Which Social Media Platforms Provide the Most Informative Data for Monitoring the Opioid Crisis?

Kristy A. Carpenter[1], Anna T. Nguyen[2], Delaney A. Smith[3], Issah A. Samori[4], Keith Humphreys[5,9], Anna Lembke[5], Mathew V. Kiang[2], Johannes C. Eichstaedt*[6], Russ B. Altman*[1,4,7,8]

Departments of [1]Biomedical Data Science, [2]Epidemiology and Population Health, [3]Biochemistry, [4]Bioengineering, [5]Psychiatry and Behavioral Sciences, [6]Psychology, [7]Genetics, and [8]Medicine, Stanford University, Stanford, CA 94305, USA

*[9]Veterans Affairs Health Care System, Palo Alto, CA, 94304, USA; *equal contribution*

ABSTRACT

Background and Aims

Social media can provide real-time insight into trends in substance use, addiction, and recovery. Prior studies have leveraged data from platforms such as Reddit and X (formerly Twitter), but evolving policies around data access have threatened their usability in opioid overdose surveillance systems. Here, we evaluate the potential of a broad set of platforms to detect emerging trends in the opioid crisis.

Design

We identified 72 online platforms with a substantial global user base or prior citations in opioid-related research. We evaluated each platform's fit with our definition of social media, size of North American user base, and volume of opioid-related discourse. We created a shortlist of 11 platforms that met our criteria. We documented basic characteristics, volume and nature of opioid discussion, official policies regulating drug-related discussion, and data accessibility of shortlisted platforms.

Setting

USA and Canada.

Measurements

We quantified the volume of opioid discussion by number of platform-specific Google search hits for opioid terms. We captured informal language by including slang generated using a large language model. We report the number of opioid-related hits and proportion of opioid-related hits to hits for common nouns.

Findings

We found that TikTok, YouTube, and Facebook have the most potential for use in opioid-related surveillance. TikTok and Facebook have the highest relative amount of drug-related discussions. Language on TikTok was predominantly informal. Many platforms offer data access tools for research, but changing company policies and user norms create instability. The demographics of users varies substantially across platforms.

Conclusions

Social media data sources hold promise for detecting trends in opioid use, but researchers must consider the utility, accessibility, and stability of data on each platform. A strategy mixing several platforms may be required to cover all demographics suffering in the epidemic.

INTRODUCTION

Real-time tracking of substance use trends is key to understanding epidemics of addiction and overdose, including the ongoing North American opioid crisis. Heterogeneity in prominent substances, drug availability, and usage patterns across the epidemic's distinct waves require different intervention strategies, and it is critical that policy makers are sensitive to these changing dynamics. Currently, academic and government surveys are standard practice for tracking usage patterns. However, difficulties with self-reported opioid use limit the reliability of these surveys, leading to inaccurate estimates of the prevalence and nature of opioid use (1) Moreover, surveys can take months or years to complete and results may not reflect quickly evolving trends in opioid use.

Social media platforms are a promising alternative data source which might be used to track patterns of opioid use. Unlike official overdose statistics, social media provides real-time, high-volume, and widely accessible streams of information (2,3) These platforms capture unfiltered experiences across diverse populations, including encounters with illicit substances that might not otherwise be disclosed. (4) Thus, social media can help researchers identify and geolocate emerging drugs and track practices among people who use opioids (PWUOs). Understanding these patterns could help policymakers anticipate future hotspots for overdoses.

Various models have been trained on social media data to predict individual- and population-level outcomes related to opioid use. By focusing on media from individual users, some researchers have attempted to detect opioid misuse (4), flag indicators of addiction (5), and assess individual risk of relapse (6). A larger body of work is dedicated to tracking population trends in opioid misuse (7) and opioid-related mortality (8,9). Some evidence suggests that language use on these platforms may be more predictive of trends in county-level deaths than factors such as demographics, healthcare access, and physical pain. (8) Other work has triangulated social media data with other surveillance datasets (such as emergency department admissions data) to develop more holistic models.(10)

Most of this work uses only a few well-known platforms, chiefly X (formerly Twitter) and Reddit. However, access to these data sources is highly dependent on corporate decisions that inadvertently impact researchers. For example, Pushshift (11) was a popular research tool for extracting Reddit posts, but when Reddit started to limit API requests for third-party data access in 2023, it effectively disabled Pushshift. A similar phenomenon occurred with the X API contemporaneously. As digital platforms evolve, it is critical for researchers to understand the scope of available datasets and whether there are feasible alternatives if a dataset becomes inaccessible. Additionally, different social media platforms have different constituent demographics; no individual platform fully represents the population at large. Therefore, an effort combining multiple platforms, rather than relying on one or two, may be better able to capture trends in this global crisis.

There are currently no comprehensive studies evaluating the nature, volume, and quality of opioid-related discussions across social media platforms. Although there are several systematic literature reviews of this field (3,12), a direct evaluation of a broad range of social media

platforms is needed for researchers seeking to leverage social media platforms to explore opioid use.

Here, we identified social media platforms that may be suitable for text-based opioid research and characterized the nature and volume of their opioid-related discussions. We created a shortlist of eleven platforms for which we investigated censorship policies, data accessibility, and prior use in opioid research. We present our findings and discuss the utility, availability, and stability of these platforms for the purpose of informing design of a social media early-warning system for trends in the opioid epidemic. Code for the analyses contained here is available at https://github.com/kristycarp/opioid-social-media-platforms.

METHODS

Identifying social media platforms

We compiled a comprehensive list of social media platforms that may contain content related to opioid use, substance use disorders, and/or addiction treatment and recovery.

We included all platforms with more than 100M monthly active users worldwide (13). We also included platforms that the Stanford Digital Economy Lab identified as "digital goods" due to their widespread use and relevance (14, personal communication).

We included platforms previously analyzed for surveillance of illicit drug use (15–17) and general adverse drug reactions (15,18–24). We also chose to include several forums dedicated to discussion of drug use and recovery, some of which have been previously used for research on illicit drug use (25). We included popular online marketplaces due to their potential for opioid-related transactions. (26–31)

Creating platform shortlist

In the first round of platform evaluation, we determined if each platform 1) had an active web domain or mobile application, 2) met the Knight First Amendment Institute definition of social media (32), 3) had a primary function other than private messaging, 4) was based in the US/Canada or had English as the default language for US-based users, and 5) returned more than 25,000 Google search hits for a set of opioid-related terms. Platforms that met all five criteria were shortlisted for detailed evaluation. A detailed description of the shortlisting process can be found in **Supplement S1.1**.

Measuring the volume of opioid-related discussion

We used the number of hits returned by Google search results when querying opioid-related terms to approximate the amount of accessible opioid-related discussion on each platform. We used this proxy method as a fast alternative to acquiring access to all content on each platform. We formatted the Google queries so as to only yield results from the specific platform's domain, *e.g.* when assessing Facebook, we limited results to only be those from facebook.com. We assembled three lists of opioid-related terms: "formal," "informal," and "algospeak."

The “formal” opioid term list includes official generic names and common brand names of opioids. These terms are the same as those used to initially select for platforms with high opioid discussion in the platform exclusion phase (**Supplement S1.1**, **Supplement S1.2**).

The “informal” opioid term list includes slang terms, misspellings, and other brand names. Because language on social media is often informal, terms like these are important to include to capture the full scale of opioid-related discussion on a given platform. We previously found that GPT-3 (33) is able to quickly and accurately generate slang for drugs of addiction at scale (34). We took validated GPT-3–generated slang terms for five opioids prominent in the opioid epidemic (codeine, fentanyl, morphine, oxycodone, and oxymorphone) which were precomputed by Carpenter *et al*, and removed terms that had common non-drug meanings in addition to drug-related meanings (*e.g.* “fenty” is a common slang term for fentanyl, but also the name of a cosmetics brand). To keep the scale similar to that of the formal term analysis, we reduced the list to the top 20 terms most commonly generated by GPT-3 (**Supplement S1.3**).

“Algospeak” is a phenomenon in which users of social media platforms purposefully alter the spelling of terms in order to evade automated censorship. (35) For example, a poster on social media may refer to fentanyl as “f3ntanyl” to avoid being banned. We noticed that the set of terms generated by GPT-3, while still relevant to present-day casual discussion, did not include algospeak. Therefore, we created an additional “algospeak” term list using GPT-4 in order to capture censorship-evading language on social media (**Supplement S1.4**). A description of how we created this list is found in **Supplement S1.5**.

We used the Google Search API with default parameters to query for the number of English-language hits specific to a given social media website for each of the formal opioid terms, informal opioid terms, and algospeak opioid terms. We tabulated the total number of hits per list to quantitatively estimate the total volume of opioid discussion on each of the shortlisted platforms. In addition, we compared the raw number of opioid-related hits to the number of hits returned for a queries of “household terms”, chosen from Corpus Of Contemporary American English to represent the most common nouns used in the English language. (**Supplement S1.6**) This normalized ratio allowed us to assess how prominent opioid discussion is relative to other content on the platform, regardless of the total number of posts on the platform; platforms with large ratios may be used for opioid discussions more than platforms with small ratios. For each of our three term lists, we calculated a per-platform normalized ratio, defined as the total number of hits for all terms in the list over the total number of hits for all household terms for a given platform, scaled by 100,000 for easier visualization (**Equations S1-3**). In the following analyses, we add informal and algospeak normalized ratios together as the algospeak terms can be considered a special case of informal terms.

Content restrictions and censorship policies

We analyzed the content restriction and censorship policies of social media sites by directly referencing the sites’ terms of use and user agreement web pages.

Evaluating data accessibility for academic research purposes

We sourced information regarding data accessibility for research purposes from each site's respective publicly-available information regarding research and API programs.

Assessing prior use of platforms in research literature

We conducted PubMed searches to identify literature relevant to each of the shortlisted platforms that focuses on applications to the opioid epidemic. We searched for articles with the name of the social media platform in the title or abstract (for less commonly studied social media platforms, this criterion was broadened to presence anywhere in the article) and either "opioids", "opioid", "opiates", or "opiate" in the title or abstract.

Example: (twitter[Title/Abstract]) AND (opioids[Title/Abstract] OR opioid[Title/Abstract] OR opiates[Title/Abstract] OR opiate[Title/Abstract])

RESULTS

Identifying social media platforms

We took the union of the platforms found from the various sources listed in **Methods** to create a superset of 72 candidate platforms (**Supplement Table 1**).

Creating platform shortlist

We applied our shortlist evaluation criteria to the 72 candidate platforms in an iterative process (**Figure 1**). Our full assessment of these platforms is included in **Table S2**.

Of the original 72 platforms, there were 4 platforms that were not active as of July 2023. There were 21 active platforms that did not meet our definition of social media and 11 active platforms that were primarily a private messaging platform. We found that 11 of the active, social media sites were not based in the United States or did not have English as the default language for a user in the United States. Among the remaining sites, there were 14 platforms that returned fewer than 25,000 query results on the selected opioid keywords.

After applying these criteria, we were left with 11 (15.3%) platforms for further evaluation: Bluelight, drugs-forum.com, Facebook, Instagram, LinkedIn, Pinterest, Reddit, TikTok, Tumblr, X (formerly Twitter), and YouTube.

Characteristics of these platforms are described in **Table 1**.

Measuring the volume of opioid-related discussion

The type and volume of publicly available data related to opioids varied widely across platforms (**Figure 2, Figure 3, Figure S1, Table S3, Table S4**). We observed the highest total volume of opioid-related discussion – in terms of both formal and informal term hits – on YouTube and Facebook (**Figure 2a, Figure S1, Table S3**). These were followed by LinkedIn, TikTok, and

Reddit. By far, the platforms with the lowest total volume of opioid-related discussion were Bluelight and drugs-forum.

Most platforms followed the same linear relationship between the amount of formal opioid terms and the amount of informal (including algospeak) opioid terms. Some slight outliers were TikTok and Instagram, which both skewed toward more informal term hits, and LinkedIn, Reddit, and X, which all skewed toward more formal term hits.

Examining the relationship between the total number of informal hits versus the total number of algospeak hits revealed largely similar trends as the comparison between formal hits and informal hits (**Figure 2b**). The outliers outside of the trendline on this plot reveal which platforms have more algospeak, indicating a response to censorship. TikTok by far had the most amount of algospeak. Instagram and X also skewed more toward algospeak than to informal, non-algospeak terms. LinkedIn and Reddit both skewed toward more informal term hits than algospeak term hits.

While having less total drug-related discussion overall, Bluelight and drugs-forum had dramatically higher rates of drug discussion relative to non-drug discussion (**Figure 3a**). This is expected due to their explicit platform focus on drugs. Among the social media platforms with more general scope, TikTok and Facebook had the highest relative amount of drug discussion when compared to non-drug discussion.

We also visualized the relative amounts of informal and algospeak term hits for the 11 platforms (**Figure 3b**). Again, Bluelight and drugs-forum show large separation from the 9 general social media platforms, especially with respect to informal (non-algospeak) terms. When looking at only the 9 general social media platforms, we saw that Facebook and TikTok again were the two platforms with the highest relative amounts of drug discussion with respect to both informal and algospeak term hits.

