

Original Paper

# Patterns of Routes of Administration and Drug Tampering for Nonmedical Opioid Consumption: Data Mining and Content Analysis of Reddit Discussions

Duilio Balsamo<sup>1</sup>, MS; Paolo Bajardi<sup>2</sup>, PhD; Alberto Salomone<sup>3</sup>, PhD; Rossano Schifanella<sup>2,4</sup>, PhD

<sup>1</sup>Department of Mathematics, University of Turin, Turin, Italy

<sup>2</sup>ISI Foundation, Turin, Italy

<sup>3</sup>Department of Chemistry, University of Turin, Turin, Italy

<sup>4</sup>Department of Computer Science, University of Turin, Turin, Italy

**Corresponding Author:**

Duilio Balsamo, MS

Department of Mathematics

University of Turin

Via Carlo Alberto, 10

Turin, 10124

Italy

Phone: 39 3331371972

Email: [duilio.balsamo@unito.it](mailto:duilio.balsamo@unito.it)

## Abstract

**Background:** The complex unfolding of the US opioid epidemic in the last 20 years has been the subject of a large body of medical and pharmacological research, and it has sparked a multidisciplinary discussion on how to implement interventions and policies to effectively control its impact on public health.

**Objective:** This study leverages Reddit, a social media platform, as the primary data source to investigate the opioid crisis. We aimed to find a large cohort of Reddit users interested in discussing the use of opioids, trace the temporal evolution of their interest, and extensively characterize patterns of the nonmedical consumption of opioids, with a focus on routes of administration and drug tampering.

**Methods:** We used a semiautomatic information retrieval algorithm to identify subreddits discussing nonmedical opioid consumption and developed a methodology based on word embedding to find alternative colloquial and nonmedical terms referring to opioid substances, routes of administration, and drug-tampering methods. We modeled the preferences of adoption of substances and routes of administration, estimating their prevalence and temporal unfolding. Ultimately, through the evaluation of odds ratios based on co-mentions, we measured the strength of association between opioid substances, routes of administration, and drug tampering.

**Results:** We identified 32 subreddits discussing nonmedical opioid usage from 2014 to 2018 and observed the evolution of interest among over 86,000 Reddit users potentially involved in firsthand opioid usage. We learned the language model of opioid consumption and provided alternative vocabularies for opioid substances, routes of administration, and drug tampering. A data-driven taxonomy of nonmedical routes of administration was proposed. We modeled the temporal evolution of interest in opioid consumption by ranking the popularity of the adoption of opioid substances and routes of administration, observing relevant trends, such as the surge in synthetic opioids like fentanyl and an increasing interest in rectal administration. In addition, we measured the strength of association between drug tampering, routes of administration, and substance consumption, finding evidence of understudied abusive behaviors, like chewing fentanyl patches and dissolving buprenorphine sublingually.

**Conclusions:** This work investigated some important consumption-related aspects of the opioid epidemic using Reddit data. We believe that our approach may provide a novel perspective for a more comprehensive understanding of nonmedical abuse of opioids substances and inform the prevention, treatment, and control of the public health effects.

(*J Med Internet Res* 2021;23(1):e21212) doi: [10.2196/21212](https://doi.org/10.2196/21212)

## KEYWORDS

routes of administration; drug tampering; Reddit; word embedding; social media; opioid; heroin; buprenorphine; oxycodone; fentanyl

## Introduction

### Background

In the last decade, the United States witnessed an unprecedented growth of deaths due to opioid drugs [1], which sparked from overprescriptions of semisynthetic opioid pain medication such as oxycodone and hydromorphone and evolved in a surge of abuse of illicit opioids like heroin [2,3] and powerful synthetic opioids like fentanyl [4,5]. Alongside traditional medical, pharmacological, and public health studies on the nonmedical adoption of prescription opioids [6-14], several phenomena related to the opioid epidemic have recently been successfully tackled through a digital epidemiology [15-18] approach. Researchers have used digital and social media data to perform various tasks, including detecting drug abuse [19,20], forecasting opioid overdose [21], studying transition into drug addiction [22], predicting opioid relapse [23], and discovering previously unknown treatments for opioid addiction [24]. A few recent studies investigated the temporal unfolding of the opioid epidemic in the United States by leveraging complementary data sources different from the official US Centers for Disease Control and Prevention data [2,25-28] and using social media like Reddit [29,30].

