medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FRONT MATTER

Title

- Full title: Which social media platforms facilitate monitoring the opioid crisis?
- Short title: Social media for opioid trend monitoring

Authors

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]†**

Affiliations

[1]Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA.

[2]Department of Epidemiology and Population Health, Stanford University, Stanford, CA 94305, USA.

[3]Department of Biochemistry, Stanford University, Stanford, CA 94305, USA.

[4]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA.

[5]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305, USA.

[6]Veterans Affairs Health Care System, Palo Alto, CA 94304, USA.

[7]Department of Psychology, Stanford University, Stanford, CA 94305, USA.

[8]Department of Genetics, Stanford University, Stanford, CA 94305, USA.

[9]Department of Medicine, Stanford University, Stanford, CA 94305, USA.

† These authors contributed equally to this work.

* Pre-publication correspondence: kcarp@stanford.edu.

** Post-publication correspondence: 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 epidemic. From these, we created a shortlist of 11 platforms, for which we documented official policies regulating drug-related discussion, data accessibility, geolocatability, and prior use in opioid-related studies. We quantified their volumes of opioid discussion, capturing informal language by including slang

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

generated using a large language model. Beyond the most commonly used Reddit and X, the platforms with high potential for use in opioid-related surveillance are TikTok, YouTube, and Facebook. Leveraging many different social platforms, instead of a single platform, safeguards against sudden changes to data access and may better capture all populations that use opioids than any single platform.

Teaser

TikTok, Facebook, and YouTube may complement Reddit and X as text sources to monitor trends in the opioid epidemic.

MAIN TEXT

Introduction

Real-time tracking of substance use trends is key to understanding epidemics of addiction, including the North American opioid crisis. 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 (*1*). Moreover, surveys can take months or years to complete and results may not reflect quickly evolving trends in opioid use.

Social media platforms are 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 (*2*, *3*). These platforms capture unfiltered experiences across diverse populations, including encounters with substances that might not otherwise be disclosed (*4*). 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 (*4*), flag indicators of addiction (*5*), and assess individual risk of relapse (*6*). A larger body of work is dedicated to tracking population trends in opioid misuse (*7*) and opioid-related mortality (*8*, *9*). Language use on these platforms may be more predictive of trends in county-level deaths than factors such as demographics, healthcare access, and physical pain (*8*). Other work has triangulated social media data with other surveillance datasets (*e.g.* emergency department admissions) to develop holistic models (*10*).

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 epidemic generally, popular culture, or the drug use of others (*2*, *8–10*). Relying only upon explicit disclosures of a social media user's illicit 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 (*12*) 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

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

phenomenon occurred with the X API contemporaneously. As digital platforms evolve, it is critical to understand the scope of available datasets and potential alternatives. Additionally, different social media platforms have different constituent demographics; no individual platform fully represents the population at large. An effort combining multiple platforms may be better able to capture trends in this crisis.

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 (*3*, *13*), a direct evaluation of a broad range of social media platforms is needed to best leverage available social media platforms.

Although opioids contribute to morbidity and mortality in many countries, the United States and Canada have an unusually serious epidemic of addiction and overdose, which was triggered by opioid overprescription and is now mainly driven by fentanyl use (*14–19*). Therefore, while social media promises to advance understanding of opioid use worldwide, it is particularly urgent to leverage to address the North American opioid epidemic.

Here, we identified social media platforms that may be suitable for text-based opioid research and characterized their opioid-related discussions. We created a shortlist of eleven platforms for which we investigated censorship policies, data accessibility, and prior use in opioid research. We 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.

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 (**Supplementary Data S1**).

Creating platform shortlist

We applied our shortlisting criteria to the 71 platforms iteratively (**Figure 1**, **Table S2**). 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.

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**.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

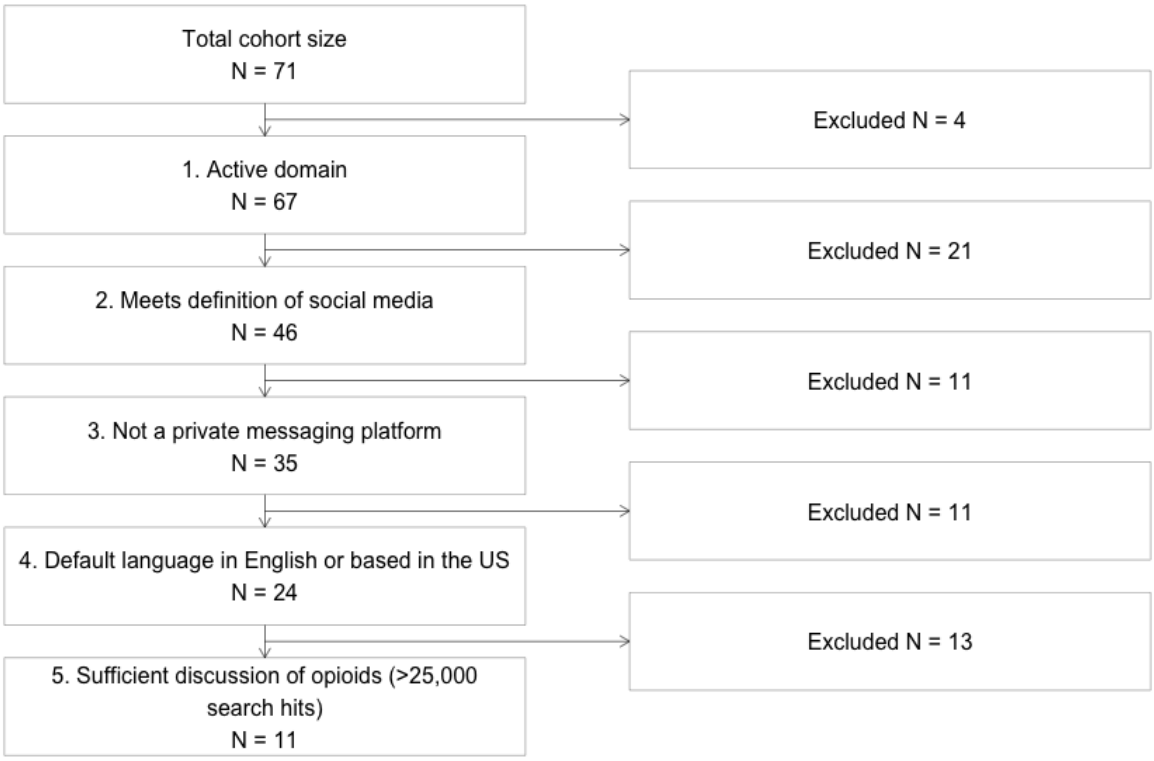

Fig. 1. Platform shortlisting consort diagram. Consort diagram of iterative exclusion criteria to attain shortlist of 11 social media platforms for further characterization.

Measuring the volume of opioid-related discussion

The type and volume of publicly available opioid-related text varied across platforms (Figure 2, Figure 3, Figure S1, Table S3, Table S4). We observed the highest total volume of opioid-related discussion on YouTube and Facebook (Figure 2a, Figure S1, Table S3). 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 (Figure 2b). The outliers reveal which platforms have more algospeak, indicating a response to censorship. TikTok by far had the most amount of algospeak. Instagram and X also skewed 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 (**Figure 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.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1. Summary of characteristics of 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 **Supplementary Text S7**). "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 **Supplementary Text S8**). "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	Posts, comme	facebook.com	✓	☐	✓	☐	☐	Consensus of

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

	and friend activity	nts, captions							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,	youtube.com	?		?		?	Location-

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

		comments, transcripts							specific hashtags

We visualized the relative amounts of informal and algospeak term hits for the 11 platforms (**Figure 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.

Content restrictions and censorship policies

All eleven shortlisted platforms state expectations regarding drug-related discussion (**Supplementary Text S7**). High-level classifications of content policies are shown in **Table 1**.

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 (*12*) 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 **Supplementary Text S8**.

An early-warning system for trends in the opioid epidemic requires geolocatability. The geolocation of social media content can be obtained directly from the platform or, if unavailable, inferred. 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 opt to do this (under 2% (*20*)). Many groups have leveraged geotagged tweets for opioid-related research (*21–27*) and other research areas (*28*, *29*). 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 (*20*, *30*, *31*).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

