "Douyin"), information quality constructs ("misinformation" OR "inaccurate" OR "accuracy" OR "content analysis" OR "quality" OR "DISCERN" OR "JAMA" OR "PEMAT" OR "Global Quality Score"), and relevant mental

health conditions ("mental health" OR "ADHD" OR "attention deficit" OR "autism" OR "depression" OR "bipolar"). The complete search strings for each database are presented in **Supplementary Table S1**.

***Eligibility criteria and study selection***

Studies were eligible for inclusion if they met all of the following criteria: (1) the study examined publicly accessible TikTok videos containing mental health or neurodevelopmental disorder-related information, encompassing any diagnostic label (e.g., ADHD, autism spectrum disorder [ASD], depression, anxiety, obsessive-compulsive disorder [OCD], post-traumatic stress disorder [PTSD], and related conditions); (2) the study used a quantitative observational cross-sectional content analysis design with structured coding or rating; (3) the primary outcome or a reportable secondary outcome included an extractable misinformation prevalence expressed as a count of misleading or misinformation videos (n) over a total number of videos assessed (N); and (4) full text was available. Studies were excluded if: they did not focus on the accuracy or misinformation content of TikTok videos (engagement-only analyses); no extractable n/N for misinformation prevalence was available; the dataset was a duplicate of another included study; or the content was non-public (e.g., private groups, closed accounts). For multi-platform studies, TikTok-specific data were required to be separately reportable.

Duplicate studies were removed prior to screening using the duplicate removal tool in EndNote X9 (Clarivate Plc., London, England) and were then further manually verified. Two reviewers (CT and NN) independently screened titles and abstracts against the eligibility criteria, followed by full-text assessment of potentially eligible studies. Disagreements at each phase were resolved through discussion. Inter-rater agreement at title/abstract screening was 98.1% (202/206), with a Cohen's kappa of 0.872, indicating an almost perfect agreement.

***Data Extraction***

Data were extracted independently by three reviewers (CT, SS, NAM) using a standardized form. For each included study, the following information was collected: author and publication year; study country and setting; content topic or mental health condition; sampling frame and search strategy (e.g., hashtags used, number of videos retrieved and sampled); definition and operationalization of misinformation or misleading content; content rating approach and instruments used (e.g., Journal of American Medical Association [JAMA] benchmark, Patient Education Materials Assessment Tool for Audiovisual Materials [PEMAT-A/V]); type of content raters; count of videos classified as misinformation or misleading (n); total number of videos assessed (N); and, where available, the count of misinformation videos produced by healthcare professionals (HCP), non-healthcare professionals (non-HCP) and the total count of non-HCP videos. Discrepancies in extracted data were resolved through discussion and reference to the original publication.

***Risk of Bias Assessment***

Study-level risk of bias was assessed using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Analytical Cross-Sectional Studies.[15] Two reviewers (CT and NN) independently evaluated each study across eight JBI domains: clarity of the study objective; specification and representativeness of the sampling frame; use of valid and standardized measurement tools; objective and consistent content classification criteria; identification of confounding factors; strategies to address confounding; reliability of outcome measurement (including inter-rater reliability); and appropriateness of statistical analysis. Each domain was rated as "Yes,"

"No," "Unclear," or "Not Applicable." Disagreements were resolved by consensus. An overall risk of bias judgment (low, moderate, or high) was assigned based on the number of affirmative responses relative to applicable items. Findings are presented in **Supplementary Table S2** and are incorporated into sensitivity analyses and the certainty of evidence assessment.

***Data synthesis and statistical analysis***

Statistical analyses were performed in R using the meta package. The primary outcome was the proportion of TikTok videos classified as misleading or misinformation in each study, pooled using a random-effects meta-analysis of proportions with logit transformation, inverse-variance pooling, restricted maximum likelihood (REML) estimation of between-study variance ($\tau^2$), and Hartung-Knapp (HK) adjusted confidence intervals (CIs).[16] Pooled estimates were back-transformed to proportions and reported with 95% CIs and 95% prediction intervals (PIs). Heterogeneity was quantified using $I^2$, $\tau^2$, and Cochran's Q. Prespecified subgroup analyses compared ADHD, ASD, and other mental health topics, and robustness was assessed using leave-one-out influence analysis. Exploratory small-study effects were assessed using funnel plots, a regression-based asymmetry test, and trim-and-fill. An exploratory meta-regression assessed whether the proportion of non-HCP content was associated with misinformation prevalence. The complete detailed analytic strategy can be seen in the **Supplementary Methods** in the **Supplementary Appendix**.

***Certainty of Evidence***

The certainty of the evidence was evaluated using the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) framework, adapted for single-arm proportion meta-analyses.[17] The GRADE assessment considered five domains: (1) risk of bias based on JBI appraisal results; (2) inconsistency, operationalized through $I^2$, $\tau^2$, and the width of the prediction interval; (3) indirectness, with downgrading if included studies used non-standard sampling frames, non-public content, or populations not representative of general TikTok users; (4) imprecision, based on the width of the 95% CI for the pooled estimate and total information size per subgroup; and (5) publication bias, based on funnel plot assessment and Egger's test. Starting certainty was set at 'Low' given the observational cross-sectional design of all included studies. An overall certainty rating was assigned.

