PLOS Digital Health

RESEARCH ARTICLE

Which social media platforms facilitate monitoring the opioid crisis?

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

1 Department of Biomedical Data Science, Stanford University, Stanford, California, United States of America, **2** Department of Epidemiology and Population Health, Stanford University, Stanford, California, United States of America, **3** Department of Biochemistry, Stanford University, Stanford, California, United States of America, **4** Department of Bioengineering, Stanford University, Stanford, California, United States of America, **5** Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, United States of America, **6** Center for Innovation to Implementation, Veterans Affairs Health Care System, Palo Alto, California, United States of America, **7** Department of Psychology, Stanford University, Stanford, California, United States of America, **8** Department of Genetics, Stanford University, Stanford, California, United States of America, **9** Department of Medicine, Stanford University, Stanford, California, United States of America

☯ These authors contributed equally to this work.
* russ.altman@stanford.edu

Abstract

Social media can provide real-time insight into trends in substance use, addiction, and recovery. Prior studies have used platforms such as Reddit and X (formerly Twitter), but evolving policies around data access have threatened these platforms' usability in research. We evaluate the potential of a broad set of platforms to detect emerging trends in the opioid use disorder and overdose epidemic. From these, we identified 11 high-potential platforms, for which we documented policies regulating drug-related discussion, data accessibility, geolocatability, and prior use in opioid-related studies. We quantified their volume of opioid discussion, including in informal language by including slang generated using a large language model. Beyond the most commonly used Reddit and X/Twitter, the platforms with high potential for use in opioid-related surveillance are TikTok, YouTube, and Facebook. Leveraging a variety of social platforms, instead of merely one, yields broader subpopulation representation and safeguards against reduced data access in any single platform.

Author summary

The opioid crisis is a devastating public health challenge, with particular severity in North America. In order to effectively address opioid addiction and overdose, we must be able to track emerging trends in opioid use in a geographic context. In addition to governmental and academic surveys, forensic laboratory drug

OPEN ACCESS

Citation: Carpenter KA, Nguyen AT, Smith DA, Samori IA, Humphreys K, Lembke A, et al. (2025) Which social media platforms facilitate monitoring the opioid crisis? PLOS Digit Health 4(4): e0000842. https://doi.org/10.1371/journal.pdig.0000842

Editor: Cleva Villanueva, Instituto Politécnico Nacional Escuela Superior de Medicina: Instituto Politecnico Nacional Escuela Superior de Medicina, MEXICO

Received: October 11, 2024

Accepted: April 1, 2025

Published: April 28, 2025

Data availability statement: Code for the analyses contained here is available at https://github.com/kristycarp/opioid-social-media-platforms. All other data are present in the main text and the Supplementary Materials.

Funding: 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; NIH GM153195 to RBA; RBA is supported by the Chan Zuckerberg Biohub. None of the listed sponsors or funders played any role in the study design, data collection, data analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

testing, and overdose death tallies, social media has emerged as a potentially useful data source for monitoring trends in the opioid epidemic. Previous work in this space predominantly relies upon Reddit and X (Twitter). In this work, we went beyond these two platforms to assess a wide range of social platforms for their utility in surveilling opioid trends. We characterized a shortlist of 11 platforms to aid selection of platforms in future opioid surveillance work. TikTok, YouTube, and Facebook emerged as particularly promising alternatives to Reddit and X. Using a wider range of social platforms may improve surveillance of opioid-related problems, in turn potentially better informing measures to mitigate harm from this crisis.

Introduction

Real-time tracking of substance use trends is key to understanding epidemics of addiction, including the opioid crisis. Opioids contribute to morbidity and mortality worldwide, but the United States and Canada have an unusually serious epidemic of addiction and overdose, triggered by opioid overprescription and now mainly driven by illicitly manufactured fentanyl [1–6]. It is therefore particularly urgent to monitor trends in the North American opioid epidemic. Heterogeneity in prominent substances, drug availability, and usage patterns across the epidemic require different intervention strategies, and policy makers must be sensitive to these dynamics. 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 [7]. Moreover, surveys can take months or years to complete and results may not reflect quickly evolving trends in opioid use.

Social media platforms are an 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 [8,9]. These platforms capture unfiltered experiences across diverse populations, including encounters with substances that might not otherwise be disclosed [10]. Thus, social media can aid in identification and geolocation of emerging drugs and in tracking practices among people who use opioids (PWUOs). Understanding these patterns could help policymakers anticipate hotspots for overdoses.

Previous work has used social media data at the individual level to detect opioid misuse [10], flag indicators of addiction [11], and assess individual risk of relapse [12]. A larger body of work is dedicated to tracking population trends in opioid misuse [13] and opioid-related mortality [14,15]. 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 [14]. Other work has triangulated social media data with other surveillance datasets (*e.g.,* emergency department admissions) to develop holistic models [16].

Some prior research found correlations between opioid mentions on social media and opioid use and overdose without distinguishing direct disclosures of opioid use from mentions referring to the opioid crisis generally, popular culture, or the drug use of others [8,14–16]. Relying only upon explicit disclosures of a social media user's opioid use may not reflect true rates of opioid usage as many PWUO will not post about their experiences, and some social media users may post false stories of drug use. Additionally, discussion of opioids in the broader community can yield insight into general themes and sentiments regarding the opioid epidemic, such as stigma or the reception of opioid policy.

Most prior work uses only a few social media platforms, chiefly X (formerly Twitter) and Reddit. However, access to these data sources is dependent on corporate decisions. For example, Pushshift [17] was a popular research tool for extracting Reddit posts, but Reddit started to limit API requests for third-party data access in 2023, effectively disabling Pushshift. A similar phenomenon occurred with the X API contemporaneously. As digital platforms evolve, it is critical to understand the scope of available datasets and potential alternatives.

There are 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 [9,18], a direct evaluation of a broad range of social media platforms is needed to best leverage available social media platforms.

Here, our research objectives were to:

1. Compile a broad list of social media platforms that may be suitable for text-based opioid research,
2. Create a shortlist of platforms to further evaluate for opioid research utility in North America,
3. Quantify the amount of opioid-related discussion present on each shortlisted platform, and
4. Evaluate the censorship policies, data accessibility, geolocatability, and prior use of each shortlisted platform.

We 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 North American opioid epidemic. We use all opioid mentions to capture the full potential of a platform to yield insight on different aspects of opioid research. While this work is limited to the United States and Canada, and the English language, we provide a framework to conduct similar work in other regions and languages. Code for the analyses contained here is available at https://github.com/kristycarp/opioid-social-media-platforms.

Materials and methods

Identifying social media platforms

We compiled a list of social media platforms that may contain content related to opioid use, substance use disorders, and/or addiction treatment and recovery. We used three criteria to select platforms: a platform must either be 1) a general-spurpose social platform with widespread use; 2) used previously in research for drug-related surveillance; or 3) a platform with a history of drug-related transactions.

With the first criterion, we sought to include all current major social media platforms, as any large general-purpose platform for online discourse is likely to include mentions of opioids. We included all platforms with more than 100M monthly active users worldwide, as per the list compiled by Wikipedia [19]. We also included all platforms that the Stanford Digital Economy Lab identified as "digital goods" due to their widespread use and relevance [20].

We anticipated that smaller platforms with focus on drug-related conversation could also be useful data sources. For the second criterion, we searched relevant literature for studies using social media as a data source for drug-related surveillance work, "social media", and "pharmacovigilance". We identified studies and reviews using social media and discussion forums for surveillance of illicit drug use [21–24] and of general adverse drug reactions [21,25–33]. We included all platforms used by the identified studies and reviews.

For the third criterion, we sought to include platforms with a history of illicit drug transactions. We identified popular online marketplaces and social messaging platforms known to have previously been used to coordinate illicit drug sales [34–41].

Creating platform shortlist

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 [42], 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 evaluation. A description of the shortlisting process can be found in Appendix A in S1 Text.

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 opioid-related discussion on each platform. While an imprecise approach, Google search matches allowed us to characterize the general volume of opioid mentions on each platform and make comparisons between platforms in an efficient, standardized manner.

We formatted the 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 generic names and brand names of opioids that are currently key drivers of the opioid epidemic in North America (prescription opioids and fentanyl). While heroin drove the second wave of the North American opioid epidemic, and remains a primary cause of opioid overdose deaths in other regions, we excluded it from our term lists in an effort to narrow our focus to the largest threats in the current North American opioid epidemic. These terms are the same as those used to select for platforms with high opioid discussion in the shortlisting phase (Appendix A in S1 Text, Appendix B in S1 Text).

The "informal" opioid term list includes slang, misspellings, and other brand names. Because language on social media is often informal, such terms are important for capturing the full scale of opioid-related discussion. We previously found that GPT-3 [43] is able to generate slang for drugs of addiction at scale [44]. We took validated GPT-3–generated slang terms for five opioids prominent in the current North American opioid epidemic (four prescription opioids: codeine, morphine, oxycodone, and oxymorphone; and fentanyl) and removed terms that had common non-drug meanings. To keep the scale similar to the formal term analysis, we reduced the list to the 20 terms most frequently generated by GPT-3 (Appendix C in S1 Text).

"Algospeak" is a phenomenon in which users of social media platforms purposefully alter words to evade banning and censorship [45]. For example, a poster on social media may refer to fentanyl as "f3ntanyl." The set of terms generated by GPT-3 did not include algospeak. We created an "algospeak" term list using GPT-4 (Appendix D in S1 Text, Appendix E in S1 Text).