Pharmacology research is interested in understanding the consequences of various routes of administration (ROA), that is, the paths by which a substance is taken into the body [6,31,32], due to the different effects and potential health-related risks tied to them [10,33,34]. Researchers have estimated the prevalence of routes of administration for nonmedical prescription opioids [9,31,32,35] and opiates [36,37]; however, they rarely consider less common ROA, such as rectal, transdermal, or subcutaneous administration [32,38], leaving the mapping of nonmedical and nonconventional administration behaviors greatly unexplored [39,40]. Many of these studies [31,32,35] acknowledge that drug tampering, that is, the intentional chemical or physical alteration of medications [41], is an important constituent of drug abuse. The alteration of the pharmacokinetics of opioids through drug-tampering methods, together with unconventional administration, may potentially lead to very different addictive patterns and ultimately have unexpected health-associated risks [33]. Research has also been focused on developing tamper-resistant and abuse-deterrent drug formulations. However, to the best of our knowledge, no large-scale empirical evidence has been found to unveil the relationships between substance manipulation, unconventional ROA, and nonmedical substance administration.

### Goals

This paper seeks to complement current studies widening the understanding of opioid consumption patterns by using Reddit, a social content aggregation website, as the primary data source. This platform is structured into subreddits, user-generated and user-moderated communities dedicated to the discussion of specific topics (Multimedia Appendix 1). Due to fair guarantees

of anonymity, no limits on the number of characters in a post, and a large variety of debated topics, this platform is often used to uninhibitedly discuss personal experiences [42]. Reddit constitutes a nonintrusive and privileged data source to study a variety of issues [43,44], including sensitive topics such as mental health [45], weight loss [46], gender issues [47], and substance abuse [22,24].

This study's contributions are manifold. First, leveraging and expanding a recent methodology proposed by Balsamo et al [30], we identified a large cohort of opioid firsthand users (ie, Reddit users showing explicit interest in firsthand opioid consumption) and characterized their habits of substance use, administration, and drug tampering over a period of 5 years. Second, using word embeddings, we identified and cataloged a large set of terms describing practices of nonmedical opioids consumption. These terms are invaluable to performing exhaustive and at-scale analyses of user-generated content from social media, as they include colloquialisms, slang, and nonmedical terminology that is established on digital platforms and hardly used in the medical literature. We provided a longitudinal perspective on online interest in the opioids discourse and a quantitative characterization of the adoption of different ROA, with a focus on the less-studied yet emerging and relevant practices. We have made available the ROA taxonomy and the corresponding vocabulary to the research community. Third, we quantified the strength of association between ROA and drug-tampering methods to better characterize emerging practices. Finally, we investigated the interplay between the previous 3 dimensions, measuring odds ratios to shed light on the "how" and "what" facets of the opioid consumption phenomenon. We studied a wide spectrum of opioid forms, referred to as "opioids" throughout, ranging from prescription opioids to opiates and illegal opioid formulations. To the best of our knowledge, our contributions are original in both breadth and depth, outlining a detailed picture of nonmedical practices and abusive behaviors of opioid consumption through the lenses of digital data.

## Methods

### Data

We refer to a publicly available Reddit data set [48] that contains all the subreddits published on the platform since 2007 [44,49]. In this work, we analyzed the textual part of the submissions and the comments collected from 2014 to 2018. We preprocessed each year separately, filtering out the subreddits with less than 100 comments in a year. We used spaCy [50] to remove English stop words, inflectional endings, and tokens with less than 100 yearly appearances. We adopted a bag-of-words model, resulting in a vocabulary of different lemmas for each year. Vocabulary sizes ranged from 300,000 to 700,000 lemmas, with a size growth of approximately 30% each year. In Table 1, the number of unique comments and unique active users per year is reported. A steady growth of

approximately 30% per year both in the volume of conversations and in the active user base is observed.

All the analyses in this work were performed on a subset of subreddits related to opioid consumption, which were identified using the procedure described here. For space constraints, we restricted the analyses of odds ratios to comments and submissions created during 2018. Similar to a vast body of

users' activities on social media platforms [51-53], the distribution of posts per user shows a heavy tail, with the majority of users posting few comments and the remaining minority (eg, core users and subreddit moderators) producing a large portion of the content. Moreover, a nonnegligible percentage of posts, respectively 25% and 7% of submissions and comments, were produced by authors who deleted their usernames.