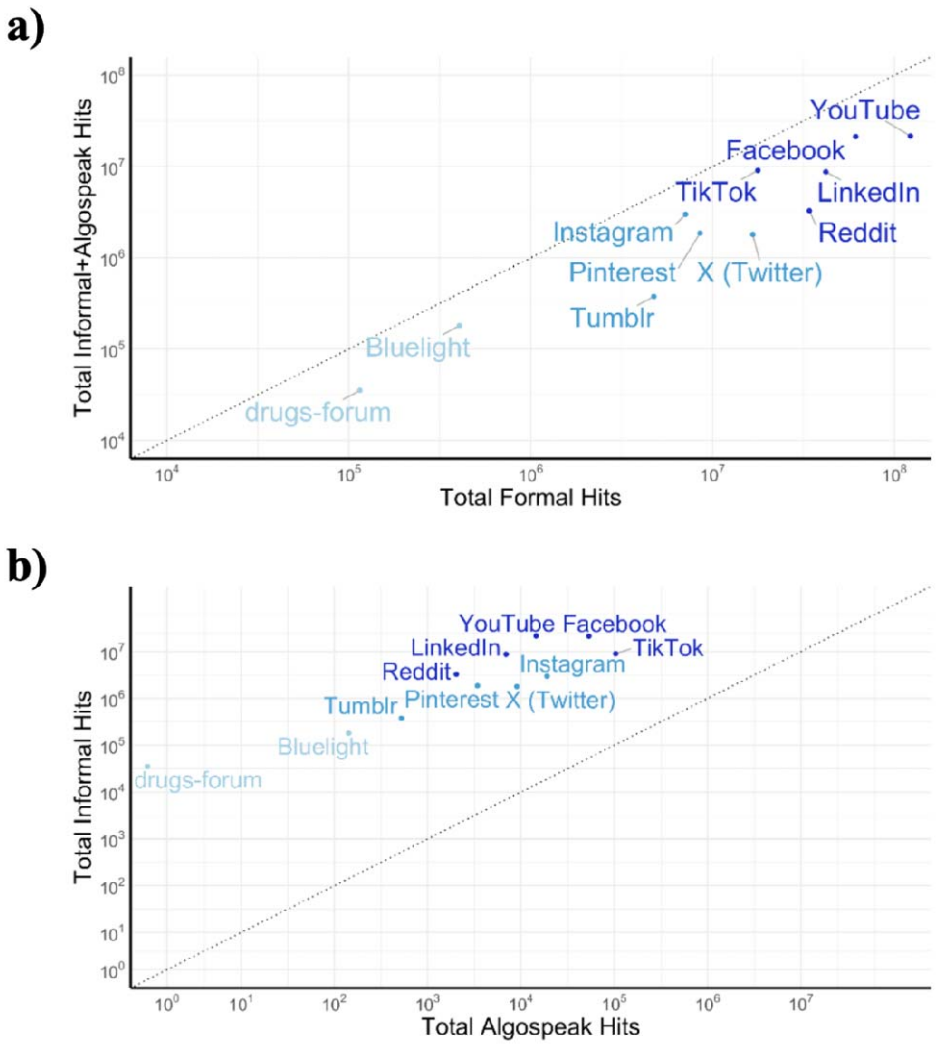

Fig. 2. Total term hits for all shortlisted platforms. Shown with a logarithmic scale. Platforms are clustered into three categories by volume of hits, shown by color (darker shade of blue indicates higher volume of hits). Dotted line shows y=x diagonal.

(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.

Similarly, Facebook (*32*), Instagram (*33–36*), and YouTube (*37–39*) give users the option to tag their content with geolocations. However, these tags can be used for purposes other than reporting the location from where the post was made (*40*). Facebook has a Data for Good program (*41*) which includes province-level GPS data from a subset of consenting users; previous work leveraged this data to study phenomena related to the COVID-19 pandemic (*42*, *43*). State *et al.* used provided user and company locations to track migration patterns using LinkedIn (*44*).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

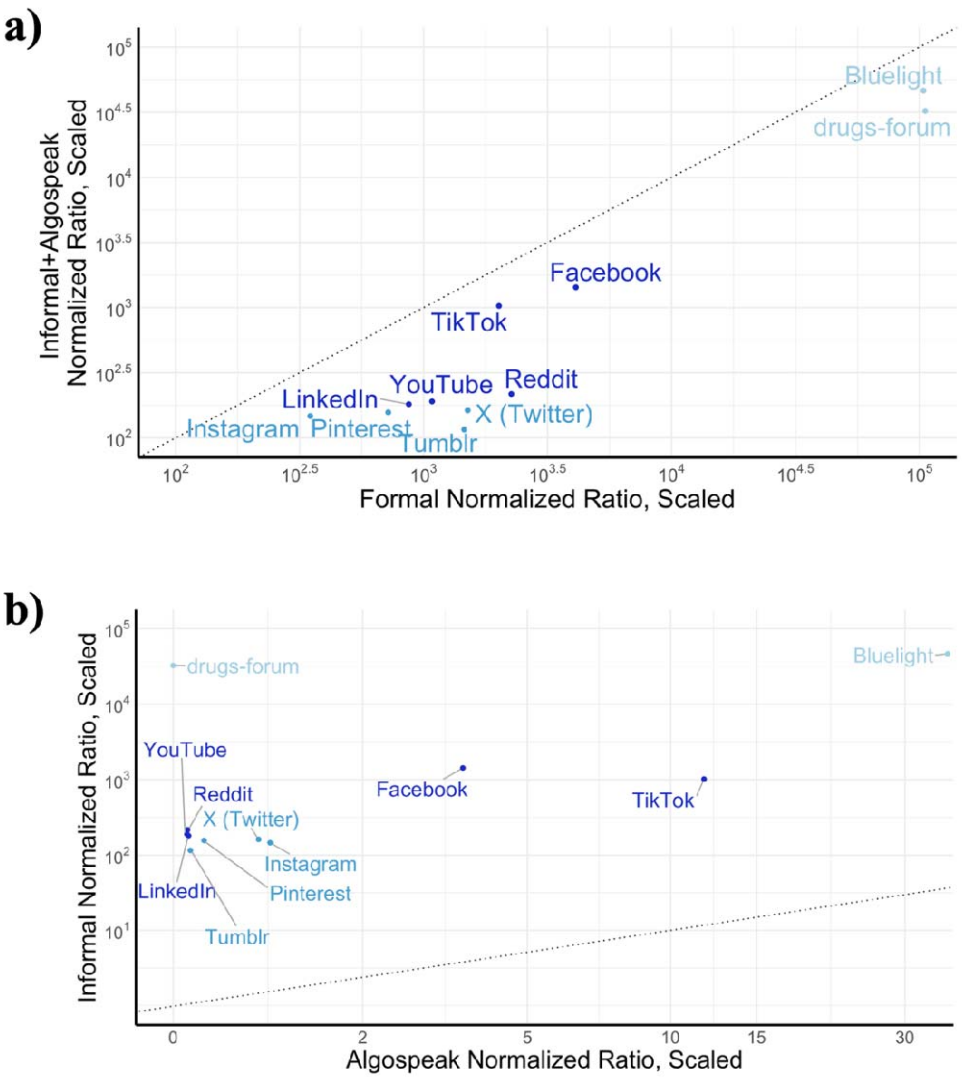

Fig. 3. Term hits normalized by number of hits for common nouns. Shown with a logarithmic scale and scaling factor of 100,000. As in Figure 2, platforms are clustered into three categories by total volume of hits, shown by color (darker shade of blue indicates higher volume of hits). Dotted line shows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

(A) Normalized hit ratios for formal opioid term list versus for informal opioid term list and algospeak opioid term list.

(B) Normalized hit ratios for informal opioid term list versus algospeak opioid term list.

The geolocatability of TikTok users has been prominently discussed in the US (*45–47*). Like other social media platforms, TikTok uses user location to personalize their content feed and allows users to add location tags to their videos (*48*). Various groups have characterized the privacy aspect of geolocation on TikTok (*49*, *50*), 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 (*51*).

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

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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- (*52*) or country-level (*53*).

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 (*54*). Several groups have leveraged the County Tweet Lexical Bank and datasets assembled by similar methods for opioid-related (*8*, *9*, *26*, *55*, *56*) and non-opioid-related (*57*, *58*) work. Others have used named entity recognition to extract unambiguous place names from social media text (*59–62*).

Several groups have used the hashtag search functionality on TikTok to obtain content relevant to geographic regions (*e.g.* specific countries) (*63*, *64*). Similarly, Hu and Conway pulled text from country-specific subreddits as a proxy for geolocation (*65*); Delbruel *et al.* examined the association between YouTube video tags and geolocation (*66*); 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 (*67*).

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 (*68–70*) 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) (*2*, *71*, *72*). Others have also searched for posts with the topic of "where are you from?" or instances of the phrase "I live in…" (*72*, *73*) and leveraged user "flairs" (tags that Reddit users can add to their username) (*72*) to geolocate Reddit users.

Numerous packages for predicting geolocation from social media data exist. Free packages used previously for geolocating social media text (*7*, *74*) include Carmen (*75*) and geopy (*76*); paid services used previously for geolocating social media text (*77–79*) include Iconosquare (*80*), Brandwatch (formerly Crimson Hexagon) (*81*), and Reputation (formerly Nuvi) (*82*). Additionally, many other groups have created geolocation inference methods (*32*, *83–86*). While these methods can facilitate geotagging, the accuracy of such methods decreases with time (*87*).

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 (*62*, *88*).

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

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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 (*2*, *7–10*, *25*, *26*, *55*, *56*), characterizing trends and themes in online discussion of opioids, opioid use, and OUD treatment (*4*, *21–24*, *55*, *59*, *72*, *89–118*), and characterizing public sentiment towards the opioid epidemic (*23*, *67*, *79*, *95*, *109*, *119*, *120*). Many groups have created models to identify posts on X and Reddit related to opioids (42,43,109,134–138), and one group created a pipeline to infer demographic information of the identified cohort (*126*). Others have used these platforms to characterize factors that influence opioid use, recovery, and the opioid epidemic generally (*6*, *8*, *127–131*), with particular interest shown to the impact of the COVID-19 pandemic (*90*, *107*, *132–140*) and co-use between opioids and other drugs (*141*, *142*).