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. To reduce variation, all search queries were executed on January 29, 2024. We tabulated the total number of hits per list to estimate the total volume of opioid discussion on each of the shortlisted platforms. Additionally, we normalized the number of opioid-related hits by the number of hits returned for queries of "household terms", chosen from the Corpus Of Contemporary American English to represent the most common nouns used in the English language (Appendix F in S1 Text, Equations A-C in S1 Text). This ratio represents how prominent opioid discussion is relative to other content on the platform, allowing for comparison between platforms with varying popularity. However, we note that because the success of downstream analysis depends upon having sufficient opioid-related text data, popular platforms with a greater total number of opioid-related hits may be preferable to smaller platforms enriched for opioid-related discussion. In the following analyses, we combine the informal and algospeak normalized ratios, as algospeak is a special case of informal language.

Evaluating content restrictions and censorship policies

We analyzed the content restriction and censorship policies of social media sites by referencing the sites' terms of use and user agreements.

Evaluating data accessibility for academic research purposes

We sourced information regarding data access for research purposes from public information on each website.

Evaluating accessibility of geolocation data

We determined whether geolocation is provided for any accessible data from the data access information pages for each platform.

Identifying geolocation inference strategies

We conducted a literature search to identify previously-used strategies for geolocation inference on social media. In addition to identifying geolocation inference strategies present in studies of shortlisted platforms for opioid-related research (see "Assessing prior use of platforms in research literature"), we conducted PubMed and Google Scholar searches for "geolocation" and each of the shortlisted platform names.

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]).

In addition, we substituted the term "social media" in place of the platform name. Search queries were performed in March 2024; our results cover papers published until that time.

We defined relevant literature as papers that describe a primary research study analyzing data originating from at least one of our 11 shortlisted platforms to obtain information related to opioid use. We excluded review papers, studies describing social media-based interventions, and studies that only used social media as a participant recruitment tool. We evaluated the abstract and, if required, the full text of all papers returned by the described PubMed queries to determine if they met these criteria.

Results

Identifying social media platforms

We took the union of the platforms found from the sources listed in Methods to create a superset of 71 candidate platforms (S1 Table).

Creating platform shortlist

We applied our shortlisting criteria to the 71 platforms iteratively (Fig 1, S2 Table). Four platforms were not active as of July 2023. Twenty-one active platforms did not meet our definition of social media and 11 active platforms were private messaging platforms. Eleven 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, 13 platforms returned fewer than 25,000 query results on the selected opioid keywords.

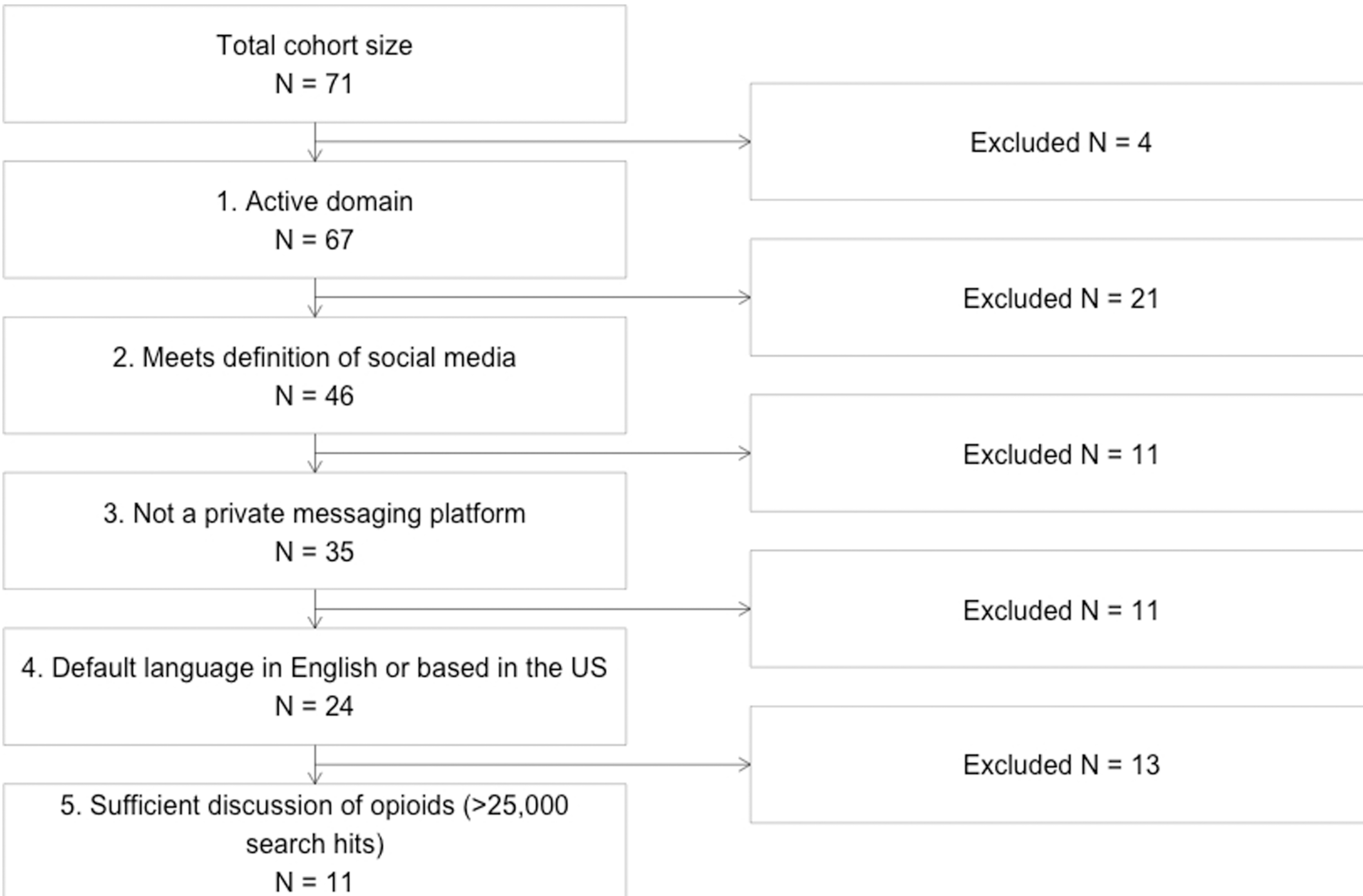

Fig 1. Platform shortlisting consort diagram. Consort diagram of iterative exclusion criteria to attain shortlist of 11 social media platforms for further characterization.

https://doi.org/10.1371/journal.pdig.0000842.g001

This process left 11 (15.5%) 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 opioid-related text varied across platforms (Fig 2, Fig 3, S1 Fig, S3 Table, S4 Table). We observed the highest total volume of opioid-related discussion on YouTube and Facebook (Fig 2a, S1 Fig, S3 Table). These were followed by LinkedIn, TikTok, and Reddit. The platforms with the lowest total volume of opioid-related discussion were Bluelight and drugs-forum.

Most platforms followed a linear relationship between the amount of formal opioid terms and the amount of informal (including algospeak) opioid terms. TikTok and Instagram skewed toward more informal term hits, and LinkedIn, Reddit, and X skewed toward more formal term hits.

The relationship between the total number of informal hits versus the total number of algospeak hits revealed similar trends (Fig 2b). The outliers reveal which platforms have more algospeak, indicating a response to censorship. TikTok by

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, as determined by manual inspection of platforms. "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, as determined by inspection of platform terms of use (provided in Appendix G in S1 Text). "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 acquisition of platform data or to receive more information about collaboration with the platform, and an X if not. API and research portal designations were determined by inspection of platform data availability (provided in Appendix H in S1 Text). "Previously researched for opioid pharmacovigilance" column reflects the relative amount of prior research related to opioid surveillance, with a checkmark indicating some prior use in the literature, two checkmarks indicating high prior use in the literature, and an X indicating no prior use in the literature (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, as determined by inspection of public-facing platform data specifications (see "Evaluating data accessibility for academic research purposes" section of Results). "Example of geolocation inference strategy" column provides one possible method for inferring geolocation of platform content, based on inference strategies previously employed in the broader literature (see "Evaluating data accessibility for academic research purposes section" of Results).