## RESULTS

***Study selection***

The database search yielded 324 records in total across five different databases. After automated and manual deduplication, 206 unique records were retained for screening. Following title and abstract review, 186 records were excluded as ineligible, most commonly because they did not report any misinformation-specific outcome, examined platforms other than TikTok without extractable TikTok-specific data, or were of other study design (e.g., commentary, editorials, protocols). The remaining 19 records were retrieved for full-text assessment. Of these, eight were excluded at full-text screening due to non-extractable misinformation prevalence outcomes, ineligible study design, non-original publication type, or major internal reporting inconsistencies. A total of 11 studies were included in the qualitative synthesis and quantitative meta-analysis (see **Supplementary Table S3** for the list of all included studies and **Supplementary Table S4** for the list of excluded studies and their reasons for exclusion during full-text review).[4,5,8,10-12,18-22] The PRISMA flow diagram illustrating the selection process is presented in **Figure 1**.

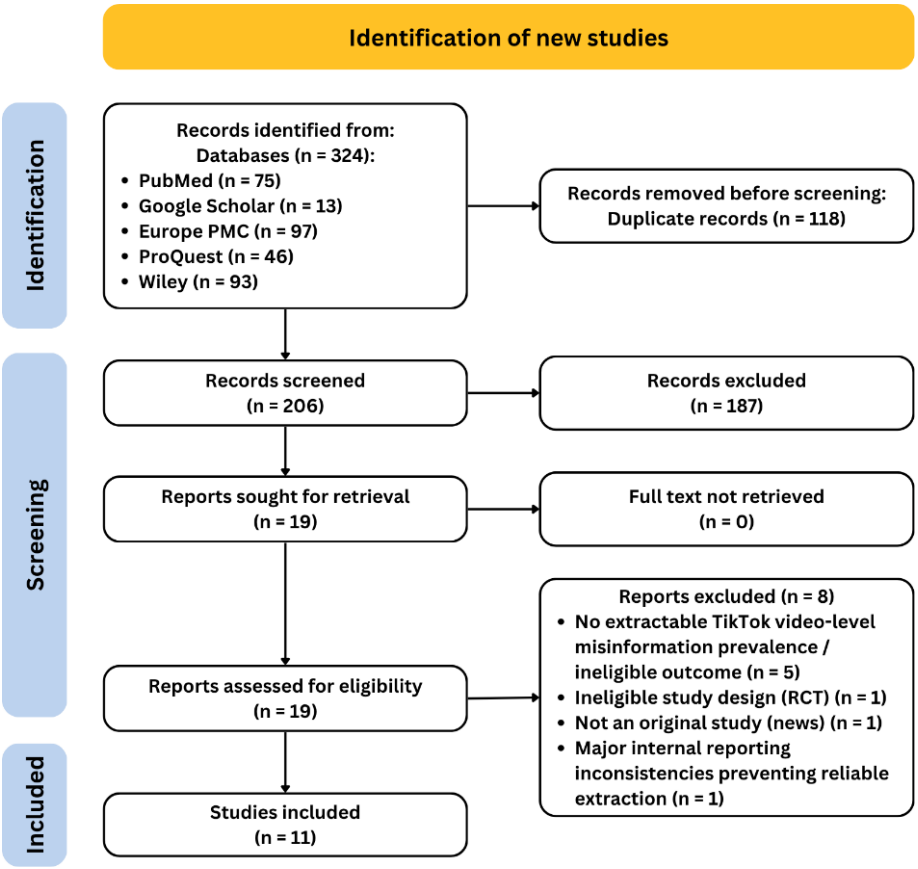



**Figure 1.** PRISMA Study Flowchart

***Characteristics of included studies***

The 11 included studies were published between 2022 and 2025 and collectively encompassed 2,978 TikTok videos. Studies originated primarily from North America and Europe, with one joint USA/UK study. All studies were cross-sectional content analyses of publicly accessible TikTok content retrieved using hashtag-based or algorithm-guided sampling. Sample sizes ranged from 100 to 1,000 videos per study (median of 100 videos). Mental health topics covered in included studies were ADHD (n = 2 studies), ASD (n = 4 studies), and other mental health topics (n = 5 studies), including general mental health content, OCD, PTSD, NSSI, and mixed psychiatric content. Among the six studies that explicitly reported video-level creator type, non-HCP videos accounted for 638 of 683 sampled videos (93.4%).

Misinformation was operationalized variously across studies. Most studies categorized videos as misleading using structured coding manuals or clinician/researcher judgment against established clinical guidelines (e.g., DSM-5 criteria), while some autism-focused studies distinguished misleading content from overgeneralization and the study by Hudon et al. separately coded disinformation and misinformation. PEMAT-A/V and JAMA benchmark tools were commonly used to assess communication quality, understandability, or actionability, but in most studies these instruments were not the sole basis for misinformation classification. Content raters included psychiatrists, clinical psychologists, trained researchers, or combinations of those experts. Detailed study characteristics are presented in **Table 1.**