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 (*143–145*), to characterize content related to drug sales (*146–149*), to identify emerging psychoactive substances (*108*), and to study public reactions to the opioid epidemic (*67*, *150*, *151*). 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 (*152*) and the opioid epidemic (*153*). YouTube and drugs-forum were both used to characterize misinformation related to OUD medications (*92*, *118*). Catalani *et al.* searched YouTube and Tumblr for content about emerging psychoactive substances (*108*). Others have used Tumblr text data to detect opioid sales (*146*, *154*, *155*). Paul *et al.* (*53*) and Lee *et al.* (*156*) used drugs-forum text data to establish correlations with NSDUH survey data. Drugs-forum has also been used to monitor other drugs (*157*, *158*).

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 (*159–166*). Vosburg *et al.* (*167*) and Soussan and Kjellgren (*168*) 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.

Discussion

Social media data signals correlate with trends in the opioid epidemic. 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 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 epidemic 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. While we highlight APIs for shortlisted platforms, not all drug-related discussion or user metadata are available through official APIs. Recently, some

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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 (*2*). Different downstream use cases will require different decisions on how to narrow the scope of opioid mentions for analysis.

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 (*169–171*).

Demographics play a key role in selecting social media platforms for surveillance of the opioid epidemic. 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 (*172*). Individuals who initiate opioid use at a younger age are more susceptible to substance use disorders (*173*); 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 (*174*). 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 (*174*). 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 (*175–178*) and in the context of substance use (*126*). 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 inferrable or explicitly-provided geolocation data. However, other platforms used less in research 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 (*174*), 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

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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 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 (*67*, *145*, *150*, *151*, *179*). 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.

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.

Data instability poses a challenge to surveilling opioid 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.

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 (*180*), 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, and harm to vulnerable populations. The ethical issues inherent to

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

analysis of social media data, and suggestions on how to conduct such research ethically, are addressed in (*181*).

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 (*182*). 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 key limitation of our analysis is its explicit focus on English text data on platforms primarily serving North America. Our choice to focus on North America was driven by the severity of the opioid 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 (*183*). 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 worldwide 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.

Materials and Methods

Experimental Design

The objective of this study was to identify social media platforms that have been underutilized in text-based research monitoring the opioid epidemic but contain opioid-related discussion and have characteristics favorable for large-scale research. The study consisted of three components: the broad platform identification stage, the platform shortlisting stage, and the platform characterization stage. The prespecified criteria and methods are detailed for each stage in their respective subsections below.

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-purpose 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

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

worldwide, as per the list compiled by Wikipedia (*184*). We also included all platforms that the Stanford Digital Economy Lab identified as "digital goods" due to their widespread use and relevance (*14*).

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 (*146*, *186–188*) and of general adverse drug reactions (*53*, *186*, *189–196*). 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 (*89*, *197–203*)

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 (*204*), 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 **Supplementary Text S1**.

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 (**Supplementary Text S1**, **Supplementary Text S2**).

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 (*205*) is able to generate slang for drugs of addiction at scale (*206*). We took validated GPT-3–generated slang terms for five opioids prominent in the current North American

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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 (**Supplementary Text S3**).

"Algospeak" is a phenomenon in which users of social media platforms purposefully alter words to evade banning and censorship (*169*). 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 (**Supplementary Text S4**, **S5**).

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 Corpus Of Contemporary American English to represent the most common nouns used in the English language (**Supplementary Text S6, S9-11**). 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.

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. We used this information to determine whether geolocation is provided for any accessible data.

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])

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

Statistical Analysis

Not applicable.

References

1. J. J. Palamar, Barriers to accurately assessing prescription opioid misuse on surveys. *Am. J. Drug Alcohol Abuse* **45**, 117–123 (2019).

2. A. Lavertu, T. Hamamsy, R. B. Altman, Monitoring the opioid epidemic via social media discussions. [Preprint] (2021). https://doi.org/10.1101/2021.04.01.21254815.

3. B. N. Rutherford, C. C. W. Lim, B. Johnson, B. Cheng, J. Chung, S. Huang, T. Sun, J. Leung, D. Stjepanović, G. C. K. Chan, #TurntTrending: a systematic review of substance use portrayals on social media platforms. *Addiction* **118**, 206–217 (2023).

4. C. L. Hanson, B. Cannon, S. Burton, C. Giraud-Carrier, An Exploration of Social Circles and Prescription Drug Abuse Through Twitter. *J. Med. Internet Res.* **15**, e189 (2013).

5. Y. Fan, Y. Zhang, Y. Ye, X. li, W. Zheng, "Social Media for Opioid Addiction Epidemiology: Automatic Detection of Opioid Addicts from Twitter and Case Studies" in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (Association for Computing Machinery, New York, NY, USA, 2017; https://dl.acm.org/doi/10.1145/3132847.3132857)*CIKM '17*, pp. 1259–1267.

6. Z. Yang, L. Nguyen, F. Jin, Predicting Opioid Relapse Using Social Media Data. arXiv arXiv:1811.12169 [Preprint] (2018). https://doi.org/10.48550/arXiv.1811.12169.

7. M. Chary, N. Genes, C. Giraud-Carrier, C. Hanson, L. S. Nelson, A. F. Manini, Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media. *J. Med. Toxicol.* **13**, 278–286 (2017).

8. S. Giorgi, D. B. Yaden, J. C. Eichstaedt, L. H. Ungar, H. A. Schwartz, A. Kwarteng, B. Curtis, Predicting U.S. county opioid poisoning mortality from multi-modal social media and psychological self-report data. *Sci. Rep.* **13**, 9027 (2023).

9. M. Matero, S. Giorgi, B. Curtis, L. H. Ungar, H. A. Schwartz, Opioid death projections with AI-based forecasts using social media language. *Npj Digit. Med.* **6**, 1–11 (2023).

10. S. A. Sumner, D. Bowen, K. Holland, M. L. Zwald, A. Vivolo-Kantor, G. P. Guy Jr, W. J. Heuett, D. P. Pressley, C. M. Jones, Estimating Weekly National Opioid Overdose Deaths in Near Real Time Using Multiple Proxy Data Sources. *JAMA Netw. Open* **5**, e2223033 (2022).

11. M. S. Jalali, M. Botticelli, R. C. Hwang, H. K. Koh, R. K. McHugh, The opioid crisis: a contextual, social-ecological framework. *Health Res. Policy Syst.* **18**, 87 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

12. J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, J. Blackburn, "The Pushshift Reddit Dataset" in *Proceedings of the International AAAI Conference on Web and Social Media* (2020; https://ojs.aaai.org/index.php/ICWSM/article/view/7347)vol. 14, pp. 830–839.

13. M. Conway, M. Hu, W. W. Chapman, Recent Advances in Using Natural Language Processing to Address Public Health Research Questions Using Social Media and ConsumerGenerated Data. *Yearb. Med. Inform.* **28**, 208–217 (2019).

14. The Lancet Regional Health – Americas, Opioid crisis: addiction, overprescription, and insufficient primary prevention. *Lancet Reg. Health - Am.* **23**, 100557 (2023).

15. H. M. A. Kaafarani, K. Han, M. El Moheb, N. Kongkaewpaisan, Z. Jia, M. W. El Hechi, S. Van Wijck, K. Breen, A. Eid, G. Rodriguez, M. Kongwibulwut, A. T. Nordestgaard, J. V. Sakran, H. Ezzeddine, B. Joseph, M. Hamidi, C. Ortega, S. L. Flores, B. J. Gutierrez-Sougarret, H. Qin, J. Yang, R. Gao, Z. Wang, Z. Gao, S. Prichayudh, S. Durmaz, G. Van Der Wilden, S. Santin, M. A. F. Ribeiro, N. Noppakunsomboom, R. Alami, L. El-Jamal, D. Naamani, G. Velmahos, K. D. Lillemoe, Opioids After Surgery in the United States Versus the Rest of the World: The International Patterns of Opioid Prescribing (iPOP) Multicenter Study. *Ann. Surg.* **272**, 879–886 (2020).

16. L. Manchikanti, A. Singh, Therapeutic Opioids: A Ten-Year Perspective on the Complexities and Complications of the Escalating Use, Abuse, and Nonmedial Use of Opioids. *Pain Physician* **11**, S63–S88.

17. K. Humphreys, C. L. Shover, C. M. Andrews, A. S. B. Bohnert, M. L. Brandeau, J. P. Caulkins, J. H. Chen, M.-F. Cuéllar, Y. L. Hurd, D. N. Juurlink, H. K. Koh, E. E. Krebs, A. Lembke, S. C. Mackey, L. Larrimore Ouellette, B. Suffoletto, C. Timko, Responding to the opioid crisis in North America and beyond: recommendations of the Stanford–Lancet Commission. *The Lancet* **399**, 555–604 (2022).

18. G. A. Kalkman, C. Kramers, W. Van Den Brink, A. F. A. Schellekens, The North American opioid crisis: a European perspective. *The Lancet* **400**, 1404 (2022).

19. R. Brown, A. Morgan, The opioid epidemic in North America: Implications for Australia. *Trends Issues Crime Crim. Justice* (2019).

20. Advanced filtering for geo data. https://developer.twitter.com/en/docs/tutorials/advanced-filtering-for-geo-data.

21. L. Flores, S. D. Young, Regional Variation in Discussion of Opioids on Social Media: A Qualitative Study. *J. Addict. Dis.* **39**, 316–321 (2021).

22. A. J. Calac, T. McMann, M. Cai, J. Li, R. Cuomo, T. K. Mackey, Exploring substance use disorder discussions in Native American communities: a retrospective Twitter infodemiology study. *Harm. Reduct. J.* **19**, 141 (2022).