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

https://doi.org/10.1371/journal.pdig.0000842.t001

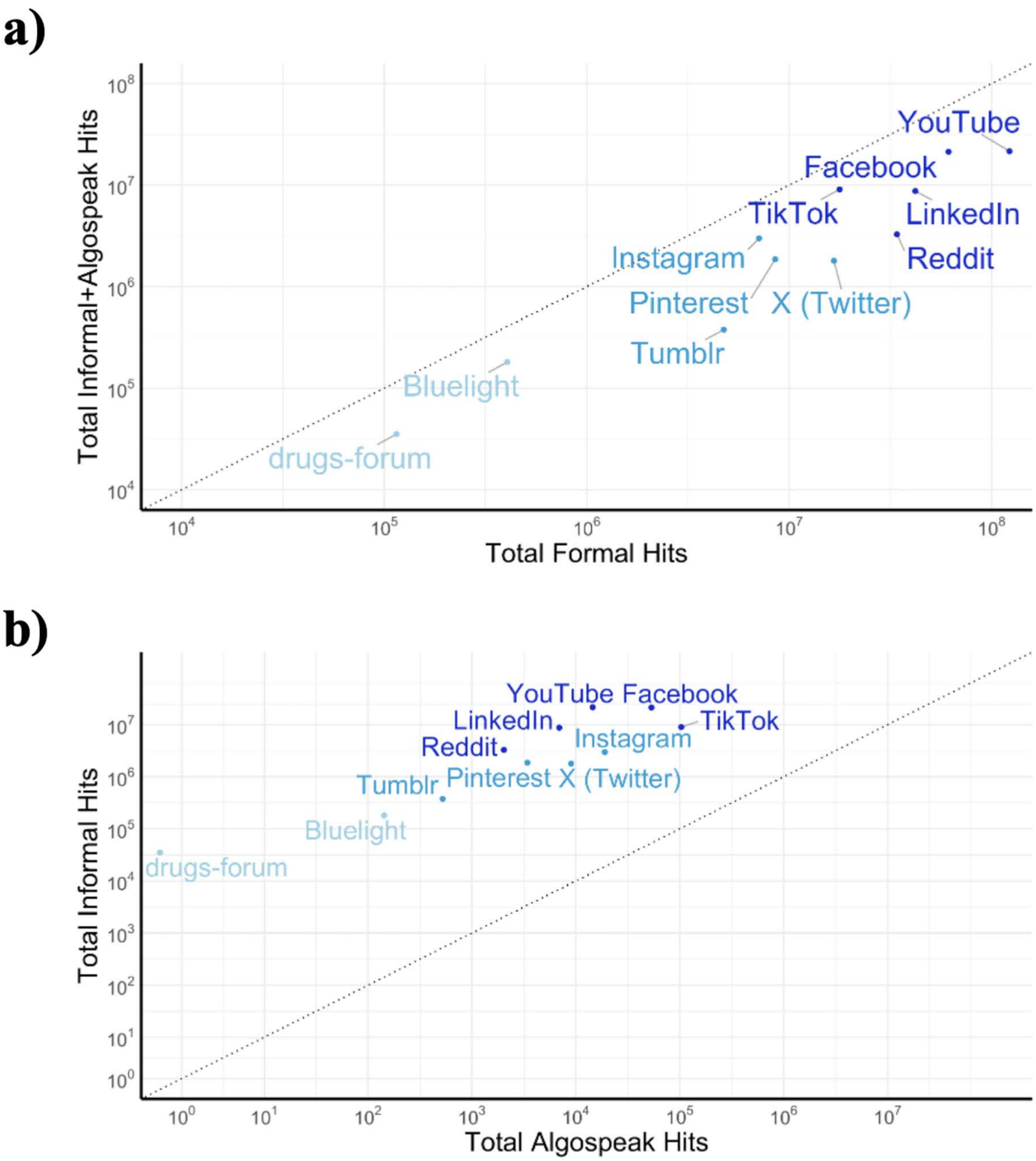

Fig 2. Total hits for each of the shortlisted platforms. Shown with a logarithmic scale. Platforms are clustered into three categories by volume of hits, shown by color (darker shade of blue indicates higher volume of hits). Dotted line shows y = x diagonal. **a)** Total hits for formal opioid term list 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.

https://doi.org/10.1371/journal.pdig.0000842.g002

far had the most amount of algospeak. Instagram and X also skewed toward algospeak. LinkedIn and Reddit both skewed away from algospeak.

While having less drug-related discussion overall, Bluelight and drugs-forum had dramatically higher rates of drug discussion relative to non-drug discussion (Fig 3a). Among the social media platforms with general scope, TikTok and Facebook had the highest relative amount of drug discussion compared to non-drug discussion.

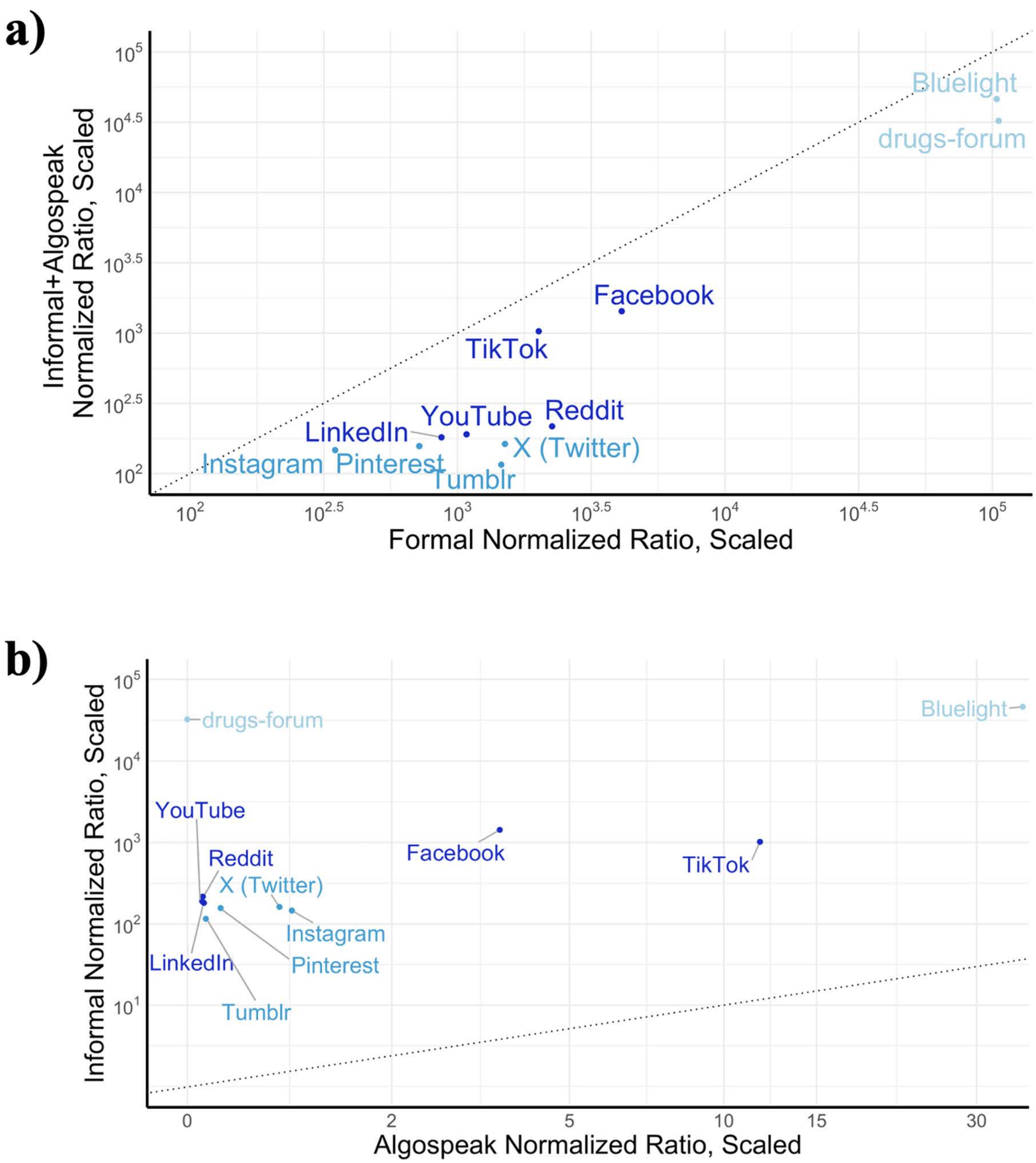

Fig 3. Term hits normalized by number of hits for common nouns. Shown with a logarithmic scale and scaling factor of 100,000. As in Fig 2, platforms are clustered into three categories by total volume of hits, shown by color (darker shade of blue indicates higher volume of hits). Dotted line shows y = 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.

https://doi.org/10.1371/journal.pdig.0000842.g003

We visualized the relative amounts of informal and algospeak term hits for the 11 platforms (Fig 3b). Bluelight and drugs-forum show large separation from the nine general social media platforms, especially with respect to informal (non-algospeak) terms. Of the nine general social media platforms, Facebook and TikTok had the highest relative amounts of drug discussion with respect to both informal and algospeak term hits.

Evaluating content restrictions and censorship policies

All eleven shortlisted platforms state expectations regarding drug-related discussion (Appendix G in S1 Text). High-level classifications of content policies are shown in Table 1, with more detailed comparison shown in Table 2.

All eleven platforms state that the sale of illicit substances is not allowed. Facebook, Instagram, TikTok, and YouTube state that posting content related to recreational use of drugs is prohibited. TikTok underscores that such content is dangerous for young people. These platforms make exceptions for recovery-oriented or educational content. LinkedIn prohibits depictions of "drug abuse." Reddit includes communities focused on drug-related discussion, which are age-restricted. Bluelight and drugs-forum are distinct as forums dedicated to drug-related discussion. Both platforms 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 discounted access to users affiliated with a research institute. In some cases, outside groups have maintained third-party APIs (*e.g.,* Pushshift [17] for Reddit). We found that smaller drug-focused forums do not have APIs that allow for easy data extraction. We provide an overview of data access capabilities in Table 1 and per-platform details in Appendix H in S1 Text.

Evaluating accessibility of geolocation data

Some social media platforms provide geolocation of users or posts. X (formerly Twitter) gives users the option to geotag their tweets. A small subset of users opts to do this (under 2% [46]). Many groups have leveraged geotagged tweets for opioid-related research [47–53] and other research areas [54,55]. It was possible to geotag a tweet with latitude and longitude coordinates until June 2019; since then, users can only tag tweets with place objects that have coordinate bounding boxes [46,56,57].