**Table 1.** Data set statistics.

Year	Reddit comments, n	Reddit authors, n	Opiates subreddits, n	Opiates comments, n	Opiates authors, n	Authors' prevalence
2014	545,720,071	8,149,234	19	386,984	12,381	0.0015
2015	699,245,245	10,673,990	19	470,609	15,888	0.0015
2016	840,575,089	12,849,603	25	612,619	21,791	0.0017
2017	1,045,425,499	14,219,062	30	866,023	28,358	0.0020
2018	1,307,123,219	18,158,464	25	919,036	33,700	0.0019

## Analytical Pipeline

The methodology adopted in this paper consists of several steps. First, we identified a cohort of opioid firsthand users and the subreddits related to opioid consumption through a semiautomatic algorithm. Second, we trained a word-embedding language model to capture the latent semantic features of the discourse on the nonmedical use of opioids. Third, we exploited the embedded vectors to extend an initial set of medical terms known from the literature, (eg, opioid substance names, ROA, and drug-tampering methods) to nonmedical and colloquial expressions. The terms were organized in a taxonomy that provides a conceptual map on the topic. Moreover, we studied the temporal evolution of the popularity of the main opioid substances and ROA. Ultimately, we measured the strength of the associations between opioid substances, routes of administration, and drug-tampering techniques in 2018.

### Identification of Firsthand Opioid Consumption on Reddit

We leveraged a semiautomatic information retrieval algorithm developed to identify relevant content related to a topic of interest [30] to collect opioid-related conversations on Reddit yearly. This approach aims at retrieving topic-specific documents by expressing a set of initial keywords of interest; here, it identified relevant subspaces of discussion via an iterative query expansion process, retaining a list of terms  $Q_y$  and a list of subreddits  $S_y$  ranked by relevance for each year.

We merged all the query terms in a set  $\bar{Q} = U_y Q_y$  containing 67 terms. To ensure that the sets  $S_y$  were effectively referring to the opioid-related topics and in particular to nonmedical opioid consumption, we performed a manual inspection on the union  $\bar{S} = U_y S_y$  of the top 150 subreddits for each year, for a total of 554 subreddits. Three independent annotators, including a domain expert specialized in antidoping analyses, read a random sample of 30 posts, checking for subreddits (1) mostly focused on discussing the use of opioids, (2) mostly focused on firsthand usage, and (3) not focused on medical treatments. This yielded a total of 32 selected subreddits, with a Fleiss  $\kappa$  interrater

agreement of  $\kappa=0.731$ , which suggests a substantial agreement, according to Landis and Koch [54]. [Multimedia Appendix 2](#) presents a complete list of the subreddits broken down by year.

Automatic language detection, performed with langdetect [55], cld2 [56], and cld3 [57], showed that the majority of posts (about 90%) were in English, approximately 5% were non-English messages, and the rest were too short or full of jargon and emojis to algorithmically detect any language. Assuming that an author who writes in one of the selected subreddits is personally interested in the topic, we identified a cohort of 86,445 unique opioid firsthand users involved in direct discussions of opioid usage across the period of study. Summary statistics are reported in [Table 1](#). In particular, for each year, we computed the number of unique active users and the volume of comments shared, as well as the user's relative prevalence over the entire amount of Reddit activity. We observed growth from 2014 to 2017, ranging from 15 to 19 users interested in opioid consumption out of every 100,000 Reddit users.