23. S. Tibebu, V. C. Chang, C.-A. Drouin, W. Thompson, M. T. Do, At-a-glance - What can social media tell us about the opioid crisis in Canada? *Health Promot. Chronic Dis. Prev. Can. Res. Policy Pract.* **38**, 263–267 (2018).

24. Z. Tacheva, A. Ivanov, Exploring the Association Between the "Big Five" Personality Traits and Fatal Opioid Overdose: County-Level Empirical Analysis. *JMIR Ment. Health* **8**, e24939 (2021).

25. R. Cuomo, V. Purushothaman, A. J. Calac, T. McMann, Z. Li, T. Mackey, Estimating County-Level Overdose Rates Using Opioid-Related Twitter Data: Interdisciplinary Infodemiology Study. *JMIR Form. Res.* **7**, e42162 (2023).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

26. A. Sarker, G. Gonzalez-Hernandez, Y. Ruan, J. Perrone, Machine Learning and Natural Language Processing for Geolocation-Centric Monitoring and Characterization of Opioid-Related Social Media Chatter. *JAMA Netw. Open* **2**, e1914672 (2019).

27. S. J. Fodeh, M. Al-Garadi, O. Elsankary, J. Perrone, W. Becker, A. Sarker, Utilizing a multi-class classification approach to detect therapeutic and recreational misuse of opioids on Twitter. *Comput. Biol. Med.* **129**, 104132 (2021).

28. J. Yin, Y. Gao, G. Chi, An Evaluation of Geo-located Twitter Data for Measuring Human Migration. *Int. J. Geogr. Inf. Sci. IJGIS* **36**, 1830–1852 (2022).

29. O. Kellert, N. H. Matlis, Geolocation of multiple sociolinguistic markers in Buenos Aires. *PLoS ONE* **17**, e0274114 (2022).

30. J. Cao, H. H. Hochmair, F. Basheeh, The Effect of Twitter App Policy Changes on the Sharing of Spatial Information through Twitter Users. *Geographies* **2**, 549–562 (2022).

31. A. Kruspe, M. Häberle, E. J. Hoffmann, S. Rode-Hasinger, K. Abdulahhad, X. X. Zhu, Changes in Twitter geolocations: Insights and suggestions for future usage. arXiv arXiv:2108.12251 [Preprint] (2021). https://doi.org/10.48550/arXiv.2108.12251.

32. Y.-C. Lin, C.-M. Lai, J. W. Chapman, S. F. Wu, G. A. Barnett, "Geo-Location Identification of Facebook Pages" in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2018; https://ieeexplore.ieee.org/abstract/document/8508816), pp. 441–446.

33. R. B. Correia, L. Li, L. M. Rocha, Monitoring Potential Drug Interactions and Reactions via Network Analysis of Instagram User Timelines. *Pac. Symp. Biocomput.* **21**, 492–503 (2016).

34. J.-P. Allem, K.-H. Chu, T. B. Cruz, J. B. Unger, Waterpipe Promotion and Use on Instagram: #Hookah. *Nicotine Tob. Res.* **19**, 1248–1252 (2017).

35. B. T. van Zanten, D. B. Van Berkel, R. K. Meentemeyer, J. W. Smith, K. F. Tieskens, P. H. Verburg, Continental-scale quantification of landscape values using social media data. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12974–12979 (2016).

36. K. D. Mukhina, S. V. Rakitin, A. A. Visheratin, Detection of tourists attraction points using Instagram profiles. *Procedia Comput. Sci.* **108**, 2378–2382 (2017).

37. Y.-C. Song, Y.-D. Zhang, J. Cao, T. Xia, W. Liu, J.-T. Li, Web Video Geolocation by Geotagged Social Resources. *IEEE Trans. Multimed.* **14**, 456–470 (2012).

38. G. Friedland, O. Vinyals, T. Darrell, "Multimodal location estimation" in *Proceedings of the 18th ACM International Conference on Multimedia* (ACM, Firenze Italy, 2010; https://dl.acm.org/doi/10.1145/1873951.1874197), pp. 1245–1252.

39. K. Gavrić, D. Ćulibrk, M. Mirković, V. Crnojević, "Using YouTube data to analyze human continent-level mobility" in *2011 International Conference on Computational Aspects of Social Networks (CASoN)* (2011; https://ieeexplore.ieee.org/abstract/document/6085945), pp. 207–210.

40. J. Baik, The Geotagging Counterpublic: The Case of Facebook Remote Check-Ins to Standing Rock. *Int. J. Commun.* **14**, 21 (2020).

41. Data For Good Tools and Data. https://dataforgood.facebook.com/dfg/tools.

42. R. Pérez-Arnal, D. Conesa, S. Alvarez-Napagao, T. Suzumura, M. Català, E. Alvarez-Lacalle, D. Garcia-Gasulla, Comparative Analysis of Geolocation Information

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

through Mobile-Devices under Different COVID-19 Mobility Restriction Patterns in Spain. *ISPRS Int. J. Geo-Inf.* **10**, 73 (2021).

43. K. Zaremba, Opening of hotels and ski facilities: Impact on mobility, spending, and Covid-19 outcomes. *Health Econ.* **32**, 1148–1180 (2023).

44. B. State, M. Rodriguez, D. Helbing, E. Zagheni, "Migration of Professionals to the U.S." in *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014. Proceedings*, L. M. Aiello, D. McFarland, Eds. (Springer International Publishing, Cham, 2014; https://doi.org/10.1007/978-3-319-13734-6_37), pp. 531–543.

45. B. Fung, TikTok collects a lot of data. But that's not the main reason officials say it's a security risk | CNN Business, *CNN* (2023). https://www.cnn.com/2023/03/24/tech/tiktok-ban-national-security-hearing/index.html.

46. N. Logan, If you use TikTok, the app is collecting a staggering amount of information about you | CBC News, *CBC* (2023). https://www.cbc.ca/news/canada/tiktok-data-collection-privacy-1.6763626.

47. D. Shepardson, TikTok CEO: App has never shared US data with Chinese government, *Reuters* (2023). https://www.reuters.com/technology/tiktok-ceo-app-has-never-shared-us-data-with-chinese-goverment-2023-03-22/.

48. Location information on TikTok | TikTok Help Center. https://support.tiktok.com/en/account-and-privacy/account-privacy-settings/location-services-on-tiktok.

49. N. Ebert, T. Geppert, J. Strycharz, M. Knieps, M. Hönig, E. Brucker-Kley, Creative beyond TikToks: Investigating Adolescents' Social Privacy Management on TikTok. *Proc. Priv. Enhancing Technol.* **2023**, 221–235 (2023).

50. V. N. L. Franqueira, J. A. Annor, O. Kafali, Age Appropriate Design: Assessment of TikTok, Twitch, and YouTube Kids. arXiv arXiv:2208.02638 [Preprint] (2022). https://doi.org/10.48550/arXiv.2208.02638.

51. S. Zannettou, O.-N. Nemeth, O. Ayalon, A. Goetzen, K. P. Gummadi, E. M. Redmiles, F. Roesner, Leveraging Rights of Data Subjects for Social Media Analysis: Studying TikTok via Data Donations. arXiv arXiv:2301.04945 [Preprint] (2023). https://doi.org/10.48550/arXiv.2301.04945.

52. F. R. Lamy, R. Daniulaityte, R. W. Nahhas, M. J. Barratt, A. G. Smith, A. Sheth, S. S. Martins, E. W. Boyer, R. G. Carlson, Increased in synthetic cannabinoids-related harms: results from a longitudinal web-based content analysis. *Int. J. Drug Policy* **44**, 121–129 (2017).

53. M. J. Paul, M. S. Chisolm, M. W. Johnson, R. G. Vandrey, M. Dredze, Assessing the Validity of Online Drug Forums as a Source for Estimating Demographic and Temporal Trends in Drug Use. *J. Addict. Med.* **10**, 324 (2016).

54. H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. A. Dziurzynski, M. Agrawal, G. Park, S. K. Lakshmikanth, S. Jha, M. E. P. Seligman, L. Ungar, R. E. Lucas, "Characterizing Geographic Variation in Well-Being Using Tweets" in *Proceedings of the International AAAI Conference on Web and Social Media* (https://aaai.org/papers/00583-14442-characterizing-geographic-variation-in-well-being-using-tweets/)vol. 7, pp. 583–591.

55. R. L. Graves, C. Tufts, Z. F. Meisel, D. Polsky, L. Ungar, R. M. Merchant, Opioid Discussion in the Twittersphere. *Subst. Use Misuse* **53**, 2132–2139 (2018).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

56. M. Anwar, D. Khoury, A. P. Aldridge, S. J. Parker, K. P. Conway, Using Twitter to Surveil the Opioid Epidemic in North Carolina: An Exploratory Study. *JMIR Public Health Surveill.* **6**, e17574 (2020).

57. S. Giorgi, D. B. Yaden, J. C. Eichstaedt, R. D. Ashford, A. E. K. Buffone, H. A. Schwartz, L. H. Ungar, B. Curtis, Cultural Differences in Tweeting about Drinking Across the US. *Int. J. Environ. Res. Public. Health* **17**, 1125 (2020).

58. B. Curtis, S. Giorgi, A. E. K. Buffone, L. H. Ungar, R. D. Ashford, J. Hemmons, D. Summers, C. Hamilton, H. A. Schwartz, Can Twitter be used to predict county excessive alcohol consumption rates? *PLoS ONE* **13**, e0194290 (2018).