**Table 1.** Characteristics of Included Studies

| Study (Year) | Topic | Country | Sampling Approach | Misinformation Definition / Operationalization | Content Raters | Misinform-ation Count (n) | Videos Identified / Sampled (N) | Proportion Misinfor-mation (%) | Non-HCP videos (%) |
|---|---|---|---|---|---|---|---|---|---|
| Yeung et al. (2022) | ADHD | Canada | #ADHD query; top 100 videos after exclusions | *JAMA benchmark + PEMAT-A/V; clinician rating using ADHD practice guidelines* | Psychiatrist + psychiatry resident (2 raters); third psychiatrist resolved disagreements | 52 | 100 | 52.0% | 89.0% |
| Sieferle et al. (2025) | ADHD | Germany | Top 125 most-liked TikTok videos from #ADHD query, uploaded after July 18, 2021 | *PEMAT-A/V; content rating using German S3 guideline and ICD-11* | Interdisciplinary team of 3 raters (physician, clinical psychologist, health scientist) | 63 | 125 | 50.4% | NR |
| Aragon-Guevara et al. (2025) | ASD | USA | #Autism hashtag; 133 informational TikTok videos | *Expert clinician fact-check against established clinical criteria for accuracy* | 2 coders under supervision of a senior clinician with ~20 years experience | 55 | 133 | 41.4% | 94.7% |
| Brennan et al. (2025) | ASD | USA | #actuallyautistic, #latediagnosedautistic, and #autism; 50 videos per hashtag (150 total) | *Structured coding by trained coders against DSM-5/clinical guidelines* | 4 trained coders | 56 | 150 | 37.3% | 99.3% |
| Brown et al. (2024) | ASD | USA | Top 100 TikTok videos from “autism spectrum disorder” query | *PEMAT-A/V; videos rated as misleading using an adapted Sood protocol* | 2 psychiatrists + 2 pre-medical researchers; senior researcher resolve ties | 40 | 100 | 40.0% | 86.0% |
| Ononuju & Ujari (2025) | ASD | USA / UK | #ASD, #Autism, #ASDinfo; Top 100 TikTok videos | *Structured codebook for ASD misinformation and stigma categories; TikTok misinformation prevalence reported separately* | 2 independent coders | 74 | 100 | 74.0% | NR |

| Study (Year) | Topic | Country | Sampling Approach | Misinformation Definition / Operationalization | Content Raters | Misinform-ation Count (n) | Videos Identified / Sampled (N) | Proportion Misinfor-mation (%) | Non-HCP videos (%) |
|---|---|---|---|---|---|---|---|---|---|
| Turuba et al. (2025) | General mental health | Canada | Top 1,000 #mentalhealth TikTok videos; 970 English videos included | *Clinical psychologist + psychiatrist consensus coding for misleading or clinically inaccurate content* | Licensed clinical psychologist + 2 psychiatrists | 106 | 970 | 10.9% | NR |
| Fitzpatrick et al. (2025) | OCD | USA | #OCD query; first 100 eligible videos discussing OCD | *PEMAT-A/V; videos categorized by expert raters as misleading* | 2 OCD researchers; clinical psychologist break ties | 29 | 100 | 29.0% | 93.0% |
| Rohl et al. (2025) | PTSD | USA | #PTSD search; 100 eligible top-liked videos | *PEMAT-A/V; video categorization using DSM-5 / clinical consistency* | 2 reviewers with PTSD research experience; third reviewer resolved disagreements | 12 | 100 | 12.0% | NR |
| Hudon et al. (2025) | Mixed mental health | Canada | 1,000 TikTok videos across 26 mental health topics in English, French, and Spanish | *Framework-based coding of disinformation and misinformation; combined disinformation + misinformation extracted for this review* | 8 trained raters; each video assessed at least twice | 220 | 1,000 | 22.0% | NR |
| Satterfield (2024) | Non-Suicidal Self-Injury | USA | #NSSI and #SH hashtags; top 100 most-liked eligible videos | *JAMA benchmark + PEMAT-A/V; videos coded by trained coders as misleading* | Study PI + additional trained coders with weekly consensus meetings | 8 | 100 | 8.0% | 95.0% |

*Note. ADHD = attention-deficit/hyperactivity disorder; ASD = autism spectrum disorder; OCD = obsessive-compulsive disorder; PTSD = post-traumatic stress disorder; NSSI = non-suicidal self-injury; JAMA = Journal of the American Medical Association benchmark; PEMAT-A/V = Patient Education Materials Assessment Tool for Audiovisual Materials; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; HCP = healthcare professional; NR = not reported at the video level*

### *Risk of Bias Assessment*

Risk of bias appraisal using the JBI Analytical Cross-Sectional Studies Checklist revealed that most included studies have a low risk of bias, with 10 rated low risk and only one rated moderate risk (see **Supplementary Table S2**). Strengths consistently observed across studies included clearly stated research objectives, transparent content sampling procedures (with the use of hashtags), and generally appropriate statistical analyses. The main concern was incomplete reporting of outcome measurement reliability or formal inter-rater reliability in some studies, and the reliance on convenience sampling using hashtags, which may not represent the full distribution of TikTok mental health content.

### *Overall Prevalence of Misinformation*

Eleven studies contributing 2,978 videos were pooled in the primary random-effects meta-analysis of proportions. The overall pooled prevalence of misleading or misinformation content on TikTok was 31% (95% CI [18%, 48%]), indicating that approximately one in three mental health videos sampled from TikTok across included studies was classified as misleading or containing misinformation. However, substantial statistical heterogeneity was observed ($I^2$ = 96.9%; Q [df = 10] = 323.42; p <.001; $\tau^2$ = 1.05), indicating that true underlying misinformation prevalence varied considerably across studies. Consistent with this, the 95% prediction interval was wide, spanning 4% to 83%. The forest plot for all included studies is presented in **Figure 2**.

### *Subgroup Analyses by Mental Health Topic*

Prespecified subgroup analysis by mental health topic category revealed clinically and statistically meaningful differences in misinformation prevalence across the three strata (Q for subgroup differences = 38.24, df = 2; p <.001). The subgroup forest plot is presented in **Figure 3**.

#### ADHD

Two studies examining ADHD-related TikTok content were included in this subgroup.[5,11] The pooled misinformation prevalence was 51% (95% CI [41%, 61%]), suggesting that approximately half of ADHD-related TikTok videos examined across these studies contained misleading content. Notably, heterogeneity within this subgroup was negligible ($I^2$ = 0%), indicating highly consistent findings across the two studies despite their differing national contexts (Canada and Germany, respectively).