Content restrictions and censorship policies

All eleven shortlisted platforms make statements in their terms of use or community standards about expectations regarding and moderation of drug-related discussion. High-level classifications of drug-related content policies are shown in **Table 1**. We provide excerpts of each platform's community standards relevant to drug-related discussion in **Supplement S1.7**.

At a minimum, all eleven platforms explicitly state that the sale of illicit substances is not allowed. Facebook, Instagram, TikTok, and YouTube go further to state that posting content related to or promoting recreational use of drugs is prohibited. TikTok additionally underscores that such content is particularly dangerous for young people. These platforms make exceptions for recovery-oriented or educational content. Similarly, LinkedIn prohibits depictions of "drug abuse." Reddit has some communities focused on drug-related discussion, and therefore is more tolerant; the Reddit content moderation system accounts for communities with mature themes, such as drug-related discussion, and places these communities under age restrictions. Bluelight and drugs-forum, as forums dedicated to drug-related discussion, are distinct from the

other, general-purpose social media platforms. Both Bluelight and drugs-forum explicitly state that they are safe spaces for discussion of all aspects of drug use and recovery.

Evaluating data accessibility for academic research purposes

Many platforms provide Application Programming Interfaces (APIs) for data extraction. Generally, APIs are available in tiers where more expensive versions provide more comprehensive access. Research APIs often provide free or heavily discounted access to users who can verify their affiliation with a research institute. In some cases, outside groups have maintained third-party APIs; for example, Pushshift was a free, popular tool to query data from Reddit. We found that smaller drug-focused forums do not have data access APIs or third-party tools that allow for easy data extraction. We provide an overview of data access capabilities in **Table 1** and per-platform details in **Supplement S1.8**. We emphasize that researchers should not violate platform terms of service when collecting data for pharmacovigilance or any other purpose.

Beyond accessing text content of these platforms, a useful early-warning system for trends in the opioid epidemic necessitates geolocatability. The geolocation of social media content can be obtained directly from the platform or, when such metadata is unavailable, inferred.

Some social media platforms explicitly provide geolocation of users or posts. X (formerly Twitter) gives platform users the option to explicitly geotag their tweets. While only a small subset of users opt to do this (under 2% (36)), the fact that this information is present on the platform in some regard means that it can be harnessed by researchers. Many groups have leveraged explicit tweet geotagging for opioid-related research (37–43) and other research areas (44,45). These studies typically only analyze tweets with geotags. It is important to note that it was possible to geotag a tweet with explicit latitude and longitude coordinates only until June 2019; since then, users can only tag tweets with place objects that have coordinate bounding boxes. (36,46,47)

Similarly, Facebook (48), Instagram (49–52), and YouTube (53–55), give users the option to tag their content with geolocations. However, it is important to note that these tags can be used for purposes other than reporting the exact location from where the post was made (56). Facebook additionally has a Data for Good program (57) which includes province-level GPS data from a subset of consenting users; researchers have leveraged this data to study phenomena related to the COVID-19 pandemic (58,59). State *et al.* used provided user and company locations to track migration patterns using LinkedIn (60)].

The geolocatability of TikTok users has been a prominent topic of discussion in the US. (61–63) Like many other prominent social media platforms, TikTok uses user location to personalize their content feed and allows users to optionally add a location tag to their videos (64). Various groups have characterized the privacy aspect of geolocation on TikTok (65,66), but there is little research using tagged locations for geographically-informed analysis. However, Zanettou *et al.* recruited consenting users to donate their geotagged data, which each user can request under the EU's GDPR regulation. (67)

Forums have a unique way of presenting geolocation information. Many online forums allow users to indicate their location in a designated field on their profile. This is required on drugs-forum and optional on Bluelight. These location entries are unstandardized free text and can be anything from cities (*e.g.* “palo alto”, “Palo Alto, CA”), states (*e.g.* “CA, USA”, “California”, “cali”), regions (*e.g.* “NorCal”, “west coast”), countries (*e.g.* “USA”, “U.S.”, “america”), or abstract concepts meant to be jokes (*e.g.* “none of your business”). Researchers have leveraged the location fields on Bluelight and drugs-forum user profiles to geolocate subsets of their data to the county- (68) or country-level (22).

Some other platforms contain similar location fields, though these are less frequently populated than on forums. Schwartz *et al.* processed these optional location entries on X to create the County Tweet Lexical Bank, a dataset of tweets geolocated at the county level. (69) Several groups have leveraged the County Tweet Lexical Bank and datasets assembled by similar methods for opioid-related (8,9,42,70,71) and non-opioid-related (72,73) work. Similarly, other groups have used named entity recognition to extract unambiguous place names within social media text content.(74–77)

A simpler method of acquiring social media content associated with a geolocation is to search for content containing location-related keywords. Several groups have used the hashtag search functionality on TikTok to obtain content relevant to geographic regions (*e.g.* specific countries). (78,79) Similarly, Hu and Conway pulled text from country-specific subreddits as a proxy for geolocation (80); Delbruel *et al.* examined the association between YouTube video tags and geolocation (81); and Goyer *et al.* used keyword search to identify Reddit and X posts relevant to Canada and manually identified Canadian Facebook groups from which to extract content (82).

A key feature of social networks is that users are connected (through “friendship”, “following”, *etc.*) to other users with whom they often share some characteristics or interests in common; this means that sometimes we can infer the location of a user based on the locations of users to whom they are connected. For example, someone who almost exclusively follows people based in New York City on Instagram is likely to also be based in New York City. This line of reasoning has been used extensively for X (83–85)because it depends upon the social network being partially labeled with geolocation.

Examining a social media platform user’s posts beyond the topic of interest and profile information beyond an explicitly stated location can also yield a proxy for geolocation. This strategy is often used for Reddit as it is facilitated by explicit subforum (subreddit) topics and post titles, key features of this platform. Several groups have used the assumption that if a Reddit user frequently posts in a city-specific subreddit, then they are likely to live or spend a significant amount of time in that city (*e.g.* posting frequently in r/sanfrancisco implies living in San Francisco) (2,86,87). Researchers have also searched for posts with the topic of “where are you from?” or instances of the phrase “I live in…” (87,87) and leveraged user “flairs” (tags with small icons that Reddit users can add to their username) (87) to geolocate Reddit users.

Numerous packages and services for predicting geolocation from social media data exist. Free packages used previously for geolocating social media text (7,88) include Carmen (89) and geopy (90); paid services used previously for geolocating social media text (91–93) include Iconosquare (94), Brandwatch (formerly Crimson Hexagon) (95), and Reputation (formerly Nuvi) (96). Beyond these resources, many other groups have created various geolocation inference methods (48,97–100). While these methods can facilitate efficient geotagging of social media data and circumvent the need to create a new geolocation pipeline from scratch, it is important to note that the accuracy of such methods decreases with the passage of time (101).

When all else fails for a given platform, a final strategy to obtain geolocation estimates is to link users to accounts on a different platform that does have geolocation. For example, while Tumblr does not provide any geolocation information, Tumblr users can opt to share their posts to X. Xu *et al.* leveraged this cross-posting to obtain geolocation information for a subset of Tumblr users (77,102).

We summarize viable geolocation inference strategies for each of the shortlisted platforms in the "Example of geolocation inference strategy" column of **Table 1.**

Prior use in research literature

There is wide variation in the presence of each of our 11 shortlisted platforms in the existing literature on text-based analysis of the opioid epidemic. Of all shortlisted social media platforms, X (formerly Twitter) and Reddit were by far the most commonly used in existing literature. The topics of studies using these two platforms range from correlating opioid-related discussion volume and opioid-related overdose death rates (2,7–10,41,42,70,71), characterizing trends and themes in online discussion of OUD and OUD treatment (4,27,37–40,70,74,87,103–121), and characterizing public sentiment towards the opioid epidemic generally (39,82,93,108,122,123). Many research groups have also created models to automatically identify posts on X and Reddit with discussion related to opioids.(42,43,109,124–127) Others have used these platforms to characterize factors that influence opioid use, recovery, and the opioid epidemic generally (6,8,128–132), with particular interest shown to the impact of the COVID-19 pandemic (103,120,133–139) and co-use between opioids and other drugs (140,141).

The platforms with the next most volume of prior work were Instagram and Facebook. Researchers used these data to examine various aspects of opioid use (142–144), to characterize content related to drug sales (16,145–147), to identify emerging psychoactive substances (121), and to study public awareness of and reactions to the opioid epidemic generally (82,148,149). Though Instagram is an image-focused platform, nearly all studies primarily used text from captions and comments as the basis of their analysis. Facebook text data was largely obtained from public groups and pages.

YouTube, Tumblr, and drugs-forum.com all have seen modest usage in text-based research focused on the opioid epidemic. YouTube comments have been analyzed with NLP techniques to characterize general sentiments toward the opioid epidemic (150). YouTube and drugs-forum

were both used in a study characterizing misinformation related to OUD medications (105). Researchers searched YouTube and Tumblr, among other platforms, for content about emerging psychoactive substances (121). Others have used Tumblr text data for detection of illicit opioid sales (16,151,152). Researchers have found correlations between trends in drugs-forum.com contents and NSDUH survey data (22). Drugs-forum.com has also been used for analysis of phenomena for other drugs (153,154).

We did not find any prior work using the remaining four shortlisted platforms (Pinterest, TikTok, Bluelight, and LinkedIn) for text-based analysis focused on the opioid epidemic. However, researchers have previously used Pinterest and TikTok to study portrayals of drugs other than opioids (155–162). Researchers have used Bluelight in the context of recruiting participants for studies related to opioid use (163,164). Our PubMed query returned no results related to opioid research for LinkedIn.

DISCUSSION

We found that all eleven shortlisted platforms contain notable volumes of opioid-related discussion. Beyond total volume of opioid-related content, other factors affecting utility in an opioid epidemic surveillance system include degree of censorship, user base demographics, and geolocatability. However, researchers must consider the accessibility and stability of these data sources before incorporating them in public health research or surveillance. While we highlight APIs and other data accession methods for shortlisted platforms, not all drug-related discussion or user metadata may be available through official APIs. Recently, some platforms have shifted from freely available APIs to paywalled versions, which may be cost-prohibitive for longitudinal large-scale projects. Quickly evolving trends in user language and norms may also require workflow updates to ensure that relevant discussions are being captured, as we show here with the emergence of algospeak. While we present all eleven shortlisted platforms as promising sources for future surveillance methods, we highlight TikTok, Facebook, and YouTube as underutilized platforms with significant opioid-related content.

Platform users, contents and dynamics affect research utility

Platforms that allow open discussion of drug-related topics (*e.g.* Bluelight, drugs-forum) or that grant pseudo-anonymity (*e.g.* X, Tumblr) afford greater freedom in discussing stigmatized topics. This increases the amount of opioid-related signal on a platform, as we observed when examining volume of opioid terms relative to common household terms. Conversely, platforms with high levels of moderation may spur increased algospeak usage. This phenomenon has been described in the literature with respect to TikTok (35,165,166), corroborating our finding of high levels of algospeak on the platform. It is therefore imperative that future work with TikTok or similar platforms accounts for opioid-related algospeak in order to fully capture relevant content.

When selecting platforms for a social media based surveillance system of the opioid epidemic, it is important to consider the demographics of each platform's user base. For example, the user base for TikTok skews to younger age groups. In a Statista survey, 67% of respondents aged 18-19 and 56% of respondents aged 20-29 reported TikTok use, compared to 38% of 40-49

year olds (167). Individuals who initiate opioid use at a younger age are more susceptible to substance use disorders (168); monitoring discussions around opioid use in these age groups could help inform preventative programs that target young individuals. By contrast, Facebook skews slightly older, with 75% of 30-49 year olds reporting that they use the platforms as opposed to 67% of 18-29 year olds (169). Platform demographics also vary by gender. A greater proportion of women than men report using Facebook, Instagram, TikTok, and Pinterest, whereas a greater proportion of men than women report using X and Reddit. (169) Given the broad demographics of the opioid epidemic, a strategy of following multiple social media platforms in order to monitor all relevant demographics may be required.

Most prior work in social media pharmacovigilance has been conducted using X and Reddit. This is almost certainly because of their high volume of content (both in total and drug-related), their formerly freely accessible APIs, and their inferrable or explicitly-provided geolocation data. However, other platforms used less in research share these attributes. We have outlined geolocation inference strategies for all 11 shortlisted platforms. Facebook, YouTube, TikTok, and LinkedIn emerged as platforms with high total volumes of opioid-related discussion but relatively low prior use in the literature. According to Pew Research Center, Facebook and YouTube are the most-used online platforms by Americans as of January 2024 (169), and TikTok has emerged as a popular platform for youth. We found that LinkedIn's high volume of opioid-related discussion was the result of people sharing research and educational resources related to opioids on the platform. At a minimum, these four platforms may provide insight supplementary to that from X and Reddit.