59. D. Jha, R. Singh, SMARTS: the social media-based addiction recovery and intervention targeting server. *Bioinforma. Oxf. Engl.*, btz800 (2019).

60. S. Milusheva, R. Marty, G. Bedoya, S. Williams, E. Resor, A. Legovini, Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning. *PLoS ONE* **16**, e0244317 (2021).

61. J. Sergeeva, A. Filatova, M. Kovalchuk, S. Teryoshkin, SemAGR: semantic method for accurate geolocations reconstruction within extensive urban sites. *Procedia Comput. Sci.* **212**, 409–417 (2022).

62. R. Compton, M. S. Keegan, J. Xu, Inferring the geographic focus of online documents from social media sharing patterns. arXiv arXiv:1406.2392 [Preprint] (2014). https://doi.org/10.48550/arXiv.1406.2392.

63. G. Pinto, K. Burghardt, K. Lerman, E. Ferrara, GET-Tok: A GenAI-Enriched Multimodal TikTok Dataset Documenting the 2022 Attempted Coup in Peru. arXiv arXiv:2402.05882 [Preprint] (2024). https://doi.org/10.48550/arXiv.2402.05882.

64. J. Bandy, N. Diakopoulos, #TulsaFlop: A Case Study of Algorithmically-Influenced Collective Action on TikTok. arXiv arXiv:2012.07716 [Preprint] (2020). https://doi.org/10.48550/arXiv.2012.07716.

65. M. Hu, M. Conway, 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* **2**, e36941 (2022).

66. S. Delbruel, D. Frey, F. Taïani, "Exploring the Geography of Tags in Youtube Views" (Technical Report 461, 2015); https://inria.hal.science/hal-01157867.

67. C. Goyer, G. Castillon, Y. Moride, 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* **19**, 5122 (2022).

68. G. Scalia, C. Francalanci, B. Pernici, CIME: Context-aware geolocation of emergency-related posts. *GeoInformatica* **26**, 125–157 (2022).

69. D. Jurgens, T. Finethy, J. McCorriston, Y. Xu, D. Ruths, "Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice" in *Proceedings of the International AAAI Conference on Web and Social Media* (2015; https://ojs.aaai.org/index.php/ICWSM/article/view/14627)vol. 9, pp. 188–197.

70. H. Pellet, S. Shiaeles, S. Stavrou, Localising social network users and profiling their movement. *Comput. Secur.* **81**, 49–57 (2019).

71. L. Bozarth, D. Quercia, L. Capra, S. Šćepanović, The role of the big geographic sort in online news circulation among U.S. Reddit users. *Sci. Rep.* **13**, 6711 (2023).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

72. D. Balsamo, P. Bajardi, A. Panisson, "Firsthand Opiates Abuse on Social Media: Monitoring Geospatial Patterns of Interest Through a Digital Cohort" in *The World Wide Web Conference* (2019; http://arxiv.org/abs/1904.00003), pp. 2572–2579.

73. K. Harrigian, "Geocoding Without Geotags: A Text-based Approach for reddit" in *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-Generated Text*, W. Xu, A. Ritter, T. Baldwin, A. Rahimi, Eds. (Association for Computational Linguistics, Brussels, Belgium, 2018; https://aclanthology.org/W18-6103), pp. 17–27.

74. H. Alhuzali, T. Zhang, S. Ananiadou, 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.* **24**, e40323 (2022).

75. M. Dredze, M. J. Paul, S. Bergsma, H. Tran, "Carmen: A twitter geolocation system with applications to public health" (2013)vol. WS-13-09, pp. 20–24.

76. geopy: Python Geocoding Toolbox, version 2.4.1; https://github.com/geopy/geopy.

77. J. A. Harris, N. A. Beck, C. J. Niedziela, G. A. Alvarez, S. A. Danquah, S. Afshar, The global reach of social media in oral and maxillofacial surgery. *Oral Maxillofac. Surg.*, 1–5 (2022).

78. S. A. Sumner, S. Galik, J. Mathieu, M. Ward, T. Kiley, B. Bartholow, A. Dingwall, P. Mork, Temporal and Geographic Patterns of Social Media Posts About an Emerging Suicide Game. *J. Adolesc. Health* **65**, 94–100 (2019).

79. E. M. Glowacki, J. B. Glowacki, G. B. Wilcox, A Text-Mining Analysis of the Public's Reactions to the Opioid Crisis. *Subst. Abuse* **39**, 129–133 (2018).

80. Iconosquare - Analytics and management for Instagram, TikTok, LinkedIn, Twitter & Facebook. https://www.iconosquare.com/.

81. Brandwatch | The social suite of the future, *Brandwatch*. https://www.brandwatch.com/.

82. Reputation | Online Reputation Management for Business, *Reputation*. https://reputation.com/.

83. R. Mahajan, V. Mansotra, Predicting Geolocation of Tweets: Using Combination of CNN and BiLSTM. *Data Sci. Eng.* **6**, 402–410 (2021).

84. P. Zola, P. Cortez, M. Carpita, Twitter user geolocation using web country noun searches. *Decis. Support Syst.* **120**, 50–59 (2019).

85. J. Bakerman, K. Pazdernik, A. Wilson, G. Fairchild, R. Bahran, Twitter Geolocation: A Hybrid Approach. *ACM Trans. Knowl. Discov. Data* **12**, 34:1-34:17 (2018).

86. M. Cha, Y. Gwon, H. Kung, "Twitter Geolocation and Regional Classification via Sparse Coding" in *Proceedings of the International AAAI Conference on Web and Social Media* (2015; https://ojs.aaai.org/index.php/ICWSM/article/view/14664)vol. 9, pp. 582–585.

87. M. Dredze, M. Osborne, P. Kambadur, "Geolocation for Twitter: Timing Matters" in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Compuational Linguistics* (2016), pp. 1064–1069.

88. J. Xu, R. Compton, T.-C. Lu, D. Allen, "Rolling through tumblr: characterizing behavioral patterns of the microblogging platform" in *Proceedings of the 2014 ACM*

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Conference on Web Science (Association for Computing Machinery, New York, NY, USA, 2014; https://dl.acm.org/doi/10.1145/2615569.2615694)*WebSci '14*, pp. 13–22.

89. B. Tofighi, A. Desai, C. Grov, J. Lee, Craigslist as a Source for Illicit Drugs: A Case Report and Review of the Literature. *Publ. Res.* (2016).

90. A. L. Nobles, D. C. Johnson, E. C. Leas, D. Goodman-Meza, M. L. Zúñiga, D. Ziedonis, S. A. Strathdee, J. W. Ayers, Characterizing Self-Reports of Self-Identified Patient Experiences with Methadone Maintenance Treatment on an Online Community during COVID-19. *Subst. Use Misuse* **56**, 2134–2140 (2021).

91. G.-G. P. Garcia, R. Dehghanpoor, E. J. Stringfellow, M. Gupta, J. Rochelle, E. Mason, T. A. Pujol, M. S. Jalali, Identifying and Characterizing Medical Advice-Seekers on a Social Media Forum for Buprenorphine Use. *Int. J. Environ. Res. Public. Health* **19**, 6281 (2022).

92. M. ElSherief, S. A. Sumner, C. M. Jones, R. K. Law, A. Kacha-Ochana, L. Shieber, L. Cordier, K. Holton, M. De Choudhury, Characterizing and Identifying the Prevalence of Web-Based Misinformation Relating to Medication for Opioid Use Disorder: Machine Learning Approach. *J. Med. Internet Res.* **23**, e30753 (2021).

93. A. Spadaro, A. Sarker, W. Hogg-Bremer, J. S. Love, N. O'Donnell, L. S. Nelson, J. Perrone, Reddit discussions about buprenorphine associated precipitated withdrawal in the era of fentanyl. *Clin. Toxicol. Phila. Pa* **60**, 694–701 (2022).

94. J. C. Black, Z. R. Margolin, R. A. Olson, R. C. Dart, Online Conversation Monitoring to Understand the Opioid Epidemic: Epidemiological Surveillance Study. *JMIR Public Health Surveill.* **6**, e17073 (2020).

95. B. Chan, A. Lopez, U. Sarkar, The Canary in the Coal Mine Tweets: Social Media Reveals Public Perceptions of Non-Medical Use of Opioids. *PloS One* **10**, e0135072 (2015).

96. M. A. Al-Garadi, Y.-C. Yang, Y. Guo, S. Kim, J. S. Love, J. Perrone, A. Sarker, Large-Scale Social Media Analysis Reveals Emotions Associated with Nonmedical Prescription Drug Use. *Health Data Sci.* **2022**, 9851989 (2022).

97. L. Shutler, L. S. Nelson, I. Portelli, C. Blachford, J. Perrone, Drug Use in the Twittersphere: A Qualitative Contextual Analysis of Tweets About Prescription Drugs. *J. Addict. Dis.* **34**, 303–310 (2015).

98. J. Kalyanam, T. Katsuki, G. R G Lanckriet, T. K. Mackey, Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. *Addict. Behav.* **65**, 289–295 (2017).

99. S. Raza, B. Schwartz, S. Lakamana, Y. Ge, A. Sarker, A framework for multi-faceted content analysis of social media chatter regarding non-medical use of prescription medications. *BMC Digit. Health* **1**, 29 (2023).