#### ASD

Four studies examining ASD-related content were included in this subgroup.[10,18,19,21] The pooled misinformation prevalence was 48% (95% CI [23%, 75%]), a finding that was broadly consistent in direction with the ADHD subgroup, though the confidence interval was considerably wider, reflecting the notable within-subgroup heterogeneity ($I^2$ = 91.4%). This heterogeneity is attributable in part to the markedly higher misinformation rate reported by Ononuju and Ujari (74%)[21], who examined both stigmatizing and misinformation content using a broad coding framework, relative to the lower rates reported by Aragon-Guevara et al. (41.4%), Brennan et al. (37.3%), and Brown et al. (40.0%).[10,18,19] Subgroup results should therefore be interpreted with caution.

#### Other Mental Health Topics

Five studies covering general mental health[8], OCD[20], PTSD[4], general psychiatric conditions[12], and NSSI[22] comprised this subgroup. The pooled misinformation prevalence was 15% (95% CI [8%, 28%]), substantially lower than in the ADHD and ASD subgroups. However, substantial residual heterogeneity persisted ($I^2$ = 93.2%), driven by the markedly lower rates observed in NSSI (Satterfield, 8%), general mental health (Turuba et al., 10.9%) and PTSD (Rohl et al., 12%) videos relative to the higher rates observed in general psychiatric videos (Hudon et al., 22%), with OCD videos showing an

intermediate rate (Fitzpatrick et al., 29%).[4,8,12,20,22] These differences likely reflect heterogeneity in content type, coding frameworks, and video selection strategy across studies in this grouping.