API access to social media data can be limited

It is not always possible to retrieve all data collected by a social media platform through its API. For example, many social media platforms infer geolocation and demographic information from user activity, but will typically omit such identifiers from available datasets. Additionally, some platforms limit the scope of posts that can be accessed and only allow queries from a small subsample of data rather than the full historical archive. Alternatives to APIs for obtaining data include manual search, web scraping of public webpages, or soliciting donation of private data directly from users, which have seen prior use (82,144,148,149,170) but these methods are more time intensive. We remind the reader that web scraping and third-party APIs may be against a platform's terms of service and that researchers should always review the terms and conditions of data use provided by platforms of interest.

Notably, many platforms allow users to post private content that is not broadly viewable. Such private content cannot be obtained through APIs or web scraping. It is plausible that users discuss opioid use on social media among their private networks, but share fewer or more filtered accounts in the public sphere. If there are substantial differences in users that engage in public versus private discourse, data will be subject to selection biases and negatively impact downstream models. On platforms with largely pseudo-anonymous users (*e.g.* Reddit, Bluelight) or where "burner" accounts are common (*e.g.* X), private posts may be less prevalent, making the dataset more representative of the platform's overall user base.

Changing business models affect platform availability and stability
The policies and practices of social media companies are rapidly evolving, leading to unstable data access from these platforms for research. Notably, we found that Reddit was one of the most commonly used platforms in prior research. This was likely mediated by the availability of Pushshift, a third-party tool that allowed researchers to easily access Reddit's post archives. However, efforts to further monetize the primary Reddit API effectively disabled Pushshift and limited access to these data by researchers.

Data instability could pose a major challenge to deploying a reliable surveillance system for opioid overdoses or conducting longitudinal studies. Surveillance systems that rely on social media data should place safeguards for scenarios where updated data is unavailable. Incorporating other data sources, such as emergency room admissions or wastewater monitoring, could make systems more robust to gaps in data access. Projecting overdoses further into the future, at the cost of larger predictive error, could also prevent gaps in hotspot detection. For future research, it would be prudent to work directly with these companies to establish current regulatory policies and opportunities for data access.

Similarly, changes in how users interact with platforms may affect how overdose hotspots can be anticipated from social media data. Emerging slang or algospeak might not be reflected in a pre-trained model, and rising discussions with these terms would not translate to predicted increases in overdoses. An effective warning system would need to be regularly maintained to reflect digital trends.

Limitations
Our study is subject to limitations that are inherent to research that uses social media data. The extent to which platforms are willing to share data for research is constantly evolving. Privacy and intellectual property concerns have pushed platforms to take a more conservative stance with data sharing. We were not able to access the full text corpus of any platform, but were able to use Google search queries as a proxy measure of how much discussion actually takes place. In practice, platforms may only provide subsamples of their text data to researchers and not all data may be accessible for studies.

The selection of specific opioid-related terms is another limitation in how we quantify opioid-related discussion. There are many phrases that have dual meanings, so we attempted to select only terms that are used exclusively to refer to opioids to avoid the false detection of opioid-related discussion. Additionally, informal language is constantly evolving and the terms used here may not be in use in the future. Here, we detailed a procedure that used generative models to create a list of terms; we believe that these methods could be used to identify updated informal language around opioids at future points in time.

We used Google search queries as a proxy estimate of the overall volume of opioid-related discussion. However, these estimates may be influenced by how Google indexes web pages. Platforms owned by Google, such as Youtube, may be overrepresented in their search results. Other platforms may restrict the access to unregistered users, which inadvertently impacts how

Google can process their webpages.(171) Similarly, Google search does not return private content that has been restricted by users. We hope that our estimates provide a crude estimate of how much opioid-related discussion is on each site, but we caution that the amount of data that can actually be accessed by researchers may vary.

We are aware that a key limitation of our analysis is its explicit focus on English text data on platforms primarily serving North America. Our choice to focus on North America was driven by the severity of the opioid addiction and overdose in the United States and Canada. However, OUD is a global problem and analysis beyond North America is needed. We chose to focus on English text and platforms with English content for our own ease of analysis as English-speaking researchers. But according to the U.S. Census Bureau, over 20% of Americans speak a language other than English at home (172), meaning that to fully characterize the opioid epidemic, text-based research must be done on other languages beyond just English. We encourage future research in leveraging social media platforms for pharmacovigilance of the opioid epidemic in regions other than North America and languages other than English for more equitable and complete assessment of this worldwide crisis.

CONCLUSION

Social media data signals correlate with trends in the ongoing opioid epidemic. Although existing research has already leveraged social media platforms to analyze such phenomena, most research has focused on only a few of the social media platforms currently in use. We have characterized the utility and accessibility of platforms with potential for monitoring opioid-related discussion to motivate future research and give a more complete perspective of emerging trends in the opioid epidemic.

ACKNOWLEDGEMENTS

The authors would like to thank Salvatore Giorgi, Shashanka Subrahmanya, Aadesh Salecha, and Rhana Hashemi for insightful conversations and assistance. This work supported by: NIH DA057598; Microsoft Accelerating Foundation Models Research Initiative; NSF GRFP DGE-1656518 to KAC; NIH NHLBI T32HL151323 to ATN; DAS is supported by the Stanford Biochemistry Department and NSF GRFP 2019286895; Stanford Summer First Fellowship to IAS; Sarafan ChEM-H CBI Program Award to IAS; KH is supported by a Senior Research Career Scientist Award (RCS 04-141-3) from the Department of Veterans Affairs Health System Research Service; JCE, RBA, AL are supported by the Stanford Institute for Human-Centered AI; RBA is supported by the Chan Zuckerberg Biohub.

REFERENCES

1. Palamar JJ. Barriers to accurately assessing prescription opioid misuse on surveys. Am J Drug Alcohol Abuse. 2019;45(2):117–23.
2. Lavertu A, Hamamsy T, Altman RB. Monitoring the opioid epidemic via social media discussions [Internet]. medRxiv; 2021 [cited 2024 Mar 14]. p. 2021.04.01.21254815. Available from: https://www.medrxiv.org/content/10.1101/2021.04.01.21254815v1
3. Rutherford BN, Lim CCW, Johnson B, Cheng B, Chung J, Huang S, et al. #TurntTrending: a systematic review of substance use portrayals on social media platforms. Addict Abingdon Engl. 2023 Feb;118(2):206–17.
4. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An Exploration of Social Circles and Prescription Drug Abuse Through Twitter. J Med Internet Res. 2013 Sep 6;15(9):e189.
5. Fan Y, Zhang Y, Ye Y, li X, Zheng W. Social Media for Opioid Addiction Epidemiology: Automatic Detection of Opioid Addicts from Twitter and Case Studies. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management [Internet]. New York, NY, USA: Association for Computing Machinery; 2017 [cited 2024 Mar 14]. p. 1259–67. (CIKM '17). Available from: https://dl.acm.org/doi/10.1145/3132847.3132857
6. Yang Z, Nguyen L, Jin F. Predicting Opioid Relapse Using Social Media Data [Internet]. arXiv; 2018 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/1811.12169
7. Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media. J Med Toxicol. 2017 Dec 1;13(4):278–86.
8. Giorgi S, Yaden DB, Eichstaedt JC, Ungar LH, Schwartz HA, Kwarteng A, et al. Predicting U.S. county opioid poisoning mortality from multi-modal social media and psychological self-report data. Sci Rep. 2023 Jun 3;13(1):9027.
9. Matero M, Giorgi S, Curtis B, Ungar LH, Schwartz HA. Opioid death projections with AI-based forecasts using social media language. Npj Digit Med. 2023 Mar 8;6(1):1–11.
10. Sumner SA, Bowen D, Holland K, Zwald ML, Vivolo-Kantor A, Guy GP Jr, et al. Estimating Weekly National Opioid Overdose Deaths in Near Real Time Using Multiple Proxy Data Sources. JAMA Netw Open. 2022 Jul 21;5(7):e2223033.
11. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The Pushshift Reddit Dataset. Proc Int AAAI Conf Web Soc Media. 2020 May 26;14:830–9.
12. Conway M, Hu M, Chapman WW. Recent Advances in Using Natural Language Processing to Address Public Health Research Questions Using Social Media and ConsumerGenerated Data. Yearb Med Inform. 2019 Aug;28(1):208–17.
13. List of social platforms with at least 100 million active users. In: Wikipedia [Internet]. 2024 [cited 2024 Mar 14]. Available from: https://en.wikipedia.org/w/index.php?title=List_of_social_platforms_with_at_least_100_million_active_users&oldid=1213672777
14. Brynjolfsson E, Collis A. How Should We Measure the Digital Economy? Harvard Business Review [Internet]. 2019 Nov 1 [cited 2024 Mar 14]; Available from: https://hbr.org/2019/11/how-should-we-measure-the-digital-economy
15. Kazemi DM, Borsari B, Levine MJ, Dooley B. Systematic review of surveillance by social media platforms for illicit drug use. J Public Health. 2017 Dec 1;39(4):763–76.
16. Haupt MR, Cuomo R, Li J, Nali M, Mackey TK. The influence of social media affordances on drug dealer posting behavior across multiple social networking sites (SNS). Comput Hum Behav Rep. 2022 Dec 1;8:100235.
17. Moyle L, Childs A, Coomber R, Barratt MJ. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs. Int J Drug Policy. 2019 Jan 1;63:101–10.
18. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing Social

Media Data for Pharmacovigilance: A Review. J Biomed Inform. 2015 Apr;54:202–12.
19. Lee JY, Lee YS, Kim DH, Lee HS, Yang BR, Kim MG. The Use of Social Media in Detecting Drug Safety–Related New Black Box Warnings, Labeling Changes, or Withdrawals: Scoping Review. JMIR Public Health Surveill. 2021 Jun 28;7(6):e30137.
20. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc. 2015 Mar 1;22(3):671–81.
21. Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: A review of the opportunities and challenges. Br J Clin Pharmacol. 2015;80(4):910–20.
22. Paul MJ, Chisolm MS, Johnson MW, Vandrey RG, Dredze M. Assessing the Validity of Online Drug Forums as a Source for Estimating Demographic and Temporal Trends in Drug Use. J Addict Med. 2016 Oct;10(5):324.
23. Correia RB, Wood IB, Bollen J, Rocha LM. Mining Social Media Data for Biomedical Signals and Health-Related Behavior. Annu Rev Biomed Data Sci. 2020;3(1):433–58.
24. Yang CC, Yang H, Jiang L. Postmarketing Drug Safety Surveillance Using Publicly Available Health-Consumer-Contributed Content in Social Media. ACM Trans Manag Inf Syst. 2014 Apr 1;5(1):2:1-2:21.
25. Wightman RS, Perrone J, Erowid F, Erowid E, Meisel ZF, Nelson LS. Comparative Analysis of Opioid Queries on Erowid.org: An Opportunity to Advance Harm Reduction. Subst Use Misuse. 2017 Aug 24;52(10):1315–9.
26. Thanki D, Frederick BJ. Social media and drug markets. In Luxembourg: Publications Office of the European Union; 2016 [cited 2024 Mar 14]. p. 115–23. Available from: http://www.emcdda.europa.eu/system/files/publications/2155/TDXD16001ENN_FINAL.pdf
27. Tofighi B, Desai A, Grov C, Lee J. Craigslist as a Source for Illicit Drugs: A Case Report and Review of the Literature. Publ Res [Internet]. 2016 Mar 9; Available from: https://academicworks.cuny.edu/sph_pubs/60
28. Laestadius L, Wang Y. Youth access to JUUL online: eBay sales of JUUL prior to and following FDA action. Tob Control. 2019 Nov;28(6):617–22.
29. Blok D, Ambrose L, Ouellette L, Seif E, Riley B, Judge B, et al. Selling poison by the bottle: Availability of dangerous substances found on eBay®. Am J Emerg Med. 2020 Apr 1;38(4):846–8.
30. Williams RS. Underage internet alcohol sales on eBay. Addiction. 2013;108(7):1346–8.
31. Haber I, Pergolizzi J, LeQuang JA. Poppy Seed Tea: A Short Review and Case Study. Pain Ther. 2019 Jun;8(1):151–5.
32. Rajendra-Nicolucci C, Zuckerman E. Top 100: The most popular social media platforms and what they can teach us [Internet]. 2021. Available from: https://knightcolumbia.org/blog/top-100-the-most-popular-social-media-platforms-and-what-they-can-teach-us
33. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language Models are Few-Shot Learners. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2020 [cited 2024 Mar 14]. p. 1877–901. Available from: https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html
34. Carpenter KA, Altman RB. Using GPT-3 to Build a Lexicon of Drugs of Abuse Synonyms for Social Media Pharmacovigilance. Biomolecules. 2023 Feb 18;13(2):387.
35. Steen E, Yurechko K, Klug D. You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok. Soc Media Soc. 2023 Jul 1;9(3):20563051231194586.
36. Advanced filtering for geo data [Internet]. [cited 2024 Mar 14]. Available from: https://developer.twitter.com/en/docs/tutorials/advanced-filtering-for-geo-data