100. S. Yoon, M. Odlum, P. Broadwell, N. Davis, H. Cho, N. Deng, M. Patrao, D. Schauer, M. E. Bales, C. Alcantara, 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.* **272**, 5–8 (2020).

101. N. A. Haug, J. Bielenberg, S. H. Linder, A. Lembke, Assessment of provider attitudes toward #naloxone on Twitter. *Subst. Abuse* **37**, 35–41 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

102. M. Chenworth, J. Perrone, J. S. Love, R. Graves, W. Hogg-Bremer, A. Sarker, Methadone and suboxone® mentions on twitter: thematic and sentiment analysis. *Clin. Toxicol. Phila. Pa* **59**, 982–991 (2021).

103. A. R. D'Agostino, A. R. Optican, S. J. Sowles, M. J. Krauss, K. E. Lee, P. A. Cavazos-Rehg, Social networking online to recover from opioid use disorder: A study of community interactions. *Drug Alcohol Depend.* **181**, 5–10 (2017).

104. S. Pandrekar, X. Chen, G. Gopalkrishna, A. Srivastava, M. Saltz, J. Saltz, F. Wang, Social Media Based Analysis of Opioid Epidemic Using Reddit. *AMIA. Annu. Symp. Proc.* **2018**, 867–876 (2018).

105. K. E. Smith, J. M. Rogers, J. C. Strickland, D. H. Epstein, When an obscurity becomes trend: social-media descriptions of tianeptine use and associated atypical drug use. *Am. J. Drug Alcohol Abuse* **47**, 455–466 (2021).

106. E. Kasson, L. M. Filiatreau, N. Kaiser, K. Davet, J. Taylor, S. Garg, M. El Sherief, T. Aledavood, M. De Choudhury, P. Cavazos-Rehg, Using Social Media to Examine Themes Surrounding Fentanyl Misuse and Risk Indicators. *Subst. Use Misuse* **58**, 920–929 (2023).

107. A. M. Bunting, N. Krawczyk, T. Lippincott, Y. Gu, S. Arya, S. Nagappala, M. C. Meacham, Trends in Fentanyl Content on Reddit Substance Use Forums, 2013-2021. *J. Gen. Intern. Med.* **38**, 3283–3287 (2023).

108. V. Catalani, D. Arillotta, J. M. Corkery, A. Guirguis, A. Vento, F. Schifano, Identifying New/Emerging Psychoactive Substances at the Time of COVID-19; A Web-Based Approach. *Front. Psychiatry* **11**, 632405 (2020).

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

110. F. Carabot, O. Fraile-Martínez, C. Donat-Vargas, J. Santoma, C. Garcia-Montero, M. Pinto da Costa, R. M. Molina-Ruiz, M. A. Ortega, M. Alvarez-Mon, M. A. Alvarez-Mon, Understanding Public Perceptions and Discussions on Opioids Through Twitter: Cross-Sectional Infodemiology Study. *J Med Internet Res* **25**, e50013 (2023).

111. M. Chenworth, J. Perrone, J. S. Love, H. A. Greller, A. Sarker, P. R. Chai, Buprenorphine Initiation in the Emergency Department: a Thematic Content Analysis of a #firesidetox Tweetchat. *J. Med. Toxicol.* **16**, 262–268 (2020).

112. F. Carabot, C. Donat-Vargas, J. Santoma-Vilaclara, M. A. Ortega, C. García-Montero, O. Fraile-Martínez, C. Zaragoza, J. Monserrat, M. Alvarez-Mon, M. A. Alvarez-Mon, Exploring Perceptions About Paracetamol, Tramadol, and Codeine on Twitter Using Machine Learning: Quantitative and Qualitative Observational Study. *J. Med. Internet Res.* **25**, e45660 (2023).

113. A. Sarker, S. Lakamana, Y. Guo, Y. Ge, A. Leslie, O. Okunromade, E. Gonzalez-Polledo, J. Perrone, A. M. McKenzie-Brown, #ChronicPain: Automated Building of a Chronic Pain Cohort from Twitter Using Machine Learning. *Health Data Sci.* **3**, 0078 (2023).

114. A. Spadaro, K. O'Connor, S. Lakamana, A. Sarker, R. Wightman, J. S. Love, J. Perrone, Self-reported Xylazine Experiences: A Mixed-methods Study of Reddit Subscribers. *J. Addict. Med.* **17**, 691–694 (2023).

115. C. B. Whitman, M. W. Reid, Corey Arnold, H. Patel, L. Ursos, R. Sa'adon, J. Pourmorady, B. M. R. Spiegel, Balancing opioid-induced gastrointestinal side effects with

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

pain management: Insights from the online community. *J. Opioid Manag.* **11**, 383–391 (2015).

116. M. Cai, N. Shah, J. Li, W.-H. Chen, R. E. Cuomo, N. Obradovich, T. K. Mackey, Identification and characterization of tweets related to the 2015 Indiana HIV outbreak: A retrospective infoveillance study. *PLOS ONE* **15**, e0235150 (2020).

117. K. E. Smith, J. M. Rogers, D. Schriefer, O. Grundmann, Therapeutic benefit with caveats?: Analyzing social media data to understand the complexities of kratom use. *Drug Alcohol Depend.* **226**, 108879 (2021).

118. M. ElSherief, S. Sumner, V. Krishnasamy, C. Jones, R. Law, A. Kacha-Ochana, L. Schieber, M. De Choudhury, Identification of Myths and Misinformation about Treatment for Opioid Use Disorder: Infodemiology Study of Social Media (Preprint). *JMIR Form. Res.*, doi: 10.2196/44726 (2022).

119. K. Rajesh, G. Wilcox, D. Ring, M. Mackert, Reactions to the opioid epidemic: A text-mining analysis of tweets. *J. Addict. Dis.* **39**, 183–188 (2021).

120. S. Ramachandran, L. Brown, D. Ring, Tones and themes in Reddits posts discussing the opioid epidemic. *J. Addict. Dis.* **40**, 552–558 (2022).

121. S. Garg, J. Taylor, M. El Sherief, E. Kasson, T. Aledavood, R. Riordan, N. Kaiser, P. Cavazos-Rehg, M. De Choudhury, Detecting risk level in individuals misusing fentanyl utilizing posts from an online community on Reddit. *Internet Interv.* **26**, 100467 (2021).

122. S. Chancellor, G. Nitzburg, A. Hu, F. Zampieri, M. De Choudhury, "Discovering Alternative Treatments for Opioid Use Recovery Using Social Media" in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, New York, NY, USA, 2019; https://dl.acm.org/doi/10.1145/3290605.3300354)*CHI '19*, pp. 1–15.

123. A. Preiss, P. Baumgartner, M. J. Edlund, G. V. Bobashev, 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.* **6**, e33919 (2022).

124. H. Yao, S. Rashidian, X. Dong, H. Duanmu, R. N. Rosenthal, F. Wang, Detection of Suicidality Among Opioid Users on Reddit: Machine Learning-Based Approach. *J. Med. Internet Res.* **22**, e15293 (2020).

125. M. A. Al-Garadi, Y.-C. Yang, H. Cai, Y. Ruan, K. O'Connor, G.-H. Graciela, J. Perrone, A. Sarker, Text classification models for the automatic detection of nonmedical prescription medication use from social media. *BMC Med. Inform. Decis. Mak.* **21**, 27 (2021).

126. Y.-C. Yang, M. A. Al-Garadi, J. S. Love, H. L. F. Cooper, J. Perrone, A. Sarker, Can accurate demographic information about people who use prescription medications nonmedically be derived from Twitter? *Proc. Natl. Acad. Sci.* **120**, e2207391120 (2023).

127. D. Jha, R. Singh, 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* **21**, 554 (2020).

128. E. Naserianhanzaei, M. Koschate-Reis, 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.* **24**, e36555 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

129. W. Kepner, M. C. Meacham, A. L. Nobles, Types and Sources of Stigma on Opioid Use Treatment and Recovery Communities on Reddit. *Subst. Use Misuse* **57**, 1511–1522 (2022).

130. A. T. Chen, S. Johnny, M. Conway, Examining stigma relating to substance use and contextual factors in social media discussions. *Drug Alcohol Depend. Rep.* **3**, 100061 (2022).

131. W. Bremer, K. Plaisance, D. Walker, M. Bonn, J. S. Love, J. Perrone, A. Sarker, Barriers to opioid use disorder treatment: A comparison of self-reported information from social media with barriers found in literature. *Front. Public Health* **11**, 1141093 (2023).

132. A. Alambo, S. Padhee, T. Banerjee, K. Thirunarayan, COVID-19 and Mental Health/Substance Use Disorders on Reddit: A Longitudinal Study. arXiv arXiv:2011.10518v1 [Preprint] (2020). https://doi.org/10.48550/arXiv.2011.10518.

133. D. Frank, N. Krawczyk, J. Arshonsky, M. A. Bragg, S. R. Friedman, A. M. Bunting, COVID-19-Related Changes to Drug-Selling Networks and Their Effects on People Who Use Illicit Opioids. *J. Stud. Alcohol Drugs* **84**, 222–229 (2023).

134. J. Arshonsky, N. Krawczyk, A. M. Bunting, D. Frank, S. R. Friedman, M. A. Bragg, Informal Coping Strategies Among People Who Use Opioids During COVID-19: Thematic Analysis of Reddit Forums. *JMIR Form. Res.* **6**, e32871 (2022).

135. N. Krawczyk, A. M. Bunting, D. Frank, J. Arshonsky, Y. Gu, S. R. Friedman, M. A. Bragg, "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* **92**, 103140 (2021).