### Vocabulary Expansion

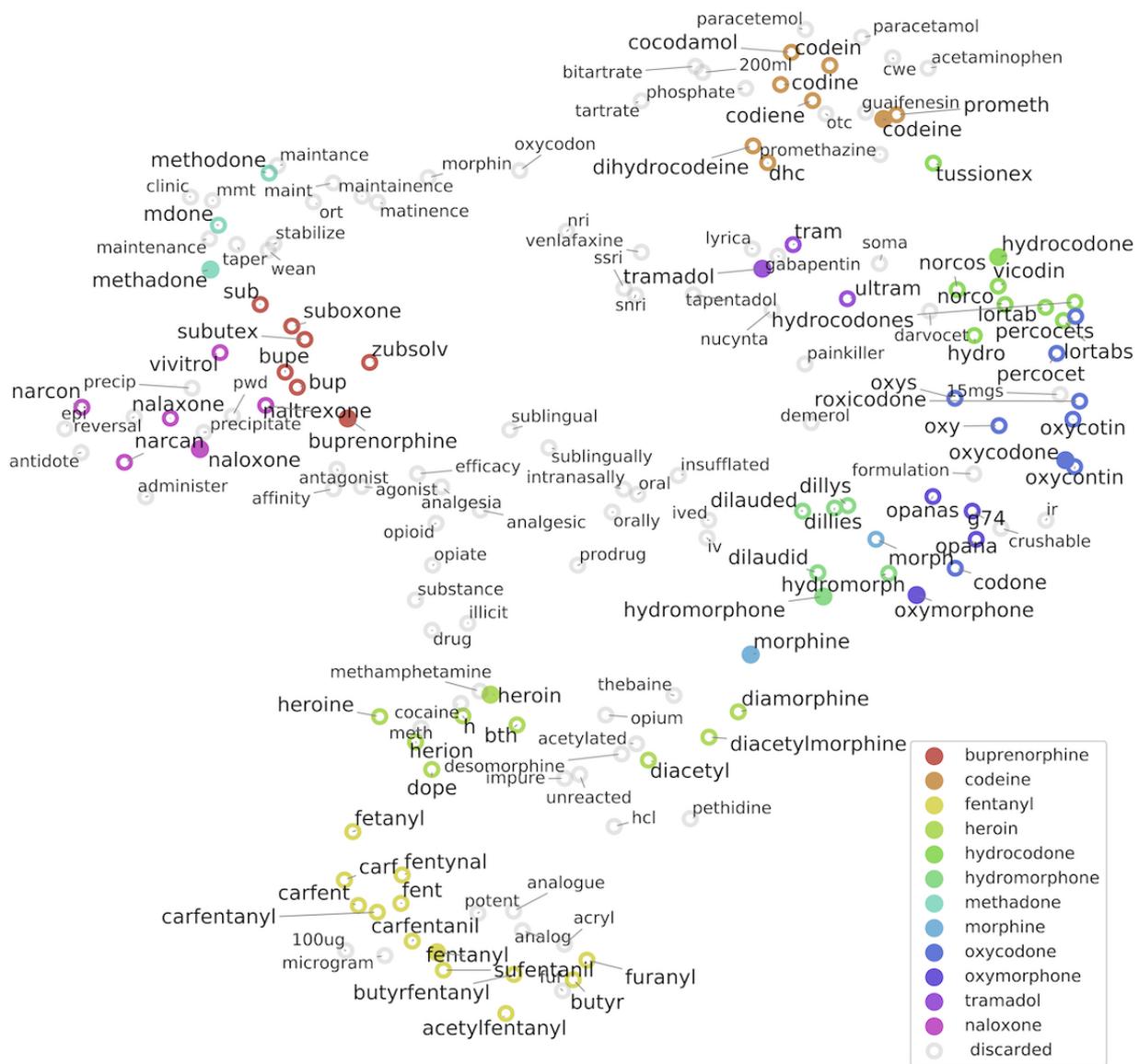
The methodology to extend the vocabulary on opioid-related domains with user-generated slang and colloquial forms was implemented in 2 steps. First, we trained a word-embedding model (word2vec [58]), which learns semantic relationships in the corpus during training and maps their terms to vectors in a latent vectorial space, with all the comments and submissions in our subreddit data set (relevant training parameters are displayed in [Multimedia Appendix 3](#)). Second, starting from a set of seed terms  $K$  (eg, a list of known opioid substances), we expanded the vocabulary by navigating the semantic neighborhood  $E_w^n = neighbors(w, n)$  of each element  $w \in K$  in the embedded space, considering the  $n=20$  semantically closest elements in terms of cosine similarity. We merged the results in a candidate expansion set,  $E = \bigcup_w E_w^n$ , together with the seed terms  $K$  if not already included. Based on the knowledge of a domain expert (a clinical and forensic toxicologist) and with the help of search engine queries and a crowdsourced online dictionary for slang words and phrases (Urban Dictionary [59]) to understand the most unusual terms, we manually selected

and categorized the relevant neighboring terms, obtaining an extended vocabulary  $V$ . Figure 1 shows an example of the expansion procedure in which the high-dimensional vectors are projected to 2 dimensions using the uniform manifold approximation and projection (UMAP) algorithm [60].

As a sensitivity analysis, we compared the effectiveness of an alternative embedding model (GloVe [61]) for topical coherence. In the case of vocabulary expansion of opioid substance terms,

that is, using  $K = \bar{Q}$  as seeds, the 2 models captured 100 terms in common out of their respective candidate terms, with word2vec showing a higher number and a larger percentage of accepted terms (Table 2). Moreover, the volume of comments that included an accepted term was almost double when using the vocabulary of word2vec rather than the vocabulary of GloVe. Hence, we chose word2vec as the reference word-embedding model.