37. Flores L, Young SD. Regional Variation in Discussion of Opioids on Social Media: A Qualitative Study. J Addict Dis. 2021;39(3):316–21.
38. Calac AJ, McMann T, Cai M, Li J, Cuomo R, Mackey TK. Exploring substance use disorder discussions in Native American communities: a retrospective Twitter infodemiology study. Harm Reduct J. 2022 Dec 14;19:141.
39. Tibebu S, Chang VC, Drouin CA, Thompson W, Do MT. At-a-glance - What can social media tell us about the opioid crisis in Canada? Health Promot Chronic Dis Prev Can Res Policy Pract. 2018 Jun;38(6):263–7.
40. Tacheva Z, Ivanov A. Exploring the Association Between the “Big Five” Personality Traits and Fatal Opioid Overdose: County-Level Empirical Analysis. JMIR Ment Health. 2021 Mar 8;8(3):e24939.
41. Cuomo R, Purushothaman V, Calac AJ, McMann T, Li Z, Mackey T. Estimating County-Level Overdose Rates Using Opioid-Related Twitter Data: Interdisciplinary Infodemiology Study. JMIR Form Res. 2023 Jan 25;7:e42162.
42. Sarker A, Gonzalez-Hernandez G, Ruan Y, Perrone J. Machine Learning and Natural Language Processing for Geolocation-Centric Monitoring and Characterization of Opioid-Related Social Media Chatter. JAMA Netw Open. 2019 Nov 6;2(11):e1914672.
43. Fodeh SJ, Al-Garadi M, Elsankary O, Perrone J, Becker W, Sarker A. Utilizing a multi-class classification approach to detect therapeutic and recreational misuse of opioids on Twitter. Comput Biol Med. 2021 Feb;129:104132.
44. Yin J, Gao Y, Chi G. An Evaluation of Geo-located Twitter Data for Measuring Human Migration. Int J Geogr Inf Sci IJGIS. 2022;36(9):1830–52.
45. Kellert O, Matlis NH. Geolocation of multiple sociolinguistic markers in Buenos Aires. PLoS ONE. 2022 Sep 9;17(9):e0274114.
46. Cao J, Hochmair HH, Basheeh F. The Effect of Twitter App Policy Changes on the Sharing of Spatial Information through Twitter Users. Geographies. 2022 Sep;2(3):549–62.
47. Kruspe A, Häberle M, Hoffmann EJ, Rode-Hasinger S, Abdulahhad K, Zhu XX. Changes in Twitter geolocations: Insights and suggestions for future usage [Internet]. arXiv; 2021 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/2108.12251
48. Lin YC, Lai CM, Chapman JW, Wu SF, Barnett GA. Geo-Location Identification of Facebook Pages. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) [Internet]. 2018 [cited 2024 Mar 14]. p. 441–6. Available from: https://ieeexplore.ieee.org/abstract/document/8508816
49. CORREIA RB, LI L, ROCHA LM. MONITORING POTENTIAL DRUG INTERACTIONS AND REACTIONS VIA NETWORK ANALYSIS OF INSTAGRAM USER TIMELINES. Pac Symp Biocomput Pac Symp Biocomput. 2016;21:492–503.
50. Allem JP, Chu KH, Cruz TB, Unger JB. Waterpipe Promotion and Use on Instagram: #Hookah. Nicotine Tob Res. 2017 Oct;19(10):1248–52.
51. van Zanten BT, Van Berkel DB, Meentemeyer RK, Smith JW, Tieskens KF, Verburg PH. Continental-scale quantification of landscape values using social media data. Proc Natl Acad Sci U S A. 2016 Nov 15;113(46):12974–9.
52. Mukhina KD, Rakitin SV, Visheratin AA. Detection of tourists attraction points using Instagram profiles. Procedia Comput Sci. 2017 Jan 1;108:2378–82.
53. Song YC, Zhang YD, Cao J, Xia T, Liu W, Li JT. Web Video Geolocation by Geotagged Social Resources. IEEE Trans Multimed. 2012 Apr;14(2):456–70.
54. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.
55. Gavrić K, Ćulibrk D, Mirković M, Crnojević V. Using YouTube data to analyze human continent-level mobility. In: 2011 International Conference on Computational Aspects of Social Networks (CASoN) [Internet]. 2011 [cited 2024 Mar 14]. p. 207–10. Available from: https://ieeexplore.ieee.org/abstract/document/6085945

56. Baik J. The Geotagging Counterpublic: The Case of Facebook Remote Check-Ins to Standing Rock. Int J Commun. 2020 Mar 13;14(0):21.
57. Data For Good Tools and Data [Internet]. [cited 2024 Mar 14]. Available from: https://dataforgood.facebook.com/dfg/tools
58. Pérez-Arnal R, Conesa D, Alvarez-Napagao S, Suzumura T, Català M, Alvarez-Lacalle E, et al. Comparative Analysis of Geolocation Information through Mobile-Devices under Different COVID-19 Mobility Restriction Patterns in Spain. ISPRS Int J Geo-Inf. 2021 Feb;10(2):73.
59. Zaremba K. Opening of hotels and ski facilities: Impact on mobility, spending, and Covid-19 outcomes. Health Econ. 2023;32(5):1148–80.
60. State B, Rodriguez M, Helbing D, Zagheni E. Migration of Professionals to the U.S. In: Aiello LM, McFarland D, editors. Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014 Proceedings [Internet]. Cham: Springer International Publishing; 2014 [cited 2024 Mar 14]. p. 531–43. Available from: https://doi.org/10.1007/978-3-319-13734-6_37
61. Fung B. CNN. 2023 [cited 2024 Mar 14]. TikTok collects a lot of data. But that's not the main reason officials say it's a security risk | CNN Business. Available from: https://www.cnn.com/2023/03/24/tech/tiktok-ban-national-security-hearing/index.html
62. News · NL· C. CBC. 2023 [cited 2024 Mar 14]. If you use TikTok, the app is collecting a staggering amount of information about you | CBC News. Available from: https://www.cbc.ca/news/canada/tiktok-data-collection-privacy-1.6763626
63. Shepardson D. TikTok CEO: App has never shared US data with Chinese government. Reuters [Internet]. 2023 Mar 22 [cited 2024 Mar 14]; Available from: https://www.reuters.com/technology/tiktok-ceo-app-has-never-shared-us-data-with-chinese-goverment-2023-03-22/
64. Location information on TikTok | TikTok Help Center [Internet]. [cited 2024 Mar 14]. Available from: https://support.tiktok.com/en/account-and-privacy/account-privacy-settings/location-services-on-tiktok
65. Ebert N, Geppert T, Strycharz J, Knieps M, Hönig M, Brucker-Kley E. Creative beyond TikToks: Investigating Adolescents' Social Privacy Management on TikTok. Proc Priv Enhancing Technol. 2023 Apr;2023(2):221–35.
66. Franqueira VNL, Annor JA, Kafali O. Age Appropriate Design: Assessment of TikTok, Twitch, and YouTube Kids [Internet]. arXiv; 2022 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/2208.02638
67. Zannettou S, Nemeth ON, Ayalon O, Goetzen A, Gummadi KP, Redmiles EM, et al. Leveraging Rights of Data Subjects for Social Media Analysis: Studying TikTok via Data Donations [Internet]. arXiv; 2023 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/2301.04945
68. Lamy FR, Daniulaityte R, Nahhas RW, Barratt MJ, Smith AG, Sheth A, et al. Increased in synthetic cannabinoids-related harms: results from a longitudinal web-based content analysis. Int J Drug Policy. 2017 Jun;44:121–9.
69. Hansen Schwartz JE. AAAI. [cited 2024 Mar 14]. Characterizing Geographic Variation in Well-Being Using Tweets. Available from: https://aaai.org/papers/00583-14442-characterizing-geographic-variation-in-well-being-using-tweets/
70. Graves RL, Tufts C, Meisel ZF, Polsky D, Ungar L, Merchant RM. Opioid Discussion in the Twittersphere. Subst Use Misuse. 2018 Nov 10;53(13):2132–9.
71. Anwar M, Khoury D, Aldridge AP, Parker SJ, Conway KP. Using Twitter to Surveil the Opioid Epidemic in North Carolina: An Exploratory Study. JMIR Public Health Surveill. 2020 Jun 24;6(2):e17574.
72. Giorgi S, Yaden DB, Eichstaedt JC, Ashford RD, Buffone AEK, Schwartz HA, et al. Cultural Differences in Tweeting about Drinking Across the US. Int J Environ Res Public Health.

2020 Feb;17(4):1125.
73. Curtis B, Giorgi S, Buffone AEK, Ungar LH, Ashford RD, Hemmons J, et al. Can Twitter be used to predict county excessive alcohol consumption rates? PLoS ONE. 2018 Apr 4;13(4):e0194290.
74. Jha D, Singh R. SMARTS: the social media-based addiction recovery and intervention targeting server. Bioinforma Oxf Engl. 2019 Oct 24;btz800.
75. Milusheva S, Marty R, Bedoya G, Williams S, Resor E, Legovini A. Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning. PLoS ONE. 2021 Feb 3;16(2):e0244317.
76. Sergeeva J, Filatova A, Kovalchuk M, Teryoshkin S. SemAGR: semantic method for accurate geolocations reconstruction within extensive urban sites. Procedia Comput Sci. 2022 Jan 1;212:409–17.
77. Compton R, Keegan MS, Xu J. Inferring the geographic focus of online documents from social media sharing patterns [Internet]. arXiv; 2014 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/1406.2392
78. Pinto G, Burghardt K, Lerman K, Ferrara E. GET-Tok: A GenAI-Enriched Multimodal TikTok Dataset Documenting the 2022 Attempted Coup in Peru [Internet]. arXiv; 2024 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/2402.05882
79. Bandy J, Diakopoulos N. #TulsaFlop: A Case Study of Algorithmically-Influenced Collective Action on TikTok [Internet]. arXiv; 2020 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/2012.07716
80. Hu M, Conway M. Perspectives of the COVID-19 Pandemic on Reddit: Comparative Natural Language Processing Study of the United States, the United Kingdom, Canada, and Australia. JMIR Infodemiology. 2022 Sep 27;2(2):e36941.
81. (PDF) Exploring the Geography of Tags in Youtube Views [Internet]. [cited 2024 Mar 14]. Available from: https://www.researchgate.net/publication/281658097_Exploring_the_Geography_of_Tags_in_Youtube_Views
82. Goyer C, Castillon G, Moride Y. Implementation of Interventions and Policies on Opioids and Awareness of Opioid-Related Harms in Canada: A Multistage Mixed Methods Descriptive Study. Int J Environ Res Public Health. 2022 Apr 22;19(9):5122.
83. Scalia G, Francalanci C, Pernici B. CIME: Context-aware geolocation of emergency-related posts. GeoInformatica. 2022 Jan 1;26(1):125–57.
84. Jurgens D, Finethy T, McCorriston J, Xu Y, Ruths D. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. Proc Int AAAI Conf Web Soc Media. 2015;9(1):188–97.
85. Pellet H, Shiaeles S, Stavrou S. Localising social network users and profiling their movement. Comput Secur. 2019 Mar 1;81:49–57.
86. Bozarth L, Quercia D, Capra L, Šćepanović S. The role of the big geographic sort in online news circulation among U.S. Reddit users. Sci Rep. 2023 Apr 25;13(1):6711.
87. Balsamo D, Bajardi P, Panisson A. Firsthand Opiates Abuse on Social Media: Monitoring Geospatial Patterns of Interest Through a Digital Cohort. In: The World Wide Web Conference [Internet]. 2019 [cited 2024 Mar 14]. p. 2572–9. Available from: http://arxiv.org/abs/1904.00003
88. Alhuzali H, Zhang T, Ananiadou S. Emotions and Topics Expressed on Twitter During the COVID-19 Pandemic in the United Kingdom: Comparative Geolocation and Text Mining Analysis. J Med Internet Res. 2022 Oct 5;24(10):e40323.
89. Dredze M, Paul MJ, Bergsma S, Tran H. Carmen: A twitter geolocation system with applications to public health. In 2013. p. 20–4.
90. geopy: Python Geocoding Toolbox [Internet]. [cited 2024 Mar 14]. Available from: https://github.com/geopy/geopy