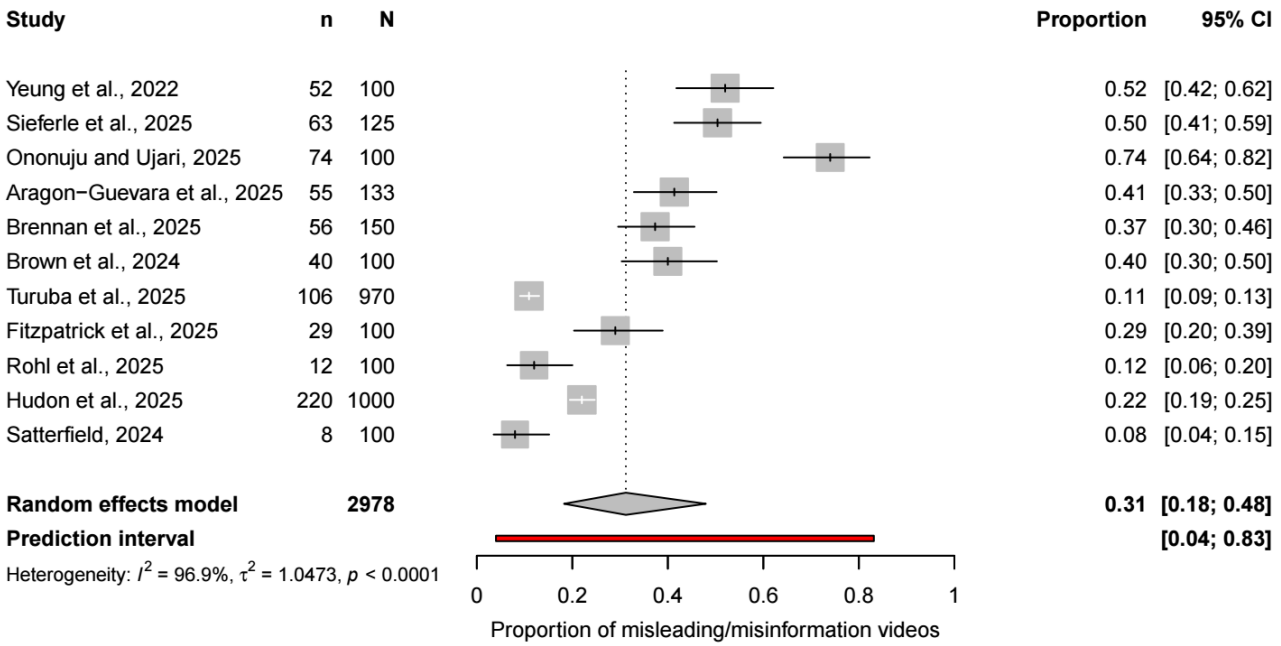



**Figure 2.** Overall Random-Effects Forest Plot

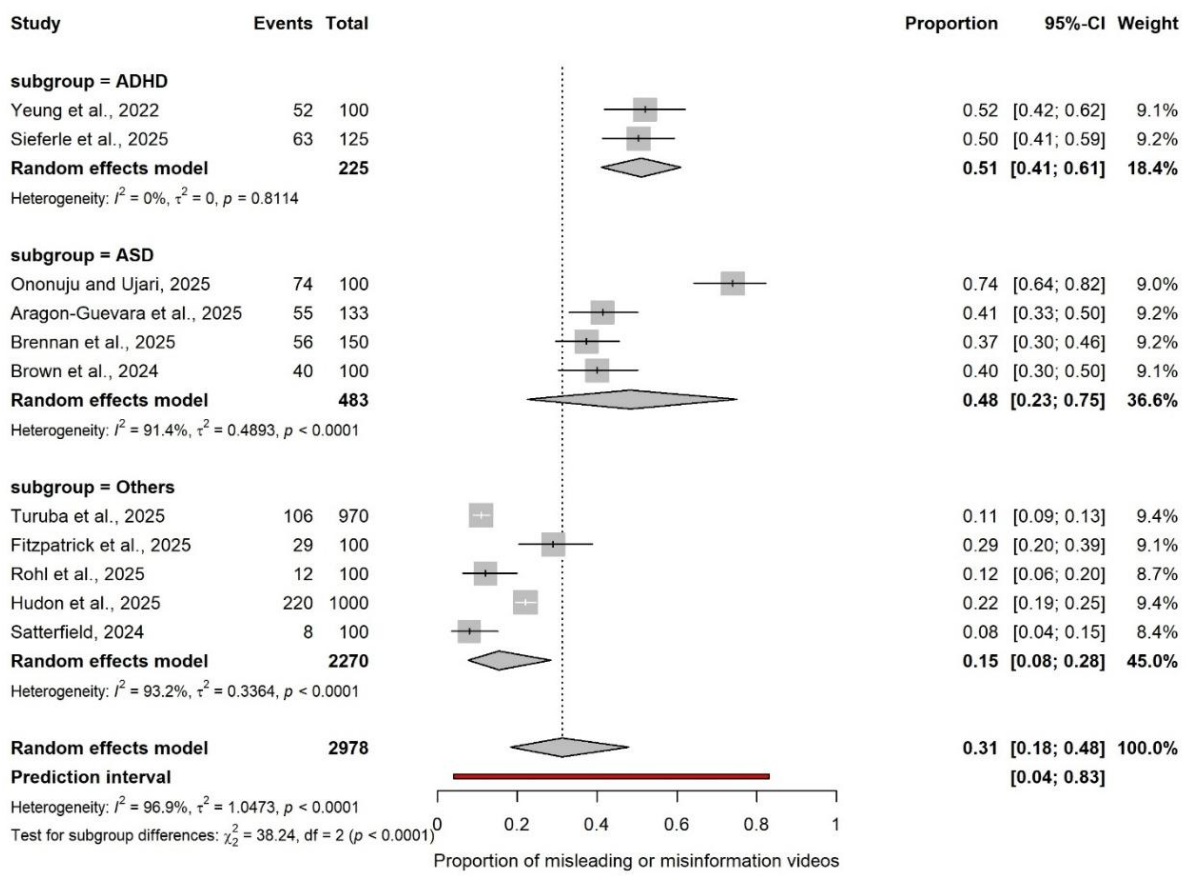



**Figure 3.** Subgroup Forest Plot by Mental-Health Topic

***Sensitivity and Influence Analyses***

Leave-one-out influence analysis demonstrated that the pooled overall estimate was robust to the sequential omission of each individual study. When any single study was excluded, the re-estimated pooled prevalence remained within the range of 28% to 35%, and the interpretation of the findings remained the same, indicating that no single study influenced the overall result. Results of the leave-one-out analysis are presented in **Supplementary Figure S1.** Because no study was rated as having a high risk of bias, the prespecified analysis excluding high-risk studies was identical to the primary model. We attempted a secondary low-risk only sensitivity analysis excluding the single moderate-risk study, and the pooled estimate remained almost identical, with a pooled estimate of 30% (95% CI [17%, 49%]; $I^2$ = 97.1%), supporting the robustness of our primary finding (**Supplementary Figure S2**).

***Small Study-Effects and Publication bias***

With the 11 included studies available for the primary analysis, both visual inspection of the funnel plot and formal testing for asymmetry were undertaken. The funnel plot suggested mild asymmetry, with smaller studies tending to report higher misinformation prevalence (**Supplementary Figure S3-S4**), but formal tests did not show statistically significant asymmetry. Egger's regression test for funnel plot asymmetry resulted in p = 0.188, providing no definitive evidence of small-study effects. The trim-and-fill method imputed two potentially missing studies and adjusted the pooled estimate to 24% (95% CI [13%, 41%]) which remained consistent with the primary analysis result. However, these analyses should be interpreted cautiously, as funnel plot asymmetry tests for proportion meta-analyses can be unstable in the presence of high heterogeneity and small-study effects may also reflect genuine between-study differences in clinical context rather than bias.

***Meta-regression***

An exploratory meta-regression was conducted to examine whether the proportion of non-HCP content (defined as the ratio of non-HCP videos to total videos, where provided) moderated misinformation prevalence. This variable was available in six of the 11 included studies. The unadjusted meta-regression coefficient for non-HCP content proportion was not statistically significant (β = -5.899; p = 0.515) indicating that the proportion of non-HCP videos did not explain a statistically meaningful portion of between-study heterogeneity in this sparse dataset. The unadjusted meta-regression bubble plot is presented in **Supplementary Figure S5**.

***Certainty of Evidence***

Using the GRADE framework adapted for prevalence meta-analyses, the overall certainty of evidence for the pooled misinformation prevalence estimate was rated as 'Very Low'. Starting certainty was 'Low' because all included studies were observational cross-sectional content analyses. Certainty was downgraded for serious inconsistency (given the substantial heterogeneity ($I^2$ = 96.9%) and wide prediction interval [PI 4%-83%]). Imprecision was also a concern because the pooled 95% CI estimate was broad (0.18-0.48), even though certainty has already reached 'Very Low'. Risk of bias was not critical as most studies were rated low risk, with one study rated moderate risk using JBI. We did not identify clear evidence of serious indirectness or publication bias. For the ADHD subgroup specifically, certainty was rated as Low because of low heterogeneity ($I^2$ = 0%) and more consistent estimates, partially offsetting the risk of bias and imprecision concerns. However, that judgment should still be interpreted cautiously because the evidence base is limited to only two studies. The complete GRADE table can be seen in **Supplementary Table S5.**

## DISCUSSION

In this systematic review and meta-analysis, approximately one in three TikTok mental health videos in the included studies contained misleading or misinformation content. This suggests that misinformation is not rare, but a substantial feature of the current mental health information environment on TikTok. However, the high heterogeneity and wide prediction interval indicate that this pooled estimate should not be interpreted as a fixed platform-wide rate. Instead, it is better understood as a summary across studies that differed in topic, sampling strategy, and misinformation classification. The very low certainty of evidence further supports cautious interpretation of the exact pooled estimate.