136. A. Sarker, N. Nataraj, W. Siu, S. Li, C. M. Jones, S. A. Sumner, 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* **17**, 16 (2022).

137. D. Arillotta, A. Guirguis, J. M. Corkery, N. Scherbaum, F. Schifano, COVID-19 Pandemic Impact on Substance Misuse: A Social Media Listening, Mixed Method Analysis. *Brain Sci.* **11**, 907 (2021).

138. N. El-Bassel, K. R. Hochstatter, M. N. Slavin, C. Yang, Y. Zhang, S. Muresan, 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.* **16**, e123–e132 (2022).

139. T. Jiang, V. Osadchiy, J. M. Weinberger, M. H. Zheng, M. H. Owen, S. A. Leonard, J. N. Mills, N. Kachroo, S. V. Eleswarapu, Impact of the COVID-19 Pandemic on Patient Preferences and Decision Making for Symptomatic Urolithiasis. *J Endourol* **35**, 1250–1256 (2021).

140. K. Wanchoo, M. Abrams, R. M. Merchant, L. Ungar, S. C. Guntuku, Reddit language indicates changes associated with diet, physical activity, substance use, and smoking during COVID-19. *PLOS ONE* **18**, e0280337 (2023).

141. A. Sarker, M. A. Al-Garadi, Y. Ge, N. Nataraj, C. M. Jones, S. A. Sumner, Signals of increasing co-use of stimulants and opioids from online drug forum data. *Harm. Reduct. J.* **19**, 51 (2022).

142. M. C. Meacham, A. L. Nobles, D. A. Tompkins, J. Thrul, "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* **17**, e0263583 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

143. V. Purushothaman, J. Li, T. K. Mackey, Detecting Suicide and Self-Harm Discussions Among Opioid Substance Users on Instagram Using Machine Learning. *Front. Psychiatry* **12**, 551296 (2021).

144. R. Cherian, M. Westbrook, D. Ramo, U. Sarkar, Representations of Codeine Misuse on Instagram: Content Analysis. *JMIR Public Health Surveill.* **4**, e22 (2018).

145. E. T. Smolev, L. Rolf, E. Zhu, S. K. Buday, M. Brody, D. M. Brogan, C. J. Dy, "Pill Pushers and CBD Oil"-A Thematic Analysis of Social Media Interactions About Pain After Traumatic Brachial Plexus Injury. *J. Hand Surg. Glob. Online* **3**, 36–40 (2021).

146. M. R. Haupt, R. Cuomo, J. Li, M. Nali, T. K. Mackey, The influence of social media affordances on drug dealer posting behavior across multiple social networking sites (SNS). *Comput. Hum. Behav. Rep.* **8**, 100235 (2022).

147. X. Yang, J. Luo, Tracking Illicit Drug Dealing and Abuse on Instagram Using Multimodal Analysis. *ACM Trans. Intell. Syst. Technol.* **8**, 58:1-58:15 (2017).

148. N. Shah, J. Li, T. K. Mackey, An unsupervised machine learning approach for the detection and characterization of illicit drug-dealing comments and interactions on Instagram. *Subst. Abuse* **43**, 273–277 (2022).

149. T. Zaprutko, D. Kopciuch, A. Paczkowska, J. Sprawka, J. Cynar, M. Pogodzińska, K. Niewczas, A. Stolecka, M. Sygit, M. Michalak, P. Ratajczak, K. Kus, Facebook as a source of access to medicines. *PloS One* **17**, e0275272 (2022).

150. D. C. Stokes, J. Purtle, Z. F. Meisel, A. K. Agarwal, State Legislators' Divergent Social Media Response to the Opioid Epidemic from 2014 to 2019: Longitudinal Topic Modeling Analysis. *J. Gen. Intern. Med.* **36**, 3373–3382 (2021).

151. D. K. Kilgo, J. Midberry, Social Media News Production, Emotional Facebook Reactions, and the Politicization of Drug Addiction. *Health Commun.* **37**, 375–383 (2022).

152. N. A. Squires, E. Soyemi, L. M. Yee, E. M. Birch, N. Badreldin, Content Quality of YouTube Videos About Pain Management After Cesarean Birth: Content Analysis. *JMIR Infodemiology* **3**, e40802 (2023).

153. M. Ittefaq, A. Zain, H. Bokhari, Opioids in Satirical News Shows: Exploring Topics, Sentiments, and Engagement in Last Week Tonight on YouTube. *J. Health Commun.* **28**, 53–63 (2023).

154. A. Al-Rawi, The convergence of social media and other communication technologies in the promotion of illicit and controlled drugs. *J. Public Health* **44**, e153–e160 (2022).

155. F. Zhao, P. Skums, A. Zelikovsky, E. L. Sevigny, M. H. Swahn, S. M. Strasser, Y. Huang, Y. Wu, Computational Approaches to Detect Illicit Drug Ads and Find Vendor Communities Within Social Media Platforms. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **19**, 180–191 (2022).

156. C. Lee, C. St Clair, C. C. Merenda, C. R. Araojo, S. Ray, D. Beasley, R. D. Hinton, Assessment of public and patient online comments in social media and food and drug administration archival data. *Res. Soc. Adm. Pharm.* **16**, 967–973 (2020).

157. R. Abouchedid, T. Gilks, P. I. Dargan, J. R. H. Archer, D. M. Wood, 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.* **14**, 134–143 (2018).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

158. D. Rhumorbarbe, M. Morelato, L. Staehli, C. Roux, D.-O. Jaquet-Chiffelle, Q. Rossy, P. Esseiva, Monitoring new psychoactive substances: Exploring the contribution of an online discussion forum. *Int. J. Drug Policy* **73**, 273–280 (2019).

159. J. W. Merten, B. T. Gordon, J. L. King, C. Pappas, Cannabidiol (CBD): Perspectives from Pinterest. *Subst. Use Misuse* **55**, 2213–2220 (2020).

160. L. I. Laestadius, J. P. D. Guidry, R. Greskoviak, J. Adams, Making "Weedish Fish": An Exploratory Analysis of Cannabis Recipes on Pinterest. *Subst. Use Misuse* **54**, 2191–2197 (2019).

161. J. Guidry, Y. Jin, L. Haddad, Y. Zhang, J. Smith, How Health Risks Are Pinpointed (or Not) on Social Media: The Portrayal of Waterpipe Smoking on Pinterest. *Health Commun.* **31**, 659–667 (2016).

162. A. S. Lee, J. L. Hart, C. G. Sears, K. L. Walker, A. Siu, C. Smith, A picture is worth a thousand words: Electronic cigarette content on Instagram and Pinterest. *Tob. Prev. Cessat.* **3**, 119 (2017).

163. T. Sun, C. C. W. Lim, J. Chung, B. Cheng, L. Davidson, C. Tisdale, J. Leung, C. E. Gartner, J. Connor, W. D. Hall, G. C. K. Chan, Vaping on TikTok: a systematic thematic analysis. *Tob. Control* **32**, 251–254 (2023).

164. B. N. Rutherford, T. Sun, B. Johnson, S. Co, T. L. Lim, C. C. W. Lim, V. Chiu, J. Leung, D. Stjepanovic, J. P. Connor, G. C. K. Chan, Getting high for likes: Exploring cannabis-related content on TikTok. *Drug Alcohol Rev.* **41**, 1119–1125 (2022).

165. J. Jancey, T. Leaver, K. Wolf, B. Freeman, K. Chai, S. Bialous, M. Bromberg, P. Adams, M. Mcleod, R. N. Carey, K. McCausland, Promotion of E-Cigarettes on TikTok and Regulatory Considerations. *Int. J. Environ. Res. Public. Health* **20**, 5761 (2023).

166. J. Whelan, G. E. Noller, R. D. Ward, Rolling through TikTok: An analysis of 3,4-methylenedioxymethamphetamine-related content. *Drug Alcohol Rev.* **43**, 36–44 (2024).

167. S. K. Vosburg, T. Dailey-Govoni, J. Beaumont, S. F. Butler, J. L. Green, Characterizing the Experience of Tapentadol Nonmedical Use: Mixed Methods Study. *JMIR Form. Res.* **6**, e16996 (2022).

168. C. Soussan, A. Kjellgren, The users of Novel Psychoactive Substances: Online survey about their characteristics, attitudes and motivations. *Int. J. Drug Policy* **32**, 77–84 (2016).

169. E. Steen, K. Yurechko, D. Klug, You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok. *Soc. Media Soc.* **9**, 20563051231194586 (2023).

170. V. Vera, Nonsuicidal Self-Injury and Content Moderation on TikTok. *Proc. Assoc. Inf. Sci. Technol.* **60**, 1164–1166 (2023).

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

172. U.S. TikTok users by age 2022, *Statista*. https://www.statista.com/statistics/1095186/tiktok-us-users-age/.

173. M. D. Stein, M. T. Conti, S. Kenney, B. J. Anderson, J. N. Flori, M. M. Risi, G. L. Bailey, Adverse childhood experience effects on opioid use initiation, injection drug use, and

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

overdose among persons with opioid use disorder. *Drug Alcohol Depend.* **179**, 325–329 (2017).

174. 1615 L. St NW, S. 800 Washington, D. 20036 U.-419-4300 | M.-857-8562 | F.-419-4372 | M. Inquiries, Social Media Fact Sheet, *Pew Research Center: Internet, Science & Tech.* https://www.pewresearch.org/internet/fact-sheet/social-media/.