91. Harris JA, Beck NA, Niedziela CJ, Alvarez GA, Danquah SA, Afshar S. The global reach of social media in oral and maxillofacial surgery. Oral Maxillofac Surg. 2022 Jun 24;1–5.
92. Sumner SA, Galik S, Mathieu J, Ward M, Kiley T, Bartholow B, et al. Temporal and Geographic Patterns of Social Media Posts About an Emerging Suicide Game. J Adolesc Health Off Publ Soc Adolesc Med. 2019 Jul;65(1):94–100.
93. Glowacki EM, Glowacki JB, Wilcox GB. A Text-Mining Analysis of the Public's Reactions to the Opioid Crisis. Subst Abuse. 2018 Apr 1;39(2):129–33.
94. Iconosquare - Analytics and management for Instagram, TikTok, LinkedIn, Twitter & Facebook [Internet]. [cited 2024 Mar 14]. Available from: https://www.iconosquare.com/
95. Brandwatch [Internet]. [cited 2024 Mar 14]. Brandwatch | The social suite of the future. Available from: https://www.brandwatch.com/
96. Reputation [Internet]. [cited 2024 Mar 14]. Reputation | Online Reputation Management for Business. Available from: https://reputation.com/
97. Mahajan R, Mansotra V. Predicting Geolocation of Tweets: Using Combination of CNN and BiLSTM. Data Sci Eng. 2021;6(4):402–10.
98. Zola P, Cortez P, Carpita M. Twitter user geolocation using web country noun searches. Decis Support Syst. 2019 May 1;120:50–9.
99. Bakerman J, Pazdernik K, Wilson A, Fairchild G, Bahran R. Twitter Geolocation: A Hybrid Approach. ACM Trans Knowl Discov Data. 2018 Mar 23;12(3):34:1-34:17.
100. Cha M, Gwon Y, Kung H. Twitter Geolocation and Regional Classification via Sparse Coding. Proc Int AAAI Conf Web Soc Media. 2015;9(1):582–5.
101. Dredze M, Osborne M, Kambadur P. Geolocation for Twitter: Timing Matters. In 2016. p. 1064–9.
102. Xu J, Compton R, Lu TC, Allen D. Rolling through tumblr: characterizing behavioral patterns of the microblogging platform. In: Proceedings of the 2014 ACM conference on Web science [Internet]. New York, NY, USA: Association for Computing Machinery; 2014 [cited 2024 Mar 14]. p. 13–22. (WebSci '14). Available from: https://dl.acm.org/doi/10.1145/2615569.2615694
103. Nobles AL, Johnson DC, Leas EC, Goodman-Meza D, Zúñiga ML, Ziedonis D, et al. Characterizing Self-Reports of Self-Identified Patient Experiences with Methadone Maintenance Treatment on an Online Community during COVID-19. Subst Use Misuse. 2021;56(14):2134–40.
104. Garcia GGP, Dehghanpoor R, Stringfellow EJ, Gupta M, Rochelle J, Mason E, et al. Identifying and Characterizing Medical Advice-Seekers on a Social Media Forum for Buprenorphine Use. Int J Environ Res Public Health. 2022 May 22;19(10):6281.
105. ElSherief M, Sumner SA, Jones CM, Law RK, Kacha-Ochana A, Shieber L, et al. Characterizing and Identifying the Prevalence of Web-Based Misinformation Relating to Medication for Opioid Use Disorder: Machine Learning Approach. J Med Internet Res. 2021 Dec 22;23(12):e30753.
106. Spadaro A, Sarker A, Hogg-Bremer W, Love JS, O'Donnell N, Nelson LS, et al. Reddit discussions about buprenorphine associated precipitated withdrawal in the era of fentanyl. Clin Toxicol Phila Pa. 2022 Jun;60(6):694–701.
107. Black JC, Margolin ZR, Olson RA, Dart RC. Online Conversation Monitoring to Understand the Opioid Epidemic: Epidemiological Surveillance Study. JMIR Public Health Surveill. 2020 Jun 29;6(2):e17073.
108. Chan B, Lopez A, Sarkar U. The Canary in the Coal Mine Tweets: Social Media Reveals Public Perceptions of Non-Medical Use of Opioids. PloS One. 2015;10(8):e0135072.
109. Al-Garadi MA, Yang YC, Guo Y, Kim S, Love JS, Perrone J, et al. Large-Scale Social Media Analysis Reveals Emotions Associated with Nonmedical Prescription Drug Use. Health Data Sci. 2022;2022:9851989.
110. Shutler L, Nelson LS, Portelli I, Blachford C, Perrone J. Drug Use in the Twittersphere: A

Qualitative Contextual Analysis of Tweets About Prescription Drugs. J Addict Dis. 2015;34(4):303–10.
111. Kalyanam J, Katsuki T, R G Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav. 2017 Feb;65:289–95.
112. Raza S, Schwartz B, Lakamana S, Ge Y, Sarker A. A framework for multi-faceted content analysis of social media chatter regarding non-medical use of prescription medications. BMC Digit Health. 2023;1:29.
113. Yoon S, Odlum M, Broadwell P, Davis N, Cho H, Deng N, et al. Application of Social Network Analysis of COVID-19 Related Tweets Mentioning Cannabis and Opioids to Gain Insights for Drug Abuse Research. Stud Health Technol Inform. 2020 Jun 26;272:5–8.
114. Haug NA, Bielenberg J, Linder SH, Lembke A. Assessment of provider attitudes toward #naloxone on Twitter. Subst Abuse. 2016;37(1):35–41.
115. 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 Pa. 2021 Nov;59(11):982–91.
116. D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Lee KE, Cavazos-Rehg PA. Social networking online to recover from opioid use disorder: A study of community interactions. Drug Alcohol Depend. 2017 Dec 1;181:5–10.
117. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz J, et al. Social Media Based Analysis of Opioid Epidemic Using Reddit. AMIA Annu Symp Proc. 2018 Dec 5;2018:867–76.
118. Smith KE, Rogers JM, Strickland JC, Epstein DH. When an obscurity becomes trend: social-media descriptions of tianeptine use and associated atypical drug use. Am J Drug Alcohol Abuse. 2021 Jul 4;47(4):455–66.
119. Kasson E, Filiatreau LM, Kaiser N, Davet K, Taylor J, Garg S, et al. Using Social Media to Examine Themes Surrounding Fentanyl Misuse and Risk Indicators. Subst Use Misuse. 2023;58(7):920–9.
120. 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 Nov;38(15):3283–7.
121. Catalani V, Arillotta D, Corkery JM, Guirguis A, Vento A, Schifano F. Identifying New/Emerging Psychoactive Substances at the Time of COVID-19; A Web-Based Approach. Front Psychiatry. 2020;11:632405.
122. Rajesh K, Wilcox G, Ring D, Mackert M. Reactions to the opioid epidemic: A text-mining analysis of tweets. J Addict Dis. 2021;39(2):183–8.
123. Ramachandran S, Brown L, Ring D. Tones and themes in Reddits posts discussing the opioid epidemic. J Addict Dis. 2022;40(4):552–8.
124. 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 Dec;26:100467.
125. Chancellor S, Nitzburg G, Hu A, Zampieri F, De Choudhury M. Discovering Alternative Treatments for Opioid Use Recovery Using Social Media. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2024 Mar 14]. p. 1–15. (CHI '19). Available from: https://dl.acm.org/doi/10.1145/3290605.3300354
126. 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 Mar 30;6(3):e33919.
127. Yao H, Rashidian S, Dong X, Duanmu H, Rosenthal RN, Wang F. Detection of Suicidality Among Opioid Users on Reddit: Machine Learning-Based Approach. J Med Internet Res.

2020 Nov 27;22(11):e15293.
128. Jha D, Singh R. Analysis of associations between emotions and activities of drug users and their addiction recovery tendencies from social media posts using structural equation modeling. BMC Bioinformatics. 2020 Dec 30;21(Suppl 18):554.
129. Naserianhanzaei E, Koschate-Reis M. Effects of Substance Use, Recovery, and Non-Drug-Related Online Community Participation on the Risk of a Use Episode During Remission From Opioid Use Disorder: Longitudinal Observational Study. J Med Internet Res. 2022 Aug 22;24(8):e36555.
130. Kepner W, Meacham MC, Nobles AL. Types and Sources of Stigma on Opioid Use Treatment and Recovery Communities on Reddit. Subst Use Misuse. 2022;57(10):1511–22.
131. Chen AT, Johnny S, Conway M. Examining stigma relating to substance use and contextual factors in social media discussions. Drug Alcohol Depend Rep. 2022 Jun;3:100061.
132. Bremer W, Plaisance K, Walker D, Bonn M, Love JS, Perrone J, et al. Barriers to opioid use disorder treatment: A comparison of self-reported information from social media with barriers found in literature. Front Public Health. 2023;11:1141093.
133. Alambo A, Padhee S, Banerjee T, Thirunarayan K. COVID-19 and Mental Health/Substance Use Disorders on Reddit: A Longitudinal Study. ArXiv. 2020 Nov 20;arXiv:2011.10518v1.
134. Frank D, Krawczyk N, Arshonsky J, Bragg MA, Friedman SR, Bunting AM. COVID-19-Related Changes to Drug-Selling Networks and Their Effects on People Who Use Illicit Opioids. J Stud Alcohol Drugs. 2023 Mar;84(2):222–9.
135. Arshonsky J, Krawczyk N, Bunting AM, Frank D, Friedman SR, Bragg MA. Informal Coping Strategies Among People Who Use Opioids During COVID-19: Thematic Analysis of Reddit Forums. JMIR Form Res. 2022 Mar 3;6(3):e32871.
136. Krawczyk N, Bunting AM, Frank D, Arshonsky J, Gu Y, Friedman SR, et al. “How will I get my next week’s script?” Reactions of Reddit opioid forum users to changes in treatment access in the early months of the coronavirus pandemic. Int J Drug Policy. 2021 Jun;92:103140.
137. 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 Mar 5;17(1):16.
138. Arillotta D, Guirguis A, Corkery JM, Scherbaum N, Schifano F. COVID-19 Pandemic Impact on Substance Misuse: A Social Media Listening, Mixed Method Analysis. Brain Sci. 2021 Jul 9;11(7):907.
139. El-Bassel N, Hochstatter KR, Slavin MN, Yang C, Zhang Y, Muresan S. Harnessing the Power of Social Media to Understand the Impact of COVID-19 on People Who Use Drugs During Lockdown and Social Distancing. J Addict Med. 2022 Apr 1;16(2):e123–32.
140. Sarker A, Al-Garadi MA, Ge Y, Nataraj N, Jones CM, Sumner SA. Signals of increasing co-use of stimulants and opioids from online drug forum data. Harm Reduct J. 2022 May 25;19(1):51.
141. Meacham MC, Nobles AL, Tompkins DA, Thrul J. “I got a bunch of weed to help me through the withdrawals”: Naturalistic cannabis use reported in online opioid and opioid recovery community discussion forums. PloS One. 2022;17(2):e0263583.
142. Purushothaman V, Li J, Mackey TK. Detecting Suicide and Self-Harm Discussions Among Opioid Substance Users on Instagram Using Machine Learning. Front Psychiatry. 2021;12:551296.
143. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of Codeine Misuse on Instagram: Content Analysis. JMIR Public Health Surveill. 2018 Mar 20;4(1):e22.
144. Smolev ET, Rolf L, Zhu E, Buday SK, Brody M, Brogan DM, et al. “Pill Pushers and CBD

Oil”-A Thematic Analysis of Social Media Interactions About Pain After Traumatic Brachial Plexus Injury. J Hand Surg Glob Online. 2021 Jan;3(1):36–40.
145. Yang X, Luo J. Tracking Illicit Drug Dealing and Abuse on Instagram Using Multimodal Analysis. ACM Trans Intell Syst Technol. 2017 Feb 24;8(4):58:1-58:15.
146. Shah N, Li J, Mackey TK. An unsupervised machine learning approach for the detection and characterization of illicit drug-dealing comments and interactions on Instagram. Subst Abuse. 2022;43(1):273–7.
147. Zaprutko T, Kopciuch D, Paczkowska A, Sprawka J, Cynar J, Pogodzińska M, et al. Facebook as a source of access to medicines. PloS One. 2022;17(10):e0275272.
148. 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 Nov;36(11):3373–82.
149. Kilgo DK, Midberry J. Social Media News Production, Emotional Facebook Reactions, and the Politicization of Drug Addiction. Health Commun. 2022 Mar;37(3):375–83.
150. Ittefaq M, Zain A, Bokhari H. Opioids in Satirical News Shows: Exploring Topics, Sentiments, and Engagement in Last Week Tonight on YouTube. J Health Commun. 2023 Jan 2;28(1):53–63.
151. Al-Rawi A. The convergence of social media and other communication technologies in the promotion of illicit and controlled drugs. J Public Health. 2022 Mar 1;44(1):e153–60.
152. Zhao F, Skums P, Zelikovsky A, Sevigny EL, Swahn MH, Strasser SM, et al. Computational Approaches to Detect Illicit Drug Ads and Find Vendor Communities Within Social Media Platforms. IEEE/ACM Trans Comput Biol Bioinform. 2022;19(1):180–91.
153. Abouchedid R, Gilks T, Dargan PI, Archer JRH, Wood DM. Assessment of the Availability, Cost, and Motivations for Use over Time of the New Psychoactive Substances-Benzodiazepines Diclazepam, Flubromazepam, and Pyrazolam-in the UK. J Med Toxicol Off J Am Coll Med Toxicol. 2018 Jun;14(2):134–43.
154. Rhumorbarbe D, Morelato M, Staehli L, Roux C, Jaquet-Chiffelle DO, Rossy Q, et al. Monitoring new psychoactive substances: Exploring the contribution of an online discussion forum. Int J Drug Policy. 2019 Nov 1;73:273–80.
155. Merten JW, Gordon BT, King JL, Pappas C. Cannabidiol (CBD): Perspectives from Pinterest. Subst Use Misuse. 2020;55(13):2213–20.
156. Laestadius LI, Guidry JPD, Greskoviak R, Adams J. Making “Weedish Fish”: An Exploratory Analysis of Cannabis Recipes on Pinterest. Subst Use Misuse. 2019;54(13):2191–7.
157. Guidry J, Jin Y, Haddad L, Zhang Y, Smith J. How Health Risks Are Pinpointed (or Not) on Social Media: The Portrayal of Waterpipe Smoking on Pinterest. Health Commun. 2016;31(6):659–67.
158. Lee AS, Hart JL, Sears CG, Walker KL, Siu A, Smith C. A picture is worth a thousand words: Electronic cigarette content on Instagram and Pinterest. Tob Prev Cessat. 2017 Jul 3;3:119.
159. Sun T, Lim CCW, Chung J, Cheng B, Davidson L, Tisdale C, et al. Vaping on TikTok: a systematic thematic analysis. Tob Control. 2023 Mar;32(2):251–4.
160. Rutherford BN, Sun T, Johnson B, Co S, Lim TL, Lim CCW, et al. Getting high for likes: Exploring cannabis-related content on TikTok. Drug Alcohol Rev. 2022 Jul;41(5):1119–25.
161. Jancey J, Leaver T, Wolf K, Freeman B, Chai K, Bialous S, et al. Promotion of E-Cigarettes on TikTok and Regulatory Considerations. Int J Environ Res Public Health. 2023 May 9;20(10):5761.
162. Whelan J, Noller GE, Ward RD. Rolling through TikTok: An analysis of 3,4-methylenedioxymethamphetamine-related content. Drug Alcohol Rev. 2024 Jan;43(1):36–44.
163. Vosburg SK, Dailey-Govoni T, Beaumont J, Butler SF, Green JL. Characterizing the