The highest pooled prevalence was observed in ADHD-related content, and this subgroup also showed the greatest internal consistency (51%; 95% CI [41%, 61%]; $I^2$ = 0%). One likely explanation is that ADHD-related TikTok content may be especially vulnerable to overpathologizing everyday experiences, overgeneralizing nonspecific psychological symptoms, and encouraging viewers to identify with ADHD diagnostic labels on the basis of brief and highly relatable videos.[5,11] Sieferle et al. (2025) explicitly noted that misleading ADHD content often frames common behaviors in the general population as ADHD-specific and may increase perceived personal susceptibility to ADHD among viewers.[5] Yeung et al. (2022) similarly observed that even videos not intended as formal medical education may describe nonspecific symptoms, overgeneralizations, and characterizations about ADHD that are misleading to viewers.[11] Users were also shown to resonate strongly with symptom narratives and sometimes explicitly attribute ADHD to themselves, raising concern that highly understandable but poorly contextualized content may promote false self-identification.[5] This interpretation is also supported by broader psychiatric literature suggesting that online self-diagnosis is growing rapidly and may carry both potential benefits and substantial risks, particularly when brief, emotionally resonant, or poorly contextualized information is consumed without professional evaluation.[3]

ASD-related content also showed a high pooled prevalence of misinformation, but unlike ADHD, the ASD subgroup showed marked heterogeneity (48%; 95% CI [23%, 75%]; $I^2$ = 91.4%). This wider variability likely reflects the fact that autism-related TikTok content is not limited to psychoeducation, but instead spans identity exploration, community building, personal storytelling, advocacy, and clinical information, all of which may be captured differently depending on the sampling frame and coding rules of a given study.[18] Brennan et al. (2025) describe TikTok as a space in which autistic adults seek not only information but also belonging, self-expression, and identity exploration, and they further discuss how some users view TikTok as a mechanism for "platformed diagnosis."[18] Ononuju and Ujari (2025) likewise found that supportive community posts and personal stories were highly prevalent in ASD-related content, while also cautioning that personal narratives may oversimplify complex ASD-related issues and introduce anecdotal bias when treated as generalizable information.[21] Aragon-Guevara et al. (2025) also showed that informational autism videos varied substantially in quality, with a large proportion classified as inaccurate or overgeneralized and with healthcare professional-created videos tending to be more accurate than those from other users.[10] Taken together, these findings suggest that the heterogeneity within the ASD subgroup likely reflects a genuinely diverse content environment, where educational, experiential, and identity-based narratives coexist and are not easily separated by simple misinformation labels.

By contrast, the subgroup of other mental health topics yielded a lower pooled

prevalence, but this should not be overinterpreted as indicating that other mental health content is 'safer' or 'more accurate'. That subgroup combined conceptually distinct topics, including general mental health, OCD, PTSD, NSSI, and mixed psychiatric content, and therefore represents a highly heterogeneous catch-all category by design. For example, Satterfield's NSSI analysis found relatively few videos classified as misleading but also emphasized the dominance of personal-experience content and the limited presence of healthcare professionals, while PTSD and OCD studies focused more directly on symptom characterization and stereotype amplification.[22] The lower pooled prevalence in this subgroup therefore likely reflects topic mixing, divergent coding frameworks, and unexplored contextual factors rather than a uniform reduction in misinformation risk.

An important strength of the present review is that the main result remained stable across multiple robustness checks. Leave-one-out analyses yielded similar pooled estimates, suggesting that no single study drove the overall result. The secondary low-risk-only sensitivity analysis also produced a very similar estimate, supporting the stability of the main finding despite study-level methodological differences. Formal asymmetry testing did not identify clear evidence of small-study effects, although publication bias analyses in meta-analyses of proportions should remain interpretively cautious in the presence of high heterogeneity and bounded outcomes.

Our exploratory meta-regression did not show that the proportion of non-HCP videos explained between-study heterogeneity, but this null finding should be interpreted cautiously. Creator type was reported at the video level in only six of the 11 included studies, and substantial residual heterogeneity persisted even after this moderator was included in the model. However, the absence of a statistically significant moderator effect at the study level does not imply that creator type (whether HCP or non-HCP) is unimportant. Several included studies still suggest descriptively that healthcare professionals tend to produce more accurate or useful content overall, including the ADHD study by Yeung et al. (2022) and the autism study by Aragon-Guevara et al. (2025), although these same studies also indicate that HCP status does not guarantee accuracy and that non-HCP creators dominate the content ecosystem.[10,11] This pattern is also relevant because among the six studies with video-level creator reporting in our sample, non-HCP videos accounted for more than 90% of all sampled content, meaning that even small differences in content quality may have outsized public impact due to the sheer volume of non-HCP material.