Similarly, Facebook [58], Instagram [59–62], and YouTube [63–65] give users the option to tag their content with geolocations. Facebook has a Data for Good program [66] which includes province-level GPS data from a subset of consenting

Table 2. Structured comparison of platform content policies.

Platform	Selling or soliciting opioids	Discussing personal recreational opioid use	Discussing recreational use of opioids in a positive manner	Discussing opioid use recovery	Discussing opioid research or policy
Bluelight	✗	✓	✓	✓	✓
Drugs-forum	✗	✓	✓	✓	✓
Facebook	✗	✗	✗	✓	✓
Instagram	✗	✗	✗	✓	✓
LinkedIn	✗	✗	✗	✓	✓
Pinterest	✗	✗	✗	✓	✓
Reddit	✗	✓*	✓*	✓*	✓
TikTok	✗	✗	✗	✓	✓
Tumblr	✗	✓	✓	✓	✓
X (Twitter)	✗	✓	✓	✓	✓
YouTube	✗	✓**	✗	✓	✓

*Only on certain subreddits

**Discussion of use is permitted, but directly showing use is not

A checkmark indicates that a given activity is permitted on a given social media platform, and an X indicates that it is not. Note that this table was constructed based on the content restriction policies as listed in Appendix G in S1 Text and that prohibited content may still be present on platforms.

https://doi.org/10.1371/journal.pdig.0000842.t002

users; previous work leveraged this data to study phenomena related to the COVID-19 pandemic [67,68]. State *et al.* used provided user and company locations to track migration patterns using LinkedIn [69].

Like other social media platforms, TikTok uses user location to personalize their content feed and allows users to add location tags to their videos [70]. Various groups have characterized the privacy aspect of geolocation on TikTok [71,72], but there is little research using tagged locations for geographically-informed analysis. Zanettou *et al.* recruited consenting users to donate their geotagged data, which each user can request under the EU's GDPR regulation [73].

Identifying geolocation inference strategies

Many online forums allow users to indicate their location 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"). Previous work has used the location fields on Bluelight and drugs-forum user profiles to geolocate data to the county- [74] or country-level [29].

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 [75]. Several groups have leveraged the County Tweet Lexical Bank and datasets assembled by similar methods for opioid-related [14,15,52,76,77] and non-opioid-related [78,79] work. Others have used named entity recognition to extract unambiguous place names from social media text [80–83].

Several groups have used the hashtag search functionality on TikTok to obtain content relevant to geographic regions (*e.g.,* specific countries) [84,85]. Similarly, Hu and Conway pulled text from country-specific subreddits as a proxy for geolocation [86]; Delbruel *et al.* examined the association between YouTube video tags and geolocation [87]; 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 [88].

A key feature of social networks is that users are connected (through "friendship", "following", *etc.*) to other users with whom they often share characteristics or interests; 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 [89–91] because it requires the social network being partially labeled with geolocation.

Examining a social media 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) [8,92,93]. Others have also searched for posts with the topic of "where are you from?" or instances of the phrase "I live in…" [93,94] and leveraged user "flairs" (tags that Reddit users can add to their username) [93] to geolocate Reddit users.

Numerous packages for predicting geolocation from social media data exist. Free packages used previously for geolocating social media text [13,95] include Carmen [96] and geopy [97]; paid services used previously for geolocating social media text [98–100] include Iconosquare [101], Brandwatch (formerly Crimson Hexagon) [102], and Reputation (formerly Nuvi) [103]. Additionally, many other groups have created geolocation inference methods [58,104–107].

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 geolocation, Tumblr users can share their posts to X. Xu *et al.* leveraged this cross-posting to obtain geolocation information for Tumblr users [83,108].

We summarize geolocation inference strategies for each of the shortlisted platforms in Table 1.

Prior use in research literature

Of all shortlisted platforms, X and Reddit were the most commonly used in existing literature. The work using these two platforms ranges from correlating opioid-related discussion volume and opioid-related overdose death rates [8,13–16,51,52,76,77], characterizing trends and themes in online discussion of opioids, opioid use, and OUD treatment [10,47–50,76,80,93,109–138], and characterizing public sentiment towards the opioid epidemic [49,88,100,114,128,139,140]. Many groups have created models to identify posts on X and Reddit related to opioids [52,53,115,141–145], and one group created a pipeline to infer demographic information of the identified cohort [146]. Others have used these platforms to characterize factors that influence opioid use, recovery, and the opioid epidemic generally [12,14,147–151], with particular interest shown to the impact of the COVID-19 pandemic [109,126,152–160] and co-use between opioids and other drugs [161,162].

The platforms with the next most volume of prior work were Instagram and Facebook. Previous work used these data to examine aspects of opioid use [163–165], to characterize content related to drug sales [22,166–168], to identify emerging psychoactive substances [127], and to study public reactions to the opioid epidemic [88,169,170]. Though Instagram is an image-focused platform, nearly all studies used text from captions and comments for analysis. Facebook text data was obtained from public groups and pages.

YouTube, Tumblr, and drugs-forum have seen modest usage in text-based research focused on the opioid epidemic. YouTube comments have been analyzed with NLP techniques to characterize sentiments toward opioid use [171] and the opioid epidemic [172]. YouTube and drugs-forum were both used to characterize misinformation related to OUD medications [111,137]. Catalani *et al.* searched YouTube and Tumblr for content about emerging psychoactive substances [127]. Others have used Tumblr text data to detect opioid sales [22,173,174]. Paul *et al.* [29] and Lee *et al.* [175] used drugs-forum text data to establish correlations with NSDUH survey data. Drugs-forum has also been used to monitor other drugs [176,177].

We did not find relevant prior work using Pinterest, TikTok, Bluelight, or LinkedIn. Previous work used Pinterest and TikTok to study portrayals of drugs other than opioids [178–185]. Vosburg *et al.* [186] and Soussan and Kjellgren [187] used Bluelight in the context of recruiting participants for studies related to opioid use. Our PubMed query returned no results related to opioid research for LinkedIn.

We provide a structured comparison of relevant existing literature using the shortlisted platforms in S6 Table.

Discussion

Social media data signals correlate with trends in the opioid crisis. Although existing research has leveraged social media platforms to analyze such phenomena, most research has focused on only a few of the 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 North American opioid addiction and overdose epidemic.

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 crisis surveillance system include degree of censorship, user base demographics, and geolocatability. In addition, the accessibility and stability of these data sources affect their utility for public health research or surveillance. Although we highlight APIs for shortlisted platforms, not all drug-related discussion or user metadata are available through official APIs. Recently, some platforms have shifted from freely available APIs to paywalled versions, which may be cost-prohibitive for large-scale projects. Evolving user language makes it challenging to capture relevant discussions, as with the emergence of algospeak. While all shortlisted platforms are potential sources for surveillance methods, we highlight TikTok, Facebook, and YouTube as underutilized platforms with significant opioid-related content. Reddit and X have a rich existing body of research and remain valuable data sources, but as their access has become restricted, additional sources are necessary.

In this work, we did not differentiate between opioid mentions in the context of self-disclosure of opioid use and other types of mentions because this distinction is not necessary for correlation with opioid mortality rates or uncovering themes in perception of opioids [8]. Different downstream use cases will require different decisions on how to narrow the scope of opioid mentions for analysis. In prior work, studies of addiction, recovery, and relapse tended to only include posts from individuals who were likely to be substance users. These users were often identified through targeted searches on specific forums (ex: the "opiates" community on Reddit) [12], manual annotation by subject matter experts [10], or automated annotations using machine learning models [11]. In contrast, studies that aim to model or forecast population-level trends in mortality and overdose typically include all detected opioid mentions, and in some cases, the full corpus of extracted posts [13–15,51]. In these cases, posts are represented by data-driven topics using methods such as latent Dirichlet allocation. There have been some efforts to classify posts that contain opioid-keywords into finer-grain categories such as specific types of opioid misuse and information sharing [52,53], but these approaches have not been widely adopted.

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 freedom in discussing stigmatized topics. This increases the amount of opioid mentions on a platform, as we observed. Platforms with high levels of moderation may spur algospeak usage. This phenomenon has been described in the literature with respect to TikTok [45,188,189].

Demographics play a key role in selecting social media platforms for surveillance of the opioid crisis. Different social media platforms have different demographic makeups (S5 Table). 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 [190]. Individuals who initiate opioid use at a younger age are more susceptible to substance use disorders [191]; monitoring discussions around opioid use in younger people could help inform preventative programs. Facebook skews older, with 75% of 30–49 year olds reporting that they use the platforms as opposed to 67% of 18–29 year olds [192]. 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 [192]. Using multiple social media platforms in order to monitor different demographics may be required in the context of the opioid epidemic. The subset of platform users discussing opioids may have a different demographic makeup than the overall platform user base; accordingly, demographic attributes such as race, age, and gender for opioid-mentioning users are critical for evaluating whether populations disproportionately impacted by the opioid epidemic are represented. Such pipelines have been developed for social media generally [193–196] and in the context of substance use [146]. Many of these tools have demonstrated good performance at the population scale. However, they are not suitable for estimating demographic attributes (especially race and gender) of individual social media accounts, as these are sensitive and may require additional human subject research protections. They may also introduce stereotype, error, and erasure of minority identities.

Most prior work in social media pharmacovigilance has been conducted using X and Reddit. They have high volume of content (both in total and drug-related), formerly freely accessible APIs, and inferable or explicitly-provided geolocation data. However, other less-researched platforms share these attributes. We have identified potential 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 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 [192], 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. These four platforms may provide insight supplementary to that from X and Reddit.