Experience of Tapentadol Nonmedical Use: Mixed Methods Study. JMIR Form Res. 2022 Jun 10;6(6):e16996.
164. Soussan C, Kjellgren A. The users of Novel Psychoactive Substances: Online survey about their characteristics, attitudes and motivations. Int J Drug Policy. 2016 Jun 1;32:77–84.
165. Vera V. Nonsuicidal Self-Injury and Content Moderation on TikTok. Proc Assoc Inf Sci Technol. 2023;60(1):1164–6.
166. Moskal M, Supernak N. Do you speak algospeak? An introduction to the recent yet prominent phenomenon of Internet discourse from a cognitive linguistics perspective (abstract). In 2023.
167. Statista [Internet]. [cited 2024 Mar 14]. U.S. TikTok users by age 2022. Available from: https://www.statista.com/statistics/1095186/tiktok-us-users-age/
168. Stein MD, Conti MT, Kenney S, Anderson BJ, Flori JN, Risi MM, et al. Adverse childhood experience effects on opioid use initiation, injection drug use, and overdose among persons with opioid use disorder. Drug Alcohol Depend. 2017 Oct 1;179:325–9.
169. NW 1615 L. St, Washington S 800, Inquiries D 20036 U 419 4300 | M 857 8562 | F 419 4372 | M. Social Media Fact Sheet [Internet]. Pew Research Center: Internet, Science & Tech. [cited 2024 Mar 14]. Available from: https://www.pewresearch.org/internet/fact-sheet/social-media/
170. van Hoof JJ, Bekkers J, van Vuuren M. Son, you're smoking on Facebook! College students' disclosures on social networking sites as indicators of real-life risk behaviors. Comput Hum Behav. 2014 May 1;34:249–57.
171. Peters J. The Verge. 2023 [cited 2024 Mar 14]. Tweets aren't showing up in Google results as often because of changes at Twitter. Available from: https://www.theverge.com/2023/7/3/23783153/google-twitter-tweets-changes-rate-limits
172. Bureau UC. Census.gov. [cited 2024 Mar 14]. Nearly 68 Million People Spoke a Language Other Than English at Home in 2019. Available from: https://www.census.gov/library/stories/2022/12/languages-we-speak-in-united-states.html
173. Chen X, Faviez C, Schuck S, Lillo-Le-Louët A, Texier N, Dahamna B, et al. Mining Patients' Narratives in Social Media for Pharmacovigilance: Adverse Effects and Misuse of Methylphenidate. Front Pharmacol. 2018 May 24;9:541.
174. Fung ICH, Blankenship EB, Ahweyevu JO, Cooper LK, Duke CH, Carswell SL, et al. Public Health Implications of Image-Based Social Media: A Systematic Review of Instagram, Pinterest, Tumblr, and Flickr. Perm J. 2020;24:18.307.
175. Nguyen T, Larsen ME, O'Dea B, Phung D, Venkatesh S, Christensen H. Estimation of the prevalence of adverse drug reactions from social media. Int J Med Inf. 2017 Jun;102:130–7.
176. Drug Enforcement Administration. Social Media Drug Trafficking Threat [Internet]. 2022. Available from: https://www.dea.gov/sites/default/files/2022-03/20220208-DEA_Social%20Media%20Drug%20Trafficking%20Threat%20Overview.pdf
177. Magge A, Tutubalina E, Miftahutdinov Z, Alimova I, Dirkson A, Verberne S, et al. DeepADEMiner: a deep learning pharmacovigilance pipeline for extraction and normalization of adverse drug event mentions on Twitter. J Am Med Inform Assoc. 2021 Sep 18;28(10):2184–92.
178. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. Psychol Sci. 2015 Feb;26(2):159–69.
179. The Lancet Regional Health – Americas. Opioid crisis: addiction, overprescription, and insufficient primary prevention. Lancet Reg Health - Am. 2023 Jul;23:100557.
180. Kaafarani HMA, Han K, El Moheb M, Kongkaewpaisan N, Jia Z, El Hechi MW, et al. Opioids After Surgery in the United States Versus the Rest of the World: The International

Patterns of Opioid Prescribing (iPOP) Multicenter Study. Ann Surg. 2020 Dec;272(6):879–86.

FIGURES & TABLES

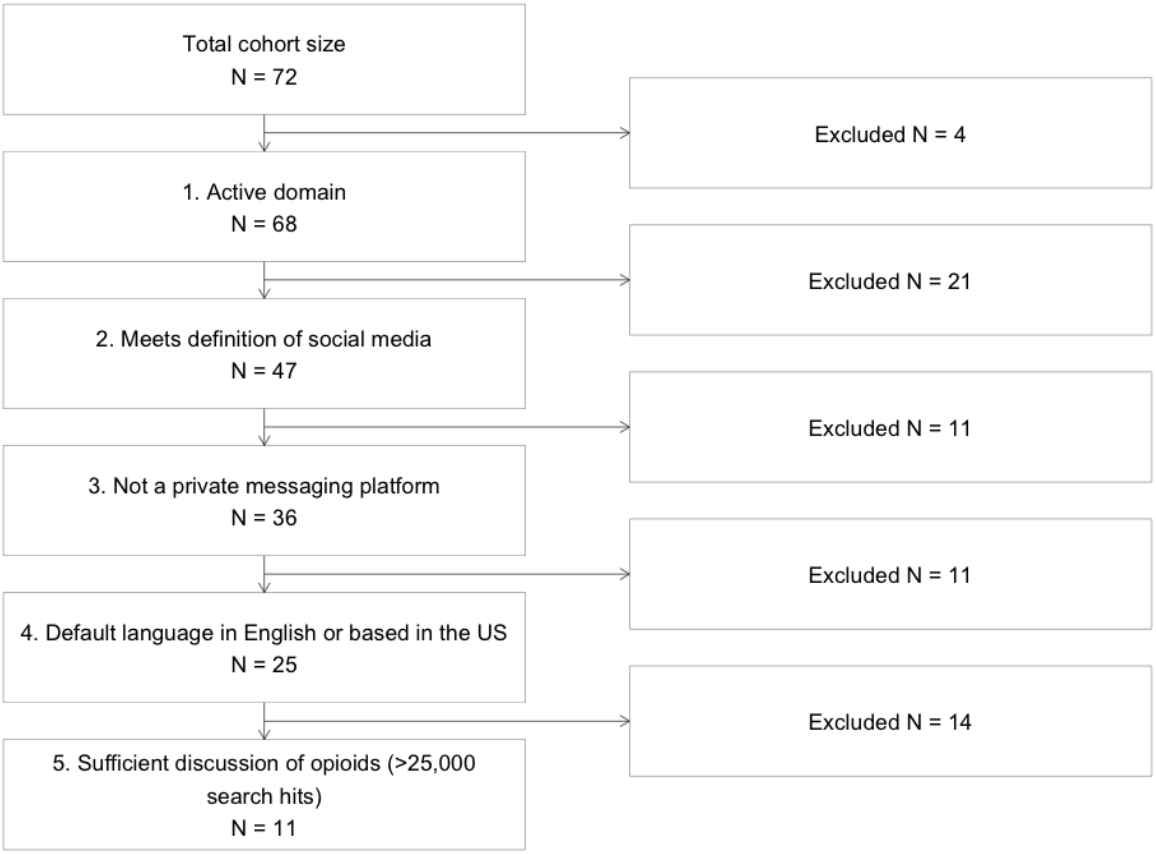


Figure 1. Consort diagram of iterative exclusion criteria to attain shortlist of 11 social media platforms for further characterization.

Table 1. Summary of characteristics of the 11 shortlisted social media platforms.
"Platform focus" gives a brief description of the primary usage of the platform. "Text data available" lists the types of text content that the platform contains. "Drug discussion not restricted" column has a checkmark if the platform does not restrict drug-related discussion, and an X if the platform has some form of restrictions. "Has API" column has a checkmark if the platform has an API available (whether the API is freely available or requires authorization before access), and an X if not. "Has research portal" column has a checkmark if the platform has a non-API platform for researchers to acquire platform data or for researchers to receive more information about collaboration with the platform, and an X if not. Number of checkmarks in "previously researched for opioid pharmacovigilance" column roughly corresponds to the relative amount of prior research, on a 4-point scale (with more checkmarks indicating relatively more prior research; see "Prior Use in Research Literature" section of Results). "Geolocation available" column has a checkmark if explicit geolocation data is provided for *any* platform content (this does not indicate explicit geolocation available for all content), and an X if not. "Example of geolocation inference strategy" column provides one possible method for inferring geolocation of platform content.

Platform Name	Platform focus	Text data available	Platform URL	Drug discussion not restricted?	Has API?	Has research portal?	Previously researched for opioid pharmacovigilance?	Geolocation available?	Example of geolocation inference strategy
Bluelight	Discussion of drug use and recovery	Forum posts	bluelight.org	✓	✗	✓	✗	✗	Self-described location of user
drugs-forum	Discussion of drug use and recovery	Forum posts	drugs-forum.com	✓	✗	✗	✓✓	✗	Self-described location of user
Facebook	Personal profiles and friend activity	Posts, comments, captions	facebook.com	✗	✓	✗	✓✓	✓	Consensus of friend locations
Instagram	Sharing photos and videos	Captions, comments	instagram.com	✗	✓	✗	✓✓	✓	Consensus of friend locations
LinkedIn	Professional networking	Posts, comments	linkedin.com	✗	✓	✗	✗	✓	Location-specific company

Pinterest	Visual curation	Captions, comments	pinterest.com	✗	✗	✓	✓	✗	Cross-posting to geolocatable platform
Reddit	Community networks and discussion	Posts, comments, captions	reddit.com	✓	✓	✗	✓✓✓✓	✗	Location-specific subreddits
TikTok	Sharing short videos	Captions, comments, transcripts	tiktok.com	✗	✓	✗	✗	✓	Location-specific hashtags
Tumblr	Microblogging	Posts, comments, captions	tumblr.com	✗	✓	✗	✓	✗	Cross-posting to geolocatable platform
X (Twitter)	Broadcasting short posts	Posts, comments	twitter.com	✗	✓	✗	✓✓✓✓	✓	Self-described location of user
YouTube	Sharing videos	Captions, comments, transcripts	youtube.com	✗	✓	✗	✓	✗	Location-specific hashtags