These findings have several clinical implications. First, clinicians should recognize that patients may arrive with symptom beliefs, diagnostic labels, and treatment expectations shaped through TikTok videos rather than formal assessment or evidence-based guidance. This is especially relevant for ADHD and autism, where brief, emotionally resonant, and highly relatable content may encourage self-identification before professional evaluation. More broadly, young people commonly seek help online for mental health concerns because online environments offer privacy, accessibility, immediacy, and lower perceived stigma than offline services, which means that inaccurate social media content may interact directly with real-world help-seeking pathways.[1] Online self-diagnosis in psychiatry is also becoming more common, which further increases the importance of asking patients where their mental health information is coming from and correcting misconceptions without dismissing the underlying reasons they sought information online in the first place.[3]

The public health implications are similarly important. Our findings support the need for targeted digital mental health literacy interventions, particularly for higher-risk topic

areas such as ADHD and ASD. This recommendation is supported by a recent systematic review and meta-analysis showing that digital mental health literacy interventions can improve mental health outcomes and related literacy constructs, especially when psychoeducation is paired with active, engaging components.[23] There is also a need for more visible, accessible, and platform-appropriate evidence-based psychiatric content, because adolescents' trust in health information on social media depends on a complex interplay between trust in the platform, the creator, and the perceived credibility of the content itself.[24] In practical terms, this means that simply having accurate information available online may be insufficient if it is not discoverable, engaging, and presented in a format that competes with highly resonant but misleading content.[23,24]

Several limitations should be considered when interpreting these findings. First, all included studies were cross-sectional content analyses, so this review estimates prevalence rather than direct clinical or behavioral impact. We cannot determine whether exposure to misleading TikTok content directly leads to self-diagnosis, delays treatment, or changes in help-seeking behavior. Future studies should examine these downstream outcomes more directly. Second, heterogeneity remained very high, which limits the precision and generalizability of the pooled estimate. However, this variability is expected with different mental health topics, coding frameworks, and sampling approaches, and our subgroup analyses suggest that content topic is one meaningful contributor to this between-study variation. Third, definitions of "misleading" or "misinformation" varied across studies, as did coding rules, instruments, and decision thresholds. More standardized and transparently reported classification frameworks would be valuable to strengthen comparability across future reviews. Fourth, many studies did not report video-level creator type, particularly HCP versus non-HCP status, and only a minority formally examined whether creator background was associated with content quality, despite some evidence that HCP-created videos may be more useful or accurate overall. This limited our ability to evaluate creator type as a moderator in a more robust way, and future studies should report creator type systematically at the video level and directly compare content quality across creator categories. Fifth, sampling strategies also differed substantially, including hashtag-based searches, top-liked or top-viewed videos, and more algorithm-informed approaches. All of these may capture different parts of the TikTok content database and may influence the estimate. Sixth, TikTok content is dynamic and platform algorithms change rapidly over time, meaning that these findings reflect sampled time points rather than a fixed platform-wide prevalence. Seventh, most included studies focused on English-language or Western-context content, which may limit generalizability to other languages and cultural settings. Finally, the overall certainty of evidence was very low, driven mainly by inconsistency and imprecision, so the pooled estimate should be interpreted as directionally informative rather than as a precise benchmark. At the same time, the consistency of the leave-one-out and sensitivity analyses supports the robustness of the overall findings that misinformation in TikTok mental health content is not rare.

## CONCLUSION

Misinformation in TikTok mental health videos appears to be relatively common in the currently available literature, with an overall pooled prevalence of approximately 31%. Prevalence varied meaningfully by topic, with the highest pooled estimates observed in ADHD and ASD related content, but those comparisons should be interpreted cautiously because heterogeneity remained high. Although sensitivity analyses yielded similar results, confidence in the exact pooled

estimate remains limited because the overall certainty of evidence was very low. These findings support the need for stronger digital mental health literacy, greater visibility of evidence-based psychiatric content on social media, and increased clinician awareness that patients' symptom beliefs and help-seeking decisions may be influenced by TikTok-derived information. Future research should use clearer and more standardized definitions of misinformation, report HCP versus non-HCP creator type at the video level, apply more consistent sampling methods, and should examine topic-specific mechanisms, geographic variation, and the real-world effects of TikTok mental health content on self-diagnosis, help-seeking, and treatment engagement.

## ACKNOWLEDGMENTS

None.

## CONFLICT OF INTERESTS

The authors declare no conflict of interest.

## FUNDINGS

This study received no external funding.