175. A. A. Morgan-Lopez, A. E. Kim, R. F. Chew, P. Ruddle, Predicting age groups of Twitter users based on language and metadata features. *PLOS ONE* **12**, e0183537 (2017).

176. A. Pandya, M. Oussalah, P. Monachesi, P. Kostakos, On the use of distributed semantics of tweet metadata for user age prediction. *Future Gener. Comput. Syst.* **102**, 437–452 (2020).

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

178. D. Preo□iuc-Pietro, L. Ungar, "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), pp. 1534–1545.

179. J. J. van Hoof, J. Bekkers, M. van Vuuren, Son, you're smoking on Facebook! College students' disclosures on social networking sites as indicators of real-life risk behaviors. *Comput. Hum. Behav.* **34**, 249–257 (2014).

180. J. Taylor, C. Pagliari, Mining social media data: How are research sponsors and researchers addressing the ethical challenges? *Res. Ethics* **14**, 1–39 (2018).

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

182. J. Peters, Tweets aren't showing up in Google results as often because of changes at Twitter, *The Verge* (2023). https://www.theverge.com/2023/7/3/23783153/google-twitter-tweets-changes-rate-limits.

183. U. C. Bureau, Nearly 68 Million People Spoke a Language Other Than English at Home in 2019, *Census.gov*. https://www.census.gov/library/stories/2022/12/languages-we-speak-in-united-states.html.

184. List of social platforms with at least 100 million active users, *Wikipedia* (2024). https://en.wikipedia.org/w/index.php?title=List_of_social_platforms_with_at_least_100_million_active_users&oldid=1213672777.

185. E. Brynjolfsson, A. Collis, How Should We Measure the Digital Economy?, *Harvard Business Review* (2019). https://hbr.org/2019/11/how-should-we-measure-the-digital-economy.

186. D. M. Kazemi, B. Borsari, M. J. Levine, B. Dooley, Systematic review of surveillance by social media platforms for illicit drug use. *J. Public Health* **39**, 763–776 (2017).

187. L. Moyle, A. Childs, R. Coomber, M. J. Barratt, #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs. *Int. J. Drug Policy* **63**, 101–110 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

188. R. S. Wightman, J. Perrone, F. Erowid, E. Erowid, Z. F. Meisel, L. S. Nelson, Comparative Analysis of Opioid Queries on Erowid.org: An Opportunity to Advance Harm Reduction. *Subst. Use Misuse* **52**, 1315–1319 (2017).

189. A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, G. Gonzalez, Utilizing Social Media Data for Pharmacovigilance: A Review. *J. Biomed. Inform.* **54**, 202–212 (2015).

190. J.-Y. Lee, Y.-S. Lee, D. H. Kim, H. S. Lee, B. R. Yang, M. G. Kim, The Use of Social Media in Detecting Drug Safety–Related New Black Box Warnings, Labeling Changes, or Withdrawals: Scoping Review. *JMIR Public Health Surveill.* **7**, e30137 (2021).

191. A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn, G. Gonzalez, Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J. Am. Med. Inform. Assoc.* **22**, 671–681 (2015).

192. R. Sloane, O. Osanlou, D. Lewis, D. Bollegala, S. Maskell, M. Pirmohamed, Social media and pharmacovigilance: A review of the opportunities and challenges. *Br. J. Clin. Pharmacol*. **80**, 910–920 (2015).

193. R. B. Correia, I. B. Wood, J. Bollen, L. M. Rocha, Mining Social Media Data for Biomedical Signals and Health-Related Behavior. *Annu. Rev. Biomed. Data Sci.* **3**, 433–458 (2020).

194. C. C. Yang, H. Yang, L. Jiang, Postmarketing Drug Safety Surveillance Using Publicly Available Health-Consumer-Contributed Content in Social Media. *ACM Trans. Manag. Inf. Syst.* **5**, 2:1-2:21 (2014).

195. I. C.-H. Fung, E. B. Blankenship, J. O. Ahweyevu, L. K. Cooper, C. H. Duke, S. L. Carswell, A. M. Jackson, J. C. Jenkins, E. A. Duncan, H. Liang, K.-W. Fu, Z. T. H. Tse, Public Health Implications of Image-Based Social Media: A Systematic Review of Instagram, Pinterest, Tumblr, and Flickr. *Perm. J.* **24**, 18.307 (2020).

196. T. Nguyen, M. E. Larsen, B. O'Dea, D. Phung, S. Venkatesh, H. Christensen, Estimation of the prevalence of adverse drug reactions from social media. *Int. J. Med. Inf.* **102**, 130–137 (2017).

197. L. Laestadius, Y. Wang, Youth access to JUUL online: eBay sales of JUUL prior to and following FDA action. *Tob. Control* **28**, 617–622 (2019).

198. D. Blok, L. Ambrose, L. Ouellette, E. Seif, B. Riley, B. Judge, A. Ziegler, J. Jones, Selling poison by the bottle: Availability of dangerous substances found on eBay®. *Am. J. Emerg. Med.* **38**, 846–848 (2020).

199. R. S. Williams, Underage internet alcohol sales on eBay. *Addiction* **108**, 1346–1348 (2013).

200. I. Haber, J. Pergolizzi, J. A. LeQuang, Poppy Seed Tea: A Short Review and Case Study. *Pain Ther.* **8**, 151–155 (2019).

201. Drug Enforcement Administration, Social Media Drug Trafficking Threat (2022). https://www.dea.gov/sites/default/files/2022-03/20220208-DEA_Social%20Media%20Drug%20Trafficking%20Threat%20Overview.pdf.

202. E. Kim, Etsy blocks sales of drugs and human remains, *CNN Business* (2012). https://money.cnn.com/2012/08/10/technology/etsy-bans-drugs/index.html.

203. R. Price, Etsy is awash with illicit products it claims to ban, from ivory to dangerous weapons and mass-produced goods, *Business Insider* (2021).

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

https://www.businessinsider.com/etsy-sells-ivory-weapons-poisonous-plants-mass-produced-products-2021-4.

204. C. Rajendra-Nicolucci, E. Zuckerman, Top 100: The most popular social media platforms and what they can teach us (2021). https://knightcolumbia.org/blog/top-100-the-most-popular-social-media-platforms-and-what-they-can-teach-us.

205. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, "Language Models are Few-Shot Learners" in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2020; https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html)vol. 33, pp. 1877–1901.

206. K. A. Carpenter, R. B. Altman, Using GPT-3 to Build a Lexicon of Drugs of Abuse Synonyms for Social Media Pharmacovigilance. *Biomolecules* **13**, 387 (2023).

207. D. Thanki, B. J. Frederick, "Social media and drug markets" (Publications Office of the European Union, Luxembourg, 2016; http://www.emcdda.europa.eu/system/files/publications/2155/TDXD16001ENN_FINAL.pdf), pp. 115–123.

208. X. Chen, C. Faviez, S. Schuck, A. Lillo-Le-Louët, N. Texier, B. Dahamna, C. Huot, P. Foulquié, S. Pereira, V. Leroux, P. Karapetiantz, A. Guenegou-Arnoux, S. Katsahian, C. Bousquet, A. Burgun, Mining Patients' Narratives in Social Media for Pharmacovigilance: Adverse Effects and Misuse of Methylphenidate. *Front. Pharmacol.* **9**, 541 (2018).

209. A. Magge, E. Tutubalina, Z. Miftahutdinov, I. Alimova, A. Dirkson, S. Verberne, D. Weissenbacher, G. Gonzalez-Hernandez, DeepADEMiner: a deep learning pharmacovigilance pipeline for extraction and normalization of adverse drug event mentions on Twitter. *J. Am. Med. Inform. Assoc.* **28**, 2184–2192 (2021).

210. J. C. Eichstaedt, H. A. Schwartz, M. L. Kern, G. Park, D. R. Labarthe, R. M. Merchant, S. Jha, M. Agrawal, L. A. Dziurzynski, M. Sap, C. Weeg, E. E. Larson, L. H. Ungar, M. E. P. Seligman, Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. *Psychol. Sci.* **26**, 159–169 (2015).

Acknowledgments

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

Funding:

National Institutes of Health grant R21DA057598
Microsoft Accelerating Foundation Models Research Initiative
National Science Foundation grant DGE-1656518 (KAC)
National Institutes of Health grant T32HL151323 (ATN)
Stanford Biochemistry Department (DAS)
National Science Foundation grant 2019286895 (DAS)
Stanford Summer First Fellowship (IAS)
Sarafan ChEM-H CBI Program Award (IAS)
Department of Veterans Affairs Health System Research Service Award RCS 04-141-3 (KH)

medRxiv preprint doi: https://doi.org/10.1101/2024.07.06.24310035; this version posted October 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Stanford Institute for Human-Centered AI (JCE, RBA, AL)
Chan Zuckerberg Biohub (RBA)

Author contributions:

Conceptualization: RBA, JCE
Methodology: KAC, ATN, DAS, JCE, RBA
Investigation: KAC, ATN, DAS, IAS
Visualization: KAC, ATN, IAS, JCE
Supervision: JCE, RBA, KH, AL, MVK
Writing—original draft: KAC, ATN, DAS
Writing—review & editing: KAC, ATN, IAS, KH, AL, MVK, JCE, RBA

Competing interests:

Authors declare that they have no competing interests.

Data and materials availability:

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.