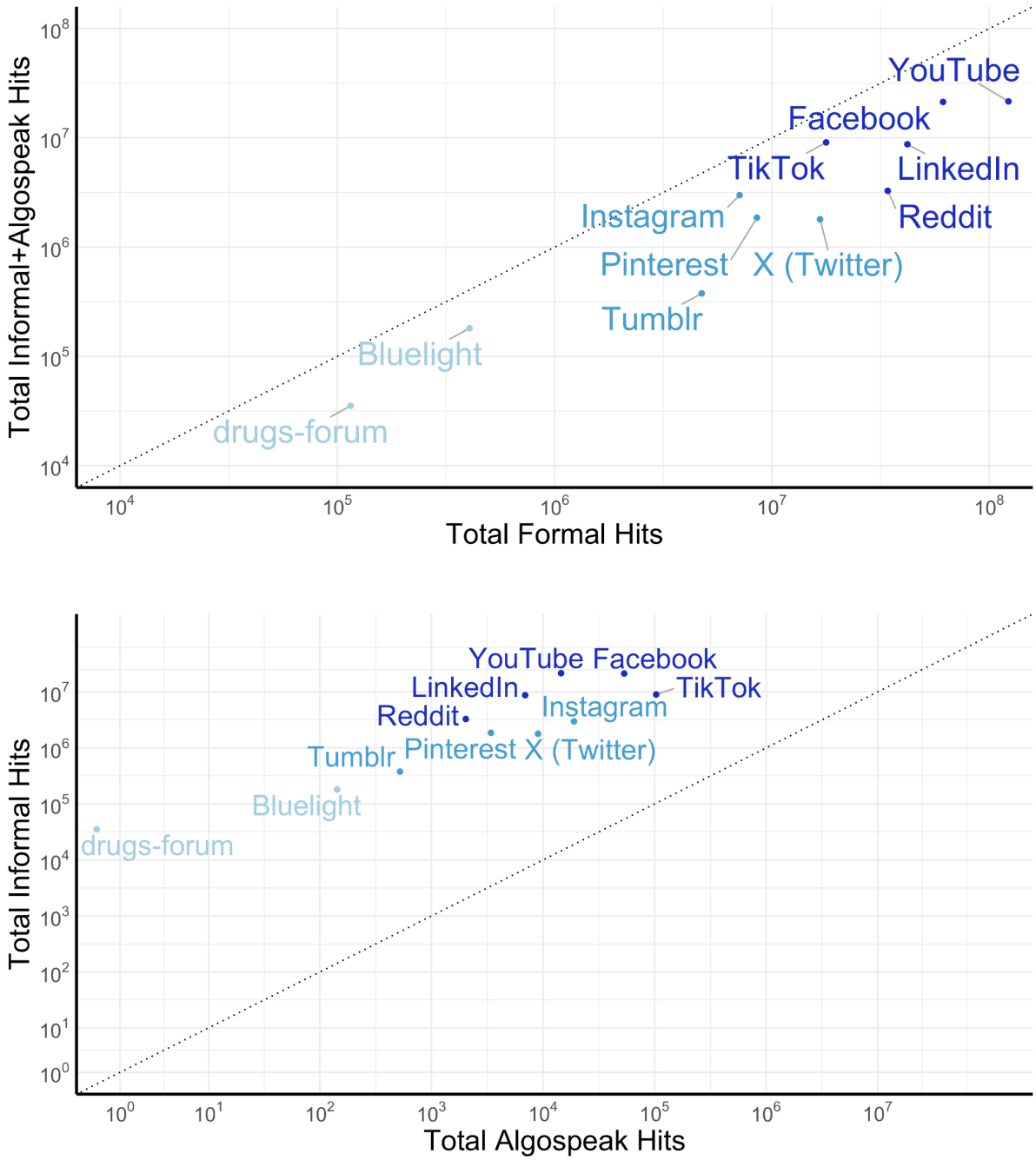


Figure 2. Total hits for each of the shortlisted platforms, shown with a logarithmic scale. Platforms are clustered into three categories by volume of hits, shown by color (darker shade of blue indicates higher volume of hits). Dotted line shows y=x diagonal.
a) Total hits for formal opioid term list versus the sum of the total hits for the informal opioid term list and the algospeak opioid term list.
b) Total hits for informal opioid term list versus algospeak opioid term list.

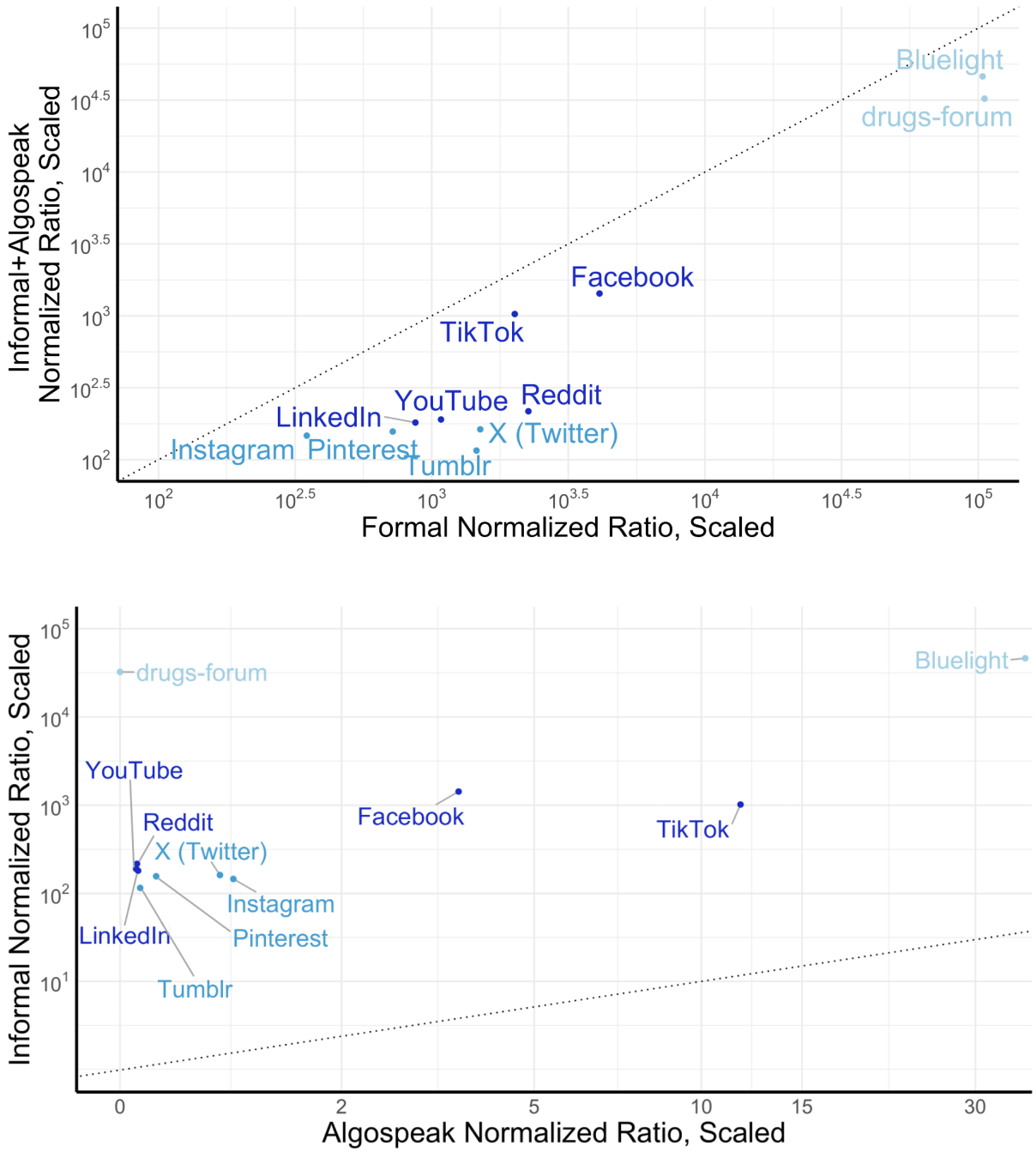
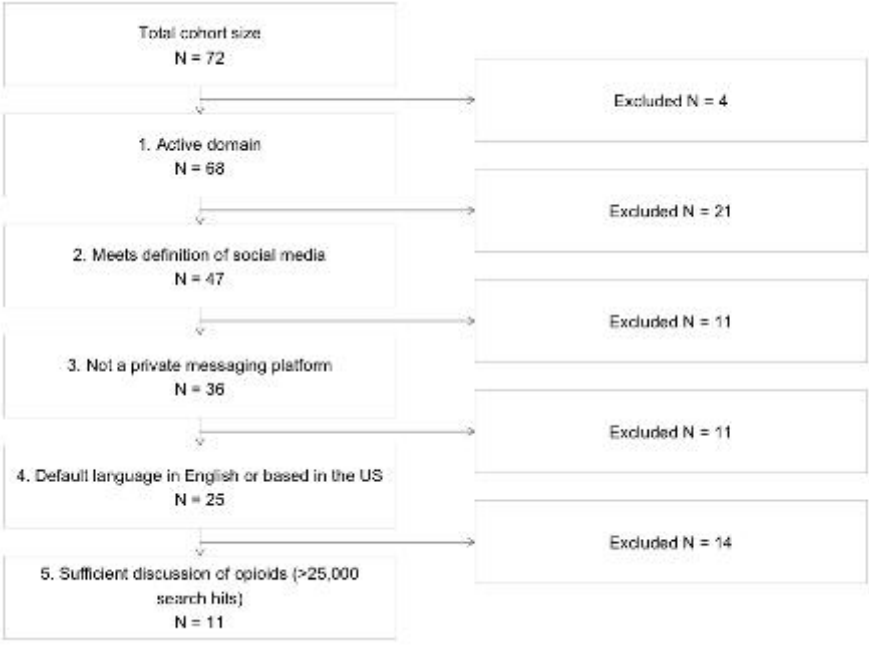
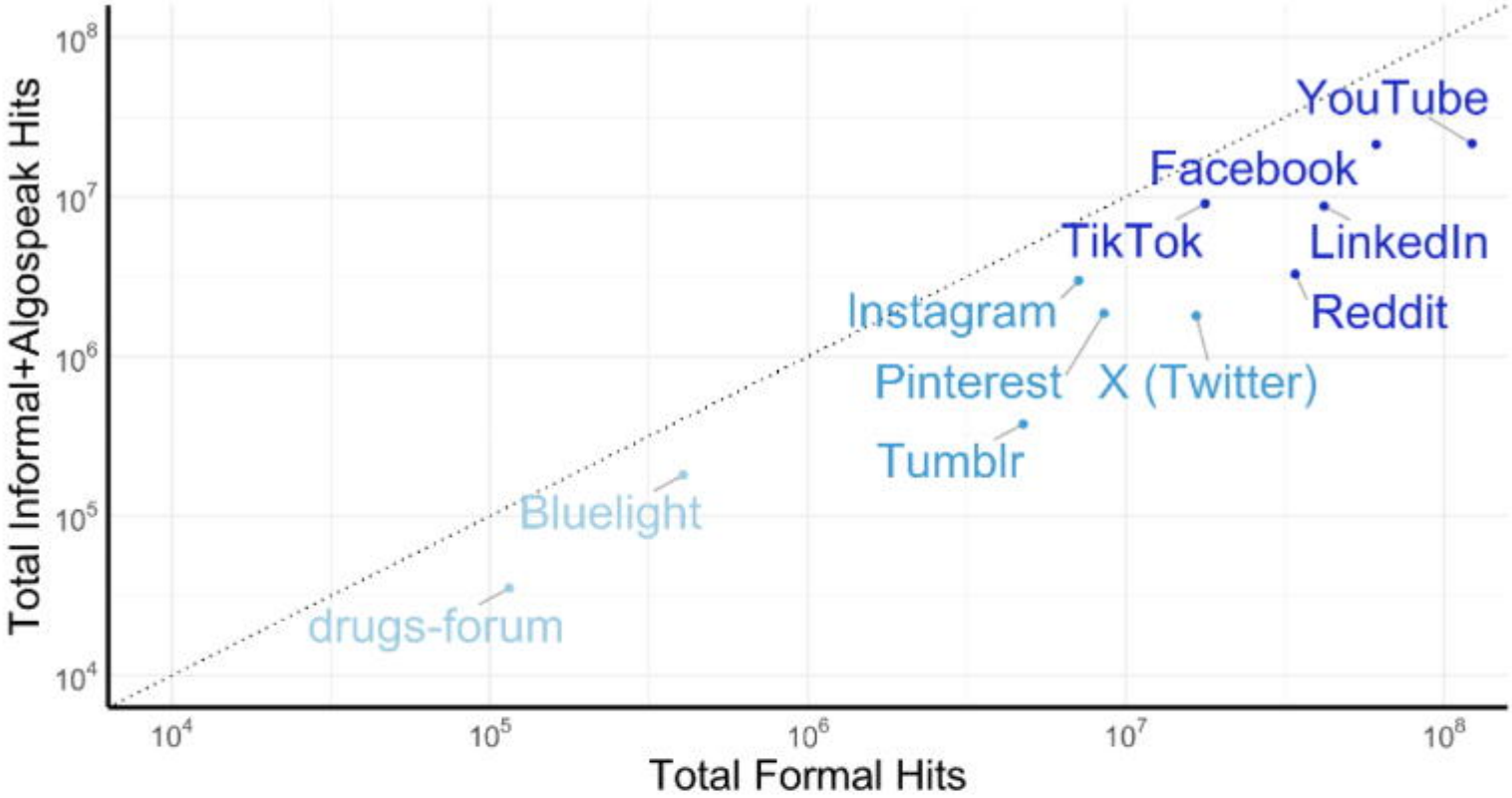