## REFERENCES

1. C. Pretorius, D. Chambers, and D. Coyle, "Young People's Online Help-Seeking and Mental Health Difficulties: Systematic Narrative Review," (in en), *Journal of Medical Internet Research,* vol. 21, no. 11, p. e13873, 2019/11/19/ 2019, doi: 10.2196/13873.
2. J. A. Naslund, A. Bondre, J. Torous, and K. A. Aschbrenner, "Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice," *J Technol Behav Sci,* vol. 5, no. 3, pp. 245-257, Sep 2020, doi: 10.1007/s41347-020-00134-x.
3. S. Monteith, T. Glenn, J. R. Geddes, P. C. Whybrow, E. D. Achtyes, and M. Bauer, "Implications of Online Self-Diagnosis in Psychiatry," (in en), *Pharmacopsychiatry,* vol. 57, no. 02, pp. 45-52, 2024/03// 2024, doi: 10.1055/a-2268-5441.
4. B. Rohl *et al.*, "Posttraumatic Stress Disorder Content on TikTok: Cross-Sectional Analysis of Popular #PTSD Posts," (in en), *Online Journal of Public Health Informatics,* vol. 17, pp. e71209-e71209, 2025/09/10/ 2025, doi: 10.2196/71209.
5. K. Sieferle, T. Guidi, F. Dorr, and E. M. Bitzer, "Quality and Perception of Attention-Deficit/Hyperactivity Disorder Content on TikTok: Cross-Sectional Study," (in en), *JMIR Infodemiology,* vol. 5, pp. e75973-e75973, 2025/11/13/ 2025, doi: 10.2196/75973.
6. C. E. Kirkpatrick and L. L. Lawrie, "TikTok as a Source of Health Information and Misinformation for Young Women in the United States: Survey Study," *JMIR Infodemiology,* vol. 4, p. e54663, May 21 2024, doi: 10.2196/54663.
7. V. Suarez-Lledo and J. Alvarez-Galvez, "Prevalence of Health Misinformation on Social Media: Systematic Review," (in en), *Journal of Medical Internet Research,* vol. 23, no. 1, p. e17187, 2021/01/20/ 2021, doi: 10.2196/17187.
8. R. Turuba *et al.*, "Do you have depression? A summative content analysis of mental health-related content on TikTok," (in en), *DIGITAL HEALTH,* vol. 11, p. 20552076241297062, 2025/01// 2025, doi: 10.1177/20552076241297062.
9. I. Starvaggi, C. Dierckman, and L. Lorenzo-Luaces, "Mental health misinformation on social media: Review and future directions," *Curr Opin Psychol,* vol. 56, p. 101738, Apr 2024, doi: 10.1016/j.copsyc.2023.101738.
10. D. Aragon-Guevara, G. Castle, E. Sheridan, and G. Vivanti, "The Reach and Accuracy of Information on Autism on TikTok," (in en), *Journal of Autism and Developmental Disorders,* vol. 55, no. 6, pp. 1953-1958, 2025/06// 2025, doi: 10.1007/s10803-023-06084-6.
11. A. Yeung, E. Ng, and E. Abi-Jaoude, "TikTok and Attention-Deficit/Hyperactivity Disorder: A Cross-Sectional Study of Social Media Content Quality," (in en), *The Canadian Journal of Psychiatry,* vol. 67, no. 12, pp. 899-906, 2022/12// 2022, doi: 10.1177/07067437221082854.
12. A. Hudon *et al.*, "Navigating the Maze of Social Media Disinformation on Psychiatric Illness and Charting Paths to Reliable Information for Mental Health Professionals: Observational Study of TikTok Videos," (in en), *Journal of Medical Internet Research,* vol. 27, pp. e64225-e64225, 2025/06/18/ 2025, doi: 10.2196/64225.
13. M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *International journal of surgery (London, England),* vol. 88, no. NA, pp. 105906-105906, 2021, doi: 10.1016/j.ijsu.2021.105906.
14. D. F. Stroup *et al.*, "Meta-analysis of observational studies in epidemiology: a

proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group," (in en), *JAMA,* vol. 283, no. 15, pp. 2008-12, Apr 19 2000, doi: 10.1001/jama.283.15.2008.
15. S. Moola *et al.*, "Chapter 7: Systematic reviews of etiology and risk," in *JBI Manual for Evidence Synthesis*: JBI, 2020.
16. J. Hartung and G. Knapp, "A refined method for the meta-analysis of controlled clinical trials with binary outcome," *Stat Med,* vol. 20, no. 24, pp. 3875-89, Dec 30 2001, doi: 10.1002/sim.1009.
17. H. J. Schünemann *et al.*, "GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence," (in en), *Journal of Clinical Epidemiology,* vol. 111, pp. 105-114, 2019/07// 2019, doi: 10.1016/j.jclinepi.2018.01.012.
18. E. Brennan, L. A. Lampinen, H. Paek, X. Wang, H. Romano, and V. H. Bal, "Deconstructing Information About Autism Diagnosis in Adults on TikTok: A Cross-Sectional, Descriptive Content Analysis," (in en), *Journal of Autism and Developmental Disorders,* 2025/11/08/ 2025, doi: 10.1007/s10803-025-07123-0.
19. E. Brown *et al.*, "A cross-sectional analysis of TikTok autism spectrum disorder content quality," (in en), *Emerging Trends in Drugs, Addictions, and Health,* vol. 4, p. 100150, 2024/12// 2024, doi: 10.1016/j.etdah.2024.100150.
20. M. Fitzpatrick, A. S. Moore, S. A. Kichuk, C. Pittenger, and B. A. Zaboski, "#OCD: A Content Analysis of Obsessive-Compulsive Disorder Stereotype Amplification and Misinformation on TikTok," (in en), *Cyberpsychology, Behavior, and Social Networking,* vol. 28, no. 9, pp. 607-615, 2025/09// 2025, doi: 10.1177/21522715251370135.
21. U. A. Ononuju and C. A. Ujari, "Stigma and Misinformation About Autism Spectrum Disorder (ASD) on Tiktok and Instagram: Content Analysis Using #ASD, #Autism and #ASDinfo," (in en), *Journal of Autism and Developmental Disorders,* 2025/09/30/ 2025, doi: 10.1007/s10803-025-07057-7.
22. S. Satterfield, "Understanding mental health content on TikTok: exploring misinformation and stereotypes among non-suicidal self-injury TikTok videos : a thesis in Psychology," Master of Arts, University of Massachusetts Dartmouth, Dartmouth, Massachusetts, United States, 2024. [Online]. Available: https://repository.lib.umassd.edu/esploro/outputs/graduate/9914424896601301
23. G. Yeo, S. M. Reich, N. A. Liaw, and E. Y. M. Chia, "The Effect of Digital Mental Health Literacy Interventions on Mental Health: Systematic Review and Meta-Analysis," (in en), *Journal of Medical Internet Research,* vol. 26, p. e51268, 2024/02/29/ 2024, doi: 10.2196/51268.
24. J. L. Freeman, P. H. Y. Caldwell, and K. M. Scott, "How Adolescents Trust Health Information on Social Media: A Systematic Review," (in en), *Academic Pediatrics,* vol. 23, no. 4, pp. 703-719, 2023/05// 2023, doi: 10.1016/j.acap.2022.12.011.