API access to social media text and geolocation data can be limited

It is not always possible to retrieve all data collected by a social media platform through its API. Many social media platforms infer geolocation and demographic information from user activity, but will omit such identifiers from available datasets. Some platforms only allow queries from a 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 [88,165,169,170,197]. Web scraping and third-party APIs may violate a platform's terms of service.

Many platforms allow users to post content that is not publicly viewable and cannot be obtained through APIs or web scraping. Users may discuss opioid use on social media among their private networks, but share more filtered accounts publicly. If there are differences in users that engage in public versus private discourse, data will be subject to selection biases. On platforms with pseudo-anonymous users, private posts may be less prevalent.

An early-warning system for trends in the opioid crisis requires geolocatability. However, even when geolocation data can be obtained directly from the platform, users sometimes use geotags for purposes other than reporting the location from where a given post was made [198]. Geolocation inference methods can predict geolocation for non-geotagged posts, but the accuracy of such methods decreases with time [199].

Changing business models affect platform availability and stability

The policies of social media companies are rapidly evolving, leading to unstable data access for research. For example, efforts to monetize the primary Reddit API effectively disabled the third-party Pushshift API, which had previously facilitated the high volume of prior research using the platform. External political factors can also affect social media platforms. Many jurisdictions, including Canada, the United Kingdom, Australia, and Taiwan banned TikTok from government devices due to security concerns, and India has banned the app nationwide [200]. The United States passed legislation to ban TikTok if the app is not sold by its Chinese parent company to a US-approved buyer [201]. Therefore, although we have spotlighted TikTok as an underexplored platform with potential to monitor the opioid epidemic, recent political developments might limit its future use in North America.

Data instability poses a challenge to surveilling opioid use and overdoses longitudinally. Incorporating other data sources could make systems more robust to gaps in data access. Ideally, mechanisms for working directly with social media companies would establish more stable data access.

Limitations

The internet is dynamic. The way in which people use social media platforms to talk about opioids will change with time, and this must be accounted for when analyzing this type of data. Informal language constantly evolves, 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 in the future to update informal opioid term lists. However, workflows based on generative models cannot capture emergent slang terms that differ from those in their training data. To circumvent this limitation of a generative AI approach, the best strategy may be to identify drug-related terms directly from the dataset.

In this review, we highlighted the challenges and opportunities of using text-based data for monitoring the opioid epidemic, but social media platforms also contain images and videos. As automated video transcription becomes more common, we believe that text-based analyses will continue to be an informative, scalable tool for monitoring opioid-related discussions on social media. The inclusion of visual features, such as the presence of drug paraphernalia or use of gestures that signal substance use, could provide valuable insights but would require extensive computational resources that may not be reasonable when conducting population-level research. Nonetheless, advancements in computer vision have allowed for the study of substance use via image and video [202,203], and these tools may become more accessible to opioid researchers in the future.

Similarly to slang terms and algospeak, emojis are commonly used to refer to drugs and bypass automated moderation. For example, the electric plug emoji can refer to a "plug," which is a slang term for a person who sells drugs. In many natural language processing workflows, emojis are processed and analyzed as if they were text, and would be captured in many of the analytical methods that we described. However, the opioid-related terms we used in our search did not include any emojis, given that they have low specificity to opioid-related discussion.

The use of social media data requires careful ethical consideration – particularly with respect to privacy. Even publicly visible social media posts related to opioids are sensitive in nature and may not be suitable to share via research publication because they relate to a vulnerable population. Social media data mining may also disrupt user trust in social media and may lead to altered behavior. Although there is no standard ethical guidance for research on mined social media data [204], good practices include redacting usernames or other identifiable information from shared data, paraphrasing posts instead of publishing direct quotes, and actively engaging with community moderators. Research should not contribute to stigmatization, individual-level surveillance, or harm to vulnerable populations. The ethical issues inherent to analysis of social media data, and suggestions on how to conduct such research ethically, are addressed by Fiesler *et al.* [205].

We used Google search queries as an estimate of the volume of public opioid-related discussion. We use this proxy rather than accessing and processing text-based content from each platform, which would require substantial computing resources. These estimates may be influenced by how Google indexes web pages. Platforms owned by Google, such as YouTube, may be overrepresented in search results. Other platforms may restrict access to unregistered users, impacting how Google can process their webpages [206]. We have provided an initial estimate of the volume of opioid-related discussion on each site, but the actual amount of accessible data may vary.

A 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 epidemic in the United States and Canada. However, OUD is a global problem and analysis beyond North America is needed. We also recognize that there are likely substantial opioid-related discussions in languages other than English within North America. According to the U.S. Census Bureau, over 20% of Americans speak a language other than English at home [207]. Nevertheless, our methods can be repurposed for other countries and languages. For example, to conduct a similar analysis for South America, one should shortlist platforms widely used in South America with significant Spanish and Portuguese language content, as well as construct Spanish and Portuguese term lists with formal and informal terms for the most common opioids in South America. Future research in leveraging social media platforms for pharmacovigilance of the opioid epidemic in regions other than North America and languages other than English would enable a more complete assessment of this international crisis.

Finally, we recognize that social media cannot fully capture all phenomena relevant to the opioid epidemic. Integrating social media data with other sources, such as emergency medical services, drug seizure records, and wastewater analysis, may supplement shortcomings and increase the efficacy of an early-warning system.

Supporting information

S1 Text. Supplementary text. Includes details of platform shortlisting methods, term lists, algospeak generation protocol, per-platform content restriction policies and data access policies, and equations.
(DOCX)

S1 Fig. Number of query results for collections of opioid-related terms. Listed by social media platform. A) Comparison between formal and informal/algospeak language. B) Comparison between informal and algospeak language.
(PNG)

S1 Table. Inclusion criteria for all platforms. Superset of all platforms initially included in analysis, in alphabetical order, with inclusion criteria for each.

(XLSX)

S2 Table. Evaluation of full platform list. Shortlisting criteria applied to all platforms initially included in analysis, in alphabetical order.
(XLSX)

S3 Table. Raw hit counts from Google Search API across the 11 shortlisted platforms. All search queries conducted on January 29 and 30, 2024.
(XLSX)

S4 Table. Formal and informal normalized ratios for each of the shortlisted platforms. Scaling factor of 100,000 applied for readability. All search queries conducted on January 29 and 30, 2024.
(XLSX)

S5 Table. Demographic data of users on social media platforms. Age and gender distributions and approximate number of American users of nine of the shortlisted platforms, with references. Bluelight and drugs-forum are excluded as such statistics are unavailable.
(XLSX)

S6 Table. Structured comparison of prior research using social media platforms. Each study is associated with a general research question or theme (one of: "Recovery/ Harm Reduction/ Treatment", "Active Use/ Production/ Promotion", "Mortality/ Suicide/ Self-harm", "Population Chatter/ Thematic Analysis", "Regulatory/ Compliance", or "Mis-information/ Myth/ Content Quality"), one or more shortlisted platforms, one or more substance categories (with "general" indicating substance use broadly), one or more data types (comments, engagement (likes, views, shares etc.) forum discussion, hashtags, pins, profiles, surveys, text posts, videos), and one or more references.
(XLSX)

Acknowledgments

The authors would like to thank Salvatore Giorgi, Shashanka Subrahmanya, Aadesh Salecha, and Rhana Hashemi for insightful conversations and assistance.

Author contributions

Conceptualization: Johannes C. Eichstaedt, Russ B. Altman.

Data curation: Kristy Carpenter, Anna T. Nguyen, Delaney A. Smith, Issah A. Samori.

Formal analysis: Kristy Carpenter, Anna T. Nguyen, Delaney A. Smith, Issah A. Samori.

Funding acquisition: Delaney A. Smith, Issah A. Samori, Keith Humphreys, Anna Lembke, Mathew V. Kiang, Johannes C. Eichstaedt, Russ B. Altman.

Investigation: Kristy Carpenter, Anna T. Nguyen, Delaney A. Smith, Issah A. Samori.

Methodology: Kristy Carpenter, Anna T. Nguyen, Delaney A. Smith, Issah A. Samori, Johannes C. Eichstaedt, Russ B. Altman.

Project administration: Russ B. Altman.

Resources: Russ B. Altman.

Software: Kristy Carpenter, Anna T. Nguyen, Delaney A. Smith, Issah A. Samori.

Supervision: Keith Humphreys, Anna Lembke, Mathew V. Kiang, Johannes C. Eichstaedt, Russ B. Altman.

Validation: Kristy Carpenter, Anna T. Nguyen, Delaney A. Smith.

Visualization: Kristy Carpenter, Anna T. Nguyen, Johannes C. Eichstaedt.

Writing – original draft: Kristy Carpenter, Anna T. Nguyen, Delaney A. Smith.

Writing – review & editing: Kristy Carpenter, Anna T. Nguyen, Issah A. Samori, Keith Humphreys, Anna Lembke, Mathew V. Kiang, Johannes C. Eichstaedt, Russ B. Altman.