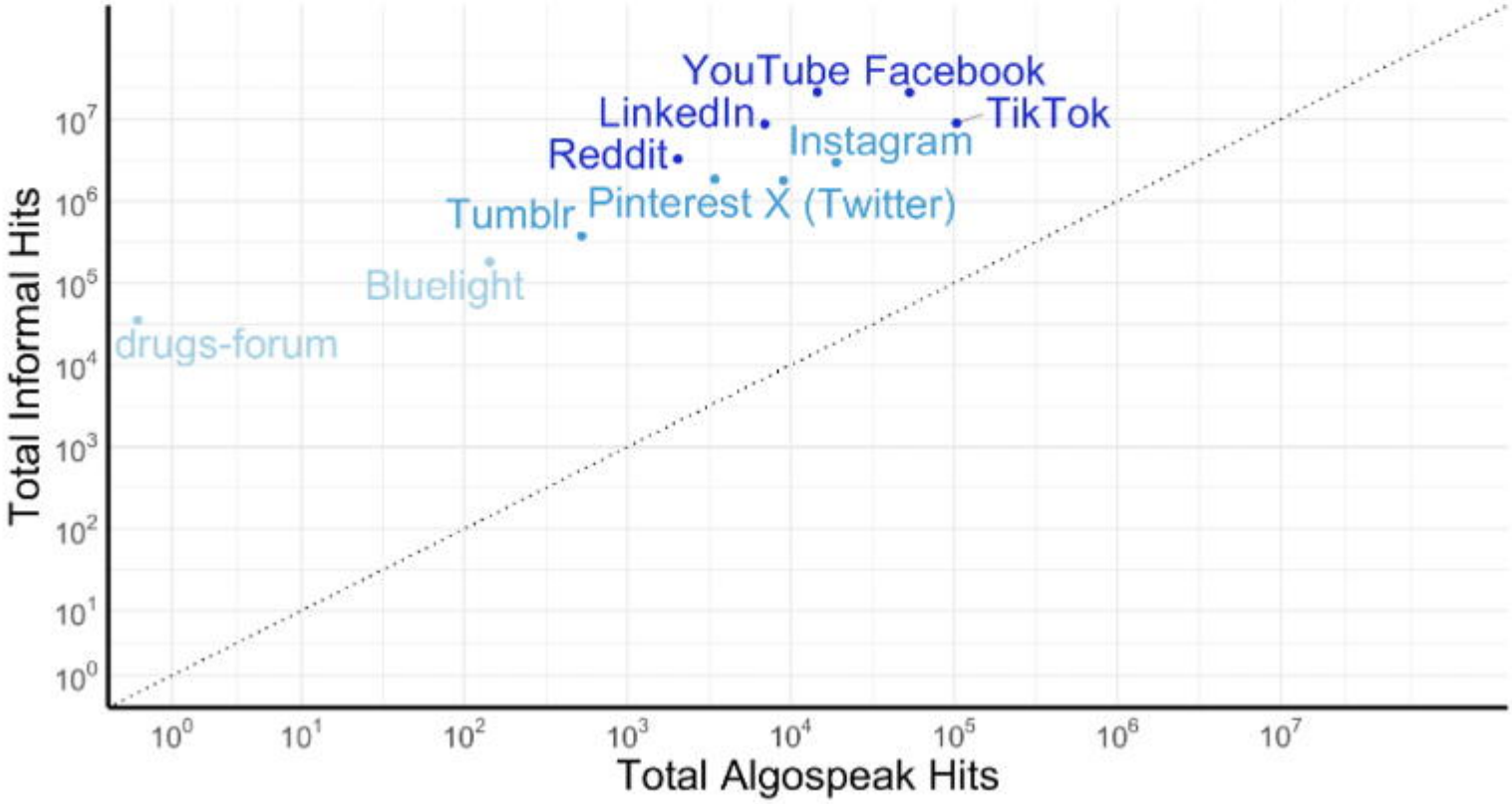
Figure 3. Term hits normalized by number of hits for common nouns, shown with a logarithmic scale and scaling factor of 100,000. As in Figure 2, platforms are clustered into three categories by total volume of hits, shown by color (darker shade of blue indicates higher volume of hits). Dotted line shows y=x diagonal.
a) Normalized hit ratios for formal opioid term list versus for informal opioid term list and algospeak opioid term list.
b) Normalized hit ratios for informal opioid term list versus algospeak opioid term list.

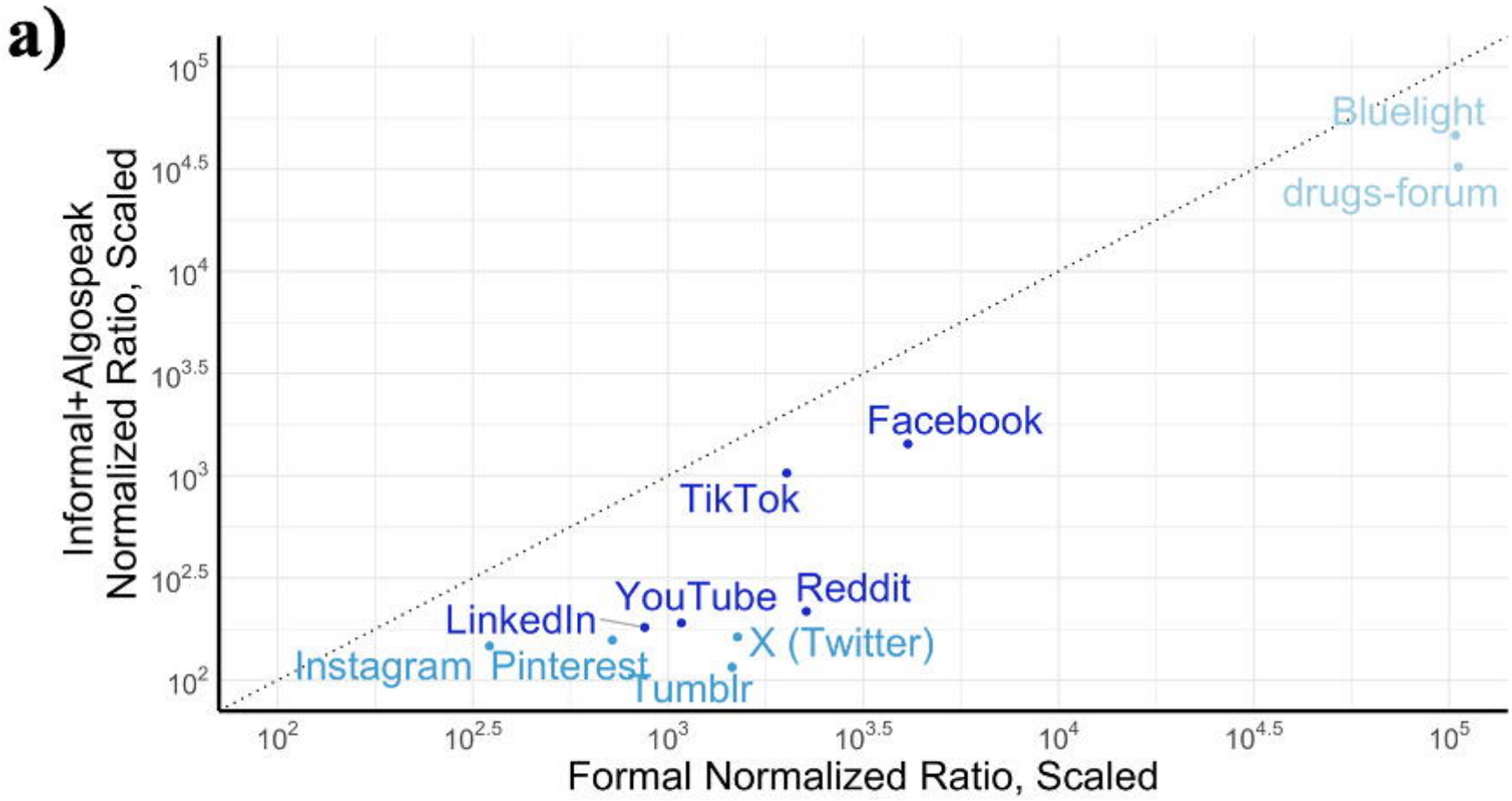

Total cohort size
N = 72
Excluded N = 4
1. Active domain
N = 68
Excluded N = 21
2. Meets definition of social media
N = 47
Excluded N = 11
3. Not a private messaging platform
N = 36
Excluded N = 11
4. Default language in English or based in the US
N = 25
Excluded N = 14
5. Sufficient discussion of opioids (>25,000 search hits)
N = 11

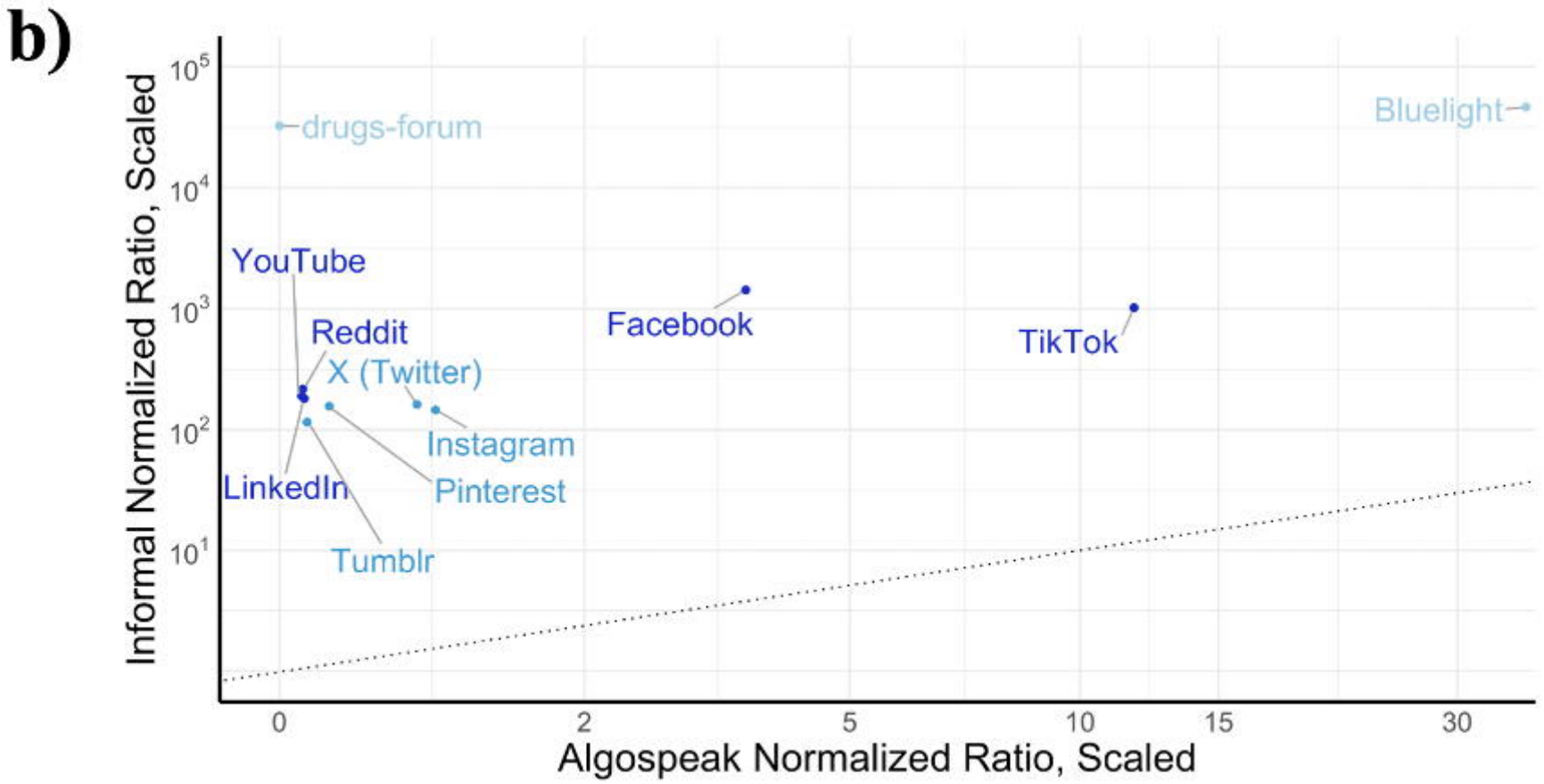
a)


Total Informal+Algospeak Hits
Total Formal Hits
YouTube
Facebook
TikTok
LinkedIn
Instagram
Reddit
Pinterest
X (Twitter)
Tumblr
Bluelight
drugs-forum

b)


Total Informal Hits
Total Algospeak Hits
YouTube
Facebook
LinkedIn
TikTok
Reddit
Instagram
Tumblr
Pinterest
X (Twitter)
Bluelight
drugs-forum


a)
Informal+Algospeak Normalized Ratio, Scaled
Formal Normalized Ratio, Scaled
Bluelight
drugs-forum
Facebook
TikTok
YouTube
Reddit
LinkedIn
X (Twitter)
Instagram
Pinterest
Tumblr


b)
Informal Normalized Ratio, Scaled
Algospeak Normalized Ratio, Scaled
drugs-forum
Bluelight
YouTube
Reddit
X (Twitter)
Facebook
TikTok
Instagram
LinkedIn
Pinterest
Tumblr