References

1. The Lancet Regional Health – Americas. Opioid crisis: addiction, overprescription, and insufficient primary prevention. Lancet Reg Health Am. 2023 Jul;23:100557. https://doi.org/10.1016/j.lana.2023.100557 PMID: 37497399
2. 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;272(6):879–86. https://doi.org/10.1097/SLA.0000000000004225 PMID: 32657939
3. Manchikanti L, Singh A. Therapeutic opioids: a ten-year perspective on the complexities and complications of the escalating use, abuse, and non-medical use of opioids. Pain Physician. 2008;11:S63–88. PMID: 18443641
4. Humphreys K, Shover CL, Andrews CM, Bohnert ASB, Brandeau ML, Caulkins JP, et al. Responding to the opioid crisis in North America and beyond: recommendations of the Stanford–Lancet Commission. The Lancet. 2022 Feb;399(10324):555–604. https://doi.org/10.1016/S0140-6736(21)02252-2 PMID: 35122753
5. Kalkman GA, Kramers C, van den Brink W, Schellekens AFA. The North American opioid crisis: a European perspective. Lancet. 2022 Oct;400(10361):1404. https://doi.org/10.1016/S0140-6736(22)01594-X PMID: 36273479
6. Brown R, Morgan A. The opioid epidemic in North America: Implications for Australia. Trends Issues Crime Crim Justice. 2019;(578).
7. Palamar JJ. Barriers to accurately assessing prescription opioid misuse on surveys. Am J Drug Alcohol Abuse. 2019;45(2):117–23. https://doi.org/10.1080/00952990.2018.1521826 PMID: 30230924
8. Lavertu A, Hamamsy T, Altman RB. Monitoring the opioid epidemic via social media discussions [Preprint]. 2021 [cited 2024 Mar 14]. p. 2021.04.01.21254815. Available from: https://www.medrxiv.org/content/10.1101/2021.04.01.21254815v1
9. 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. Addiction. 2023 Feb;118(2):206–17. https://doi.org/10.1111/add.16020 PMID: 36075258
10. 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.
11. Fan Y, Zhang Y, Ye Y, li X, Zheng W. Social Media for Opioid Addiction Epidemiology. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017:1259–67. https://doi.org/10.1145/3132847.3132857
12. Yang Z, Nguyen L, Jin F. Predicting Opioid Relapse Using Social Media Data [Preprint]. arXiv; 2018 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/1811.12169
13. 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. https://doi.org/10.1007/s13181-017-0625-5 PMID: 28831738
14. 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. https://doi.org/10.1038/s41598-023-34468-2 PMID: 37270657
15. 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.
16. Sumner SA, Bowen D, Holland K, Zwald ML, Vivolo-Kantor A, Guy GP, 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.
17. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift reddit dataset. Proceedings of the International AAAI Conference on Web and Social Media. 2020:830–9. https://ojs.aaai.org/index.php/ICWSM/article/view/7347
18. 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. https://doi.org/10.1055/s-0039-1677918 PMID: 31419834
19. List of social platforms with at least 100 million active users. Wikipedia. 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
20. 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
21. 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.
22. 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;8:100235.

23. 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.
24. 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. https://doi.org/10.1080/10826084.2016.1276600 PMID: 28394703
25. 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. https://doi.org/10.1016/j.jbi.2015.02.004 PMID: 25720841
26. Lee J-Y, Lee Y-S, 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;7(6):e30137. https://doi.org/10.2196/30137 PMID: 34185021
27. 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.
28. 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. https://doi.org/10.1111/bcp.12717 PMID: 26147850
29. 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. https://doi.org/10.1097/ADM.0000000000000238 PMID: 27466069
30. 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.
31. 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.
32. Fung IC-H, 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. https://doi.org/10.7812/TPP/18.307 PMID: 31852039
33. 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;102:130–7
34. Tofighi B, Perna M, Desai A, Grov C, Lee J. Craigslist as a source for heroin: a report of two cases. J Subst Use. 2016;21(5):543–6.
35. 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. https://doi.org/10.1136/tobaccocontrol-2018-054499 PMID: 30185531
36. 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 ; 38(4):846–8.
37. Williams RS. Underage internet alcohol sales on eBay. Addiction. 2013;108(7):1346–8. https://doi.org/10.1111/add.12182 PMID: 23587121
38. Haber I, Pergolizzi J, LeQuang JA. Poppy seed tea: a short review and case study. Pain Ther. 2019 Jun; 8(1):151–155.
39. 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
40. Kim E. Etsy blocks sales of drugs and human remains. CNN Business [Internet]. 2012 Aug 10; Available from: https://money.cnn.com/2012/08/10/technology/etsy-bans-drugs/index.html
41. Price R. Etsy is awash with illicit products it claims to ban, from ivory to dangerous weapons and mass-produced goods. Business Insider [Internet]. 2021 Apr 30; Available from: https://www.businessinsider.com/etsy-sells-ivory-weapons-poisonous-plants-mass-produced-products-2021-4
42. 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
43. 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. Curran Associates, Inc; 2020 [cited 2024 Mar 14]. p. 1877–901. Available from: https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html
44. 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.
45. 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.
46. Advanced filtering for geo data [Internet]. [cited 2024 Mar 14]. Available from: https://developer.twitter.com/en/docs/tutorials/advanced-filtering-for-geo-data
47. Flores L, Young SD. Regional variation in discussion of opioids on social media. J Addict Dis. 2021;39(3):316–21. https://doi.org/10.1080/10550887.2021.1874804 PMID: 33570000
48. 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.
49. Tibebu S, Chang VC, Drouin C-A, 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. 2018 Jun;38(6):263–7. https://doi.org/10.24095/hpcdp.38.6.08 PMID: 29911824

50. 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.
51. 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.
52. 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. https://doi.org/10.1001/jamanetworkopen.2019.14672 PMID: 31693125
53. 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. https://doi.org/10.1016/j.compbiomed.2020.104132 PMID: 33290931
54. Yin J, Gao Y, Chi G. An evaluation of geo-located twitter data for measuring human migration. Int J Geogr Inf Sci. 2022;36(9):1830–52. https://doi.org/10.1080/13658816.2022.2075878 PMID: 36643847
55. Kellert O, Matlis NH. Geolocation of multiple sociolinguistic markers in Buenos Aires. PLoS One. 2022 Sep 9;17(9):e0274114.
56. 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.
57. Kruspe A, Häberle M, Hoffmann EJ, Rode-Hasinger S, Abdulahhad K, Zhu XX. Changes in Twitter geolocations: Insights and suggestions for future usage [Preprint]. arXiv; 2021 [cited 2024 Mar 14]. http://arxiv.org/abs/2108.12251
58. 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), 2018 [cited 2024 Mar 14]. p. 441–6. https://ieeexplore.ieee.org/abstract/document/8508816
59. Correia RB, Li L, Rocha LM. Monitoring potential drug interactions and reactions via network analysis of instagram user timelines. Pac Symp Biocomput. 2016;21:492–503. PMID: 26776212
60. Allem JP, Chu KH, Cruz TB, Unger JB. Waterpipe Promotion and Use on Instagram: #Hookah. Nicotine Tob Res. 2017;19(10):1248–52.
61. 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. https://doi.org/10.1073/pnas.1614158113 PMID: 27799537
62. Mukhina KD, Rakitin SV, Visheratin AA. Detection of tourists attraction points using Instagram profiles. Procedia Comput Sci. 2017 Jan 1; 108:2378–82.
63. 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.
64. Friedland G, Vinyals O, Darrell T. Multimodal location estimation. In: Proceedings of the 18th ACM international conference on Multimedia. Firenze Italy: ACM; 2010 [cited 2024 Sep 24]. p. 1245–52. https://doi.org/10.1145/1873951.1874197
65. 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). 2011 [cited 2024 Mar 14]. p. 207–10. Available from: https://ieeexplore.ieee.org/abstract/document/6085945
66. Data for good tools and data. [cited 2024 Mar 14]. Available from: https://dataforgood.facebook.com/dfg/tools.
67. 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. https://doi.org/10.3390/ijgi10020073
68. ZarembaK.Opening of hotels and ski facilities: Impact on mobility, spending, and Covid-19 outcomes.Health Econ. 2023;32(5):1148–80.
69. 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. 2014; Cham: Springer International Publishing; 531–543. [cited 2024 Mar 14]. https://doi.org/10.1007/978-3-319-13734-6_37
70. 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
71. 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.
72. Franqueira VNL, Annor JA, Kafali O. Age Appropriate Design: Assessment of TikTok, Twitch, and YouTube Kids [Preprint]. arXiv; 2022 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/2208.02638
73. 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 [Preprint]. arXiv; 2023 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/2301.04945
74. 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;44:121–9. https://doi.org/10.1016/j.drugpo.2017.05.007 PMID: 28578250
75. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski LA, Agrawal M, Park G, et al. Characterizing Geographic Variation in Well-Being Using Tweets. In: Proceedings of the International AAAI Conference on Web and Social Media. [cited 2024 Mar 14]. p. 583–91. https://aaai.org/papers/00583-14442-characterizing-geographic-variation-in-well-being-using-tweets/

76. 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.
77. 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;6(2):e17574. https://doi.org/10.2196/17574 PMID: 32469322
78. 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;17(4):1125. https://doi.org/10.3390/ijerph17041125 PMID: 32053866
79. 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. https://doi.org/10.1371/journal.pone.0194290 PMID: 29617408
80. Jha D, Singh R. SMARTS: the social media-based addiction recovery and intervention targeting server. Bioinformatics. 2019 Oct 24; btz800.
81. 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. https://doi.org/10.1371/journal.pone.0244317 PMID: 33534801
82. 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.
83. Compton R, Keegan MS, Xu J. Inferring the geographic focus of online documents from social media sharing patterns [Preprint]. arXiv; 2014 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/1406.2392
84. Pinto G, Burghardt K, Lerman K, Ferrara E. GET-Tok: A GenAI-Enriched Multimodal TikTok Dataset Documenting the 2022 Attempted Coup in Peru [Preprint]. arXiv; 2024 [cited 2024 Mar 14]. http://arxiv.org/abs/2402.05882
85. Bandy J, Diakopoulos N. #TulsaFlop: A Case Study of Algorithmically-Influenced Collective Action on TikTok [Preprint]. arXiv; 2020 [cited 2024 Mar 14]. Available from: http://arxiv.org/abs/2012.07716
86. 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. https://doi.org/10.2196/36941 PMID: 36196144
87. Delbruel S, Frey D, Taïani F. Exploring the Geography of Tags in Youtube Views [Internet]. 2015 May [cited 2024 Mar 14]. Report No.: 461. Available from: https://inria.hal.science/hal-01157867
88. 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.
89. Scalia G, Francalanci C, Pernici B. CIME: Context-aware geolocation of emergency-related posts. GeoInformatica. 2022 Jan 1; 26(1): 125–57.
90. 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. Proceedings of the International AAAI Conference on Web and Social Media 2015 [cited 2024 Mar 14]. p. 188–97. https://ojs.aaai.org/index.php/ICWSM/article/view/14627
91. Pellet H, Shiaeles S, Stavrou S. Localising social network users and profiling their movement. Comput Secur. 2019 Mar 1; 81: 49–57.
92. 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.
93. 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. 2019 [cited 2024 Mar 14]. p. 2572–9. Available from: http://arxiv.org/abs/1904.00003
94. Harrigian K. Geocoding Without Geotags: A Text-based Approach for reddit. In: Xu W, Ritter A, Baldwin T, Rahimi A, editors. Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text. Brussels, Belgium: Association for Computational Linguistics; 2018 [cited 2024 Mar 14]. p. 17–27. Available from: https://aclanthology.org/W18-6103
95. 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.
96. Dredze M, Paul MJ, Bergsma S, Tran H. Carmen: A Twitter geolocation system with applications to public health. In 2013. p. 20–4.
97. geopy: Python Geocoding Toolbox [Internet]. [cited 2024 Mar 14]. Available from: https://github.com/geopy/geopy
98. 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;:1–5. PMID: 35739365
99. 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. 2019 Jul;65(1):94–100. https://doi.org/10.1016/j.jadohealth.2018.12.025 PMID: 30819581
100. 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.
101. Iconosquare - Analytics and management for Instagram, TikTok, LinkedIn, Twitter & Facebook. [cited 2024 Mar 14]. Available from: https://www.iconosquare.com/
102. Brandwatch [Internet]. [cited 2024 Mar 14]. Brandwatch | The social suite of the future. Available from: https://www.brandwatch.com/
103. Reputation [Internet]. [cited 2024 Mar 14]. Reputation | Online Reputation Management for Business. Available from: https://reputation.com/

104. Mahajan R, Mansotra V. Predicting Geolocation of Tweets: Using Combination of CNN and BiLSTM. Data Sci Eng. 2021;6(4):402–10. https://doi.org/10.1007/s41019-021-00165-1 PMID: 34254044

105. Zola P, Cortez P, Carpita M. Twitter user geolocation using web country noun searches. Decis Support Syst. 2019 May 1; 120:50–9.

106. 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.

107. Cha M, Gwon Y, Kung H. Twitter Geolocation and Regional Classification via Sparse Coding. Proceedings of the International AAAI Conference on Web and Social Media. 2015 [cited 2024 Mar 14]. p. 582–5. Available from: https://ojs.aaai.org/index.php/ICWSM/article/view/14664

108. Xu J, Compton R, Lu T-C, Allen D. Rolling through tumblr: characterizing behavioral patterns of the microblogging platform. In: Proceedings of the 2014 ACM conference on Web science. New York, NY, USA: Association for Computing Machinery; 2014 [cited 2024 Mar 14]. 13–22. (Web-Sci'14). https://dl.acm.org/doi/10.1145/2615569.2615694

109. 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. https://doi.org/10.1080/10826084.2021.1972317 PMID: 34486471

110. 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.

111. 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. https://doi.org/10.2196/30753 PMID: 34941555

112. 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). 2022;60(6):694–701. https://doi.org/10.1080/15563650.2022.2032730 PMID: 35119337

113. 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.

114. 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. https://doi.org/10.1371/journal.pone.0135072 PMID: 26252774

115. 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. https://doi.org/10.34133/2022/9851989 PMID: 37621877

116. Shutler L, Nelson L S, 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.

117. 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. https://doi.org/10.1016/j.addbeh.2016.08.019 PMID: 27568339

118. 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. https://doi.org/10.1186/s44247-023-00029-w PMID: 37680768

119. 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;272:5–8. https://doi.org/10.3233/SHTI200479 PMID: 32604586

120. Haug NA, Bielenberg J, Linder SH, Lembke A. Assessment of provider attitudes toward #naloxone on Twitter. Subst Abus. 2016;37(1):35–41. https://doi.org/10.1080/08897077.2015.1129390 PMID: 26860229

121. Chenworth M, Perrone J, Love JS, Graves R, Hogg-Bremer W, Sarker A. Methadone and suboxone mentions on twitter: thematic and sentiment analysis. Clin Toxicol (Phila). 2021 Nov;59(11):982–91. https://doi.org/10.1080/15563650.2021.1893742 PMID: 33821724

122. 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;181:5–10. https://doi.org/10.1016/j.drugalcdep.2017.09.010 PMID: 29024875

123. 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.

124. 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.

125. 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. https://doi.org/10.1080/10826084.2023.2196574 PMID: 37021375

126. 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.

127. 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. https://doi.org/10.3389/fpsyt.2020.632405 PMID: 33633599

128. Nasralah T, El-Gayar O, Wang Y. Social Media Text Mining Framework for Drug Abuse: Development and Validation Study With an Opioid Crisis Case Analysis. J Med Internet Res. 2022 Aug 13; 22(8):e18350.

129. Carabot F, Fraile-Martínez O, Donat-Vargas C, Santoma J, Garcia-Montero C, Pinto da Costa M, et al. Understanding Public Perceptions and Discussions on Opioids Through Twitter: Cross-Sectional Infodemiology Study. J Med Internet Res. 2023 Oct 31; 25:e50013.

130. Chenworth M, Perrone J, Love JS, Greller HA, Sarker A, Chai PR. Buprenorphine Initiation in the Emergency Department: a Thematic Content Analysis of a #firesidetox Tweetchat. J Med Toxicol. 2020;16(3):262–8. https://doi.org/10.1007/s13181-019-00754-7 PMID: 31898154

131. Carabot F, Donat-Vargas C, Santoma-Vilaclara J, Ortega MA, García-Montero C, Fraile-Martínez O, et al. Exploring Perceptions About Paracetamol, Tramadol, and Codeine on Twitter Using Machine Learning: Quantitative and Qualitative Observational Study. J Med Internet Res. 2023 Nov 14;25: e45660.

132. Sarker A, Lakamana S, Guo Y, Ge Y, Leslie A, Okunromade O, et al. #ChronicPain: Automated Building of a Chronic Pain Cohort from Twitter Using Machine Learning. Health Data Sci. 2023;3:0078. https://doi.org/10.34133/hds.0078 PMID: 38333075

133. Spadaro A, O'Connor K, Lakamana S, Sarker A, Wightman R, Love JS, et al. Self-reported Xylazine Experiences: A Mixed-methods Study of Reddit Subscribers. J Addict Med. 2023;17(6):691–4. https://doi.org/10.1097/ADM.0000000000001216 PMID: 37934533

134. Whitman CB, Reid MW, Corey A, Patel H, Ursos L, Sa'adon R, et al. Balancing opioid-induced gastrointestinal side effects with pain management: Insights from the online community. J Opioid Manag. 2015 Sep 1;11(5):383–91.

135. Cai M, Shah N, Li J, Chen W-H, Cuomo RE, Obradovich N, et al. Identification and characterization of tweets related to the 2015 Indiana HIV outbreak: A retrospective infoveillance study. PLoS One. 2020 Aug 26;15(8):e0235150. https://doi.org/10.1371/journal.pone.0235150 PMID: 32845882

136. Smith KE, Rogers JM, Schriefer D, Grundmann O. Therapeutic benefit with caveats?: Analyzing social media data to understand the complexities of kratom use. Drug Alcohol Depend. 2021;226:108879. https://doi.org/10.1016/j.drugalcdep.2021.108879 PMID: 34216869

137. ElSherief M, Sumner S, Krishnasamy V, Jones C, Law R, Kacha-Ochana A, et al. Identification of Myths and Misinformation About Treatment for Opioid Use Disorder on Social Media: Infodemiology Study. JMIR Form Res. 2024;8:e44726. https://doi.org/10.2196/44726 PMID: 38393772

138. Tofighi B, El Shahawy O, Segoshi A, Moreno KP, Badiei B, Sarker A, et al. Assessing perceptions about medications for opioid use disorder and Naloxone on Twitter. J Addict Dis. 2021;39(1):37–45. https://doi.org/10.1080/10550887.2020.1811456 PMID: 32835641

139. 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.

140. Ramachandran S, Brown L, Ring D. Tones and themes in Reddits posts discussing the opioid epidemic. J Addict Dis. 2022;40(4):552–8. https://doi.org/10.1080/10550887.2022.2049170 PMID: 35274598

141. Garg S, Taylor J, El Sherief M, Kasson E, Aledavood T, Riordan R, et al. Detecting risk level in individuals misusing fentanyl utilizing posts from an online community on Reddit. Internet Interv. 2021;26:100467. https://doi.org/10.1016/j.invent.2021.100467 PMID: 34804810

142. 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. 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

143. 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. https://doi.org/10.2196/33919 PMID: 35353047

144. 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.

145. Al-Garadi MA, Yang YC, Cai H, Ruan Y, O'Connor K, Graciela GH, et al. Text classification models for the automatic detection of nonmedical prescription medication use from social media. BMC Med Inform Decis Mak. 2021 Dec;21(1):27.

146. Yang Y-C, Al-Garadi MA, Love JS, Cooper HLF, Perrone J, Sarker A. Can accurate demographic information about people who use prescription medications nonmedically be derived from Twitter?. Proc Natl Acad Sci U S A. 2023 Feb 21;120(8):e2207391120. https://doi.org/10.1073/pnas.2207391120 PMID: 36787355

147. 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.

148. 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;24(8):e36555. https://doi.org/10.2196/36555 PMID: 35994333

149. 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. https://doi.org/10.1080/10826084.2022.2091786 PMID: 35815614

150. Chen AT, Johnny S, Conway M. Examining stigma relating to substance use and contextual factors in social media discussions. Drug Alcohol Depend Rep. 2022;3:100061. https://doi.org/10.1016/j.dadr.2022.100061 PMID: 36845987

151. 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. https://doi.org/10.3389/fpubh.2023.1141093 PMID: 37151596

152. Alambo A, Padhee S, Banerjee T, Thirunarayan K. COVID-19 and Mental Health/Substance Use Disorders on Reddit: A Longitudinal Study [Preprint]. arXiv; 2020 [cited 2024 Mar 14]. https://arxiv.org/abs/2011.10518v1

153. 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;84(2):222–9. https://doi.org/10.15288/jsad.21-00438 PMID: 36971722

154. 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.

155. 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;92:103140. https://doi.org/10.1016/j.drugpo.2021.103140 PMID: 33558165

156. Sarker A, Nataraj N, Siu W, Li S, Jones CM, Sumner SA. Concerns among people who use opioids during the COVID-19 pandemic: a natural language processing analysis of social media posts. Subst Abuse Treat Prev Policy. 2022;17(1):16. https://doi.org/10.1186/s13011-022-00442-w PMID: 35248103

157. 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.

158. 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. https://doi.org/10.1097/ADM.0000000000000883 PMID: 34145186

159. Jiang T, Osadchiy V, Weinberger JM, Zheng MH, Owen MH, Leonard SA, et al. Impact of the COVID-19 Pandemic on Patient Preferences and Decision Making for Symptomatic Urolithiasis. J Endourol. 2021;35(8):1250–6. https://doi.org/10.1089/end.2020.1141 PMID: 33478351

160. Wanchoo K, Abrams M, Merchant RM, Ungar L, Guntuku SC. Reddit language indicates changes associated with diet, physical activity, substance use, and smoking during COVID-19. PLoS One. 2023 Feb 3;18(2):e0280337. https://doi.org/10.1371/journal.pone.0280337 PMID: 36735708

161. 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.

162. 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. https://doi.org/10.1371/journal.pone.0263583 PMID: 35134074

163. 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. https://doi.org/10.3389/fpsyt.2021.551296 PMID: 34135778

164. 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. https://doi.org/10.2196/publichealth.8144 PMID: 29559422

165. 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. https://doi.org/10.1016/j.jhsg.2020.10.005 PMID: 33537664

166. 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.

167. 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 Abus. 2022;43(1):273–7. https://doi.org/10.1080/08897077.2021.1941508 PMID: 34214410

168. 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. https://doi.org/10.1371/journal.pone.0275272 PMID: 36227918

169. 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. https://doi.org/10.1007/s11606-021-06678-9 PMID: 33782896

170. 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.

171. Squires NA, Soyemi E, Yee LM, Birch EM, Badreldin N. Content Quality of YouTube Videos About Pain Management After Cesarean Birth: Content Analysis. JMIR Infodemiology. 2023 Jun 23;3:e40802. https://doi.org/10.2196/40802 PMID: 37351938

172. 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;28(1):53–63. https://doi.org/10.1080/10810730.2023.2176575 PMID: 36755488

173. 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.

174. 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. https://doi.org/10.1109/TCBB.2020.2978476 PMID: 32149652

175. Lee C, St Clair C, Merenda CC, Araojo CR, Ray S, Beasley D, et al. Assessment of public and patient online comments in social media and food and drug administration archival data. Res Soc Adm Pharm. 2020 Jul;16(7):967–73. https://doi.org/10.1016/j.sapharm.2019.10.009 PMID: 31668550

176. 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. 2018;14(2):134–43. https://doi.org/10.1007/s13181-018-0659-3 PMID: 29671244

177. 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.

178. Merten JW, Gordon BT, King JL, Pappas C. Cannabidiol (CBD): Perspectives from Pinterest. Subst Use Misuse. 2020;55(13):2213–20. https://doi.org/10.1080/10826084.2020.1797808 PMID: 32715862

179. 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.

180. 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. https://doi.org/10.1080/10410236.2014.987468 PMID: 26512916

181. 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.

182. 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. https://doi.org/10.1136/tobaccocontrol-2021-056619 PMID: 34312317

183. 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. https://doi.org/10.1111/dar.13433 PMID: 35073422

184. 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.

185. Whelan J, Noller GE, Ward RD. Rolling through TikTok: An analysis of 3,4-methylenedioxymethamphetamine-related content. Drug Alcohol Rev. 2024;43(1):36–44. https://doi.org/10.1111/dar.13640 PMID: 36917507

186. 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.

187. 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

188. VeraVNonsuicidal Self-Injury and Content Moderation on TikTok. Proc Assoc Inf Sci Technol. 2023;60(1)1164–6

189. Moskal M, Supernak N. Do you speak algospeak? An introduction to the recent yet prominent phenomenon of Internet discourse from a cognitive linguistics perspective. Proceedings of Poznańskie Forum Kognitywistyczne. 2023.

190. 190.StatistaU.S. TikTok users by age 2022[cited 2024 Mar 14]https://www.statista.com/statistics/1095186/tiktok-us-users-age/

191. 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.

192. NW 1615L. 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/

193. Morgan-Lopez AA, Kim AE, Chew RF, Ruddle P. Predicting age groups of Twitter users based on language and metadata features. PLoS One. 2017 Aug 29;12(8):e0183537. https://doi.org/10.1371/journal.pone.0183537 PMID: 28850620

194. Pandya A, Oussalah M, Monachesi P, Kostakos P. On the use of distributed semantics of tweet metadata for user age prediction. Future Gener Comput Syst. 2020 Jan;102:437–52.

195. Liu W, Ruths D. What's in a Name? Using First Names as Features for Gender Inference in Twitter. Washington, DC, USA: Association for the Advancement of Artificial Intelligence; 2013.

196. Preoţiuc-Pietro D, Ungar L. User-level race and ethnicity predictors from Twitter text. In: Proceedings of the 27th International Conference on Computational Linguistics. Sante Fe, New Mexico, USA; 2018. p. 1534–45.

197. 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.

198. Baik J. The Geotagging Counterpublic: The Case of Facebook Remote Check-Ins to Standing Rock. Int J Commun. 2020 Mar 13 14:(0):21

199. Dredze M, Osborne M, Kambadur P. Geolocation for Twitter: Timing Matters. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics. 2016. p. 1064–9.

200. Taylor A, Hassan J, Francis E, Mellen R. Where is TikTok banned? These countries restrict the app. The Washington Post [Internet]. 2025 Jan 17; Available from: https://www.washingtonpost.com/world/2025/01/17/countries-banned-tiktok/

201. HaderoH.The rise - and potential fall - of TikTok in the USThe Associated Press2025 Jan 17

202. Roy A, Paul A, Pirsiavash H, Pan S. Automated Detection of Substance Use-Related Social Media Posts Based on Image and Text Analysis. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). Boston, MA: IEEE; 2017 [cited 2025 Jan 17]. p. 772–9. Available from: https://ieeexplore.ieee.org/document/8372025/

203. Hu C, Yin M, Liu B, Li X, Ye Y. Detection of Illicit Drug Trafficking Events on Instagram. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021 [cited 2025 Jan 17]. p. 3838–46. Available from: https://doi.org/10.1145/3459637.3481908

204. Taylor J, Pagliari C. Mining social media data: How are research sponsors and researchers addressing the ethical challenges?. Research Ethics. 2018;14(2):1–39. https://doi.org/10.1177/1747016117738559

205. Fiesler C, Zimmer M, Proferes N, Gilbert S, Jones N. Remember the Human: A Systematic Review of Ethical Considerations in Reddit Research. Proc ACM Hum-Comput Interact. 2024 Feb 16; 8(GROUP):1–33.

206. Peters J. Tweets aren't showing up in Google results as often because of changes at Twitter. The Verge. 2023. [cited 2024 Mar 14]. Available from: https://www.theverge.com/2023/7/3/23783153/google-twitter-tweets-changes-rate-limits

207. Bureau UC. Census.gov. Nearly 68 Million People Spoke a Language Other Than English at Home in 2019. [cited 2024 Mar 14]. Available from: https://www.census.gov/library/stories/2022/12/languages-we-speak-in-united-states.html