**Figure 1.** Two-dimensional projection of the word2vec embedding, modeling the semantic relationships among terms in the Reddit opioids data set. Filled markers represent the seed terms  $K$ . Expansion terms, represented with hollow markers, are colored according to their respective initial term if accepted or in gray if discarded. The nature of the relationships between neighboring terms varies, representing (1) equivalence (eg, synonyms), (2) common practices (eg, the use of methadone for addiction maintenance), or (3) co-use (eg, the cluster of heroin, cocaine, and methamphetamine).



**Table 2.** Comparison of term expansions of opioid substances for the 2 trained models.

Model	Candidate terms, n	Accepted terms, n (%)	Comments <sup>a</sup> , n
word2vec	297	128 (43.1)	225,165
GloVe	369	110 (29.8)	144,564

<sup>a</sup>Comments in the corpus mentioning at least one term of the respective accepted terms for vocabulary expansion.

### ***Strength of Association Between Opioid Substances, ROA, and Drug Tampering***

We evaluated the odds ratios (ORs) to quantify the pairwise strength of the association between substance use and ROA, substance use and drug-tampering methods, and ROA and drug-tampering methods. Under the assumption that co-mention was a proxy for associating a substance to its ROA (or drug-tampering method), we focused on the posts that contained a reference to terms in each domain, evaluating contingency tables and odds ratios. Odds ratios, significance, and confidence intervals were estimated using chi-square tests implemented in the statsmodel Python package [62], with the significance level set to  $\alpha=.01$ . As a sensitivity analysis, we assessed the effect of the proximity of terms on the characterization of odds ratios. We modified the definition of co-occurrence, introducing a distance threshold  $\rho$  at sentence level. We explored the range  $\rho \in \{0..5\}$ , where  $\rho=0$  indicates that co-occurrence appears within the same sentence, and  $\rho>0$  measures the distance in both directions (eg,  $\rho=1$  for the preceding and consecutive sentences). The value  $\rho=\infty$  indicates the scenario in which we considered the entire post as reference. Accordingly, given a threshold  $\rho$  in the construction of the contingency table, the co-occurrence event between 2 terms is conditioned to their distance being less than or equal to  $\rho$ . Conversely, we considered terms to be separate events in cases of distance above the threshold. It is important to consider that the OR measures do not imply any form of causation but rather surface correlations that could be used in hypothesis formation. To better interpret the results of this analysis, in some cases, manual inspection of the comments mentioning the variables under investigation was performed following the directives on privacy and ethics (see the “Ethics and Privacy” section).

## **Results**

### **Characterizing Interest in Opioids, ROA, and Drug-Tampering Methods**

We applied the methodology described in the “Vocabulary Expansion” section to extract and expand domain-specific vocabularies and to characterize the temporal unfolding of interest in different opioid substances, routes of administration, and drug-tampering methodologies. We started from a review of the relevant medical research, collecting an initial set of terms referring to the most common opioid substances, ROA [6,10,31,34,38,39,41,63,64], and drug-tampering methods [41,63]. We expanded the original set with neighboring terms in a low-dimensional embedding space, and the outputs were reviewed and organized by a domain expert. The resulting vocabulary for opioid substances is shown in Table 3. It is worth noting that the vocabulary expansion procedure considerably increased the richness of the terminology related to the domain of interest and, consequently, the volume of conversations on Reddit that contained these terms. For example, for the heroin category, we observed a 62% growth in the retrieved relevant conversations (Table 3). We investigated the temporal unfolding of the popularity of the opioid substances, measured as the fraction of authors mentioning a substance over the entire opioid firsthand user base, for each trimester from 2014 to 2018. A binary characterization of the mentioning behavior at the user level was considered to discount potential biases due to users with high activity. We also provided a relative measure of popularity to account for the constantly increasing volume of active users on Reddit. Figure 2 shows a decrease in the usage of heroin and a rise in fentanyl and codeine.

The resulting vocabulary for routes of administration was further organized in a 2-level hierarchical structure, reported in Table 4. It is worth noting that the taxonomy does not have a strict medical interpretation, nor was it intended to be a comprehensive review. However, it can give structure to otherwise unstructured collections of words and help in the interpretation of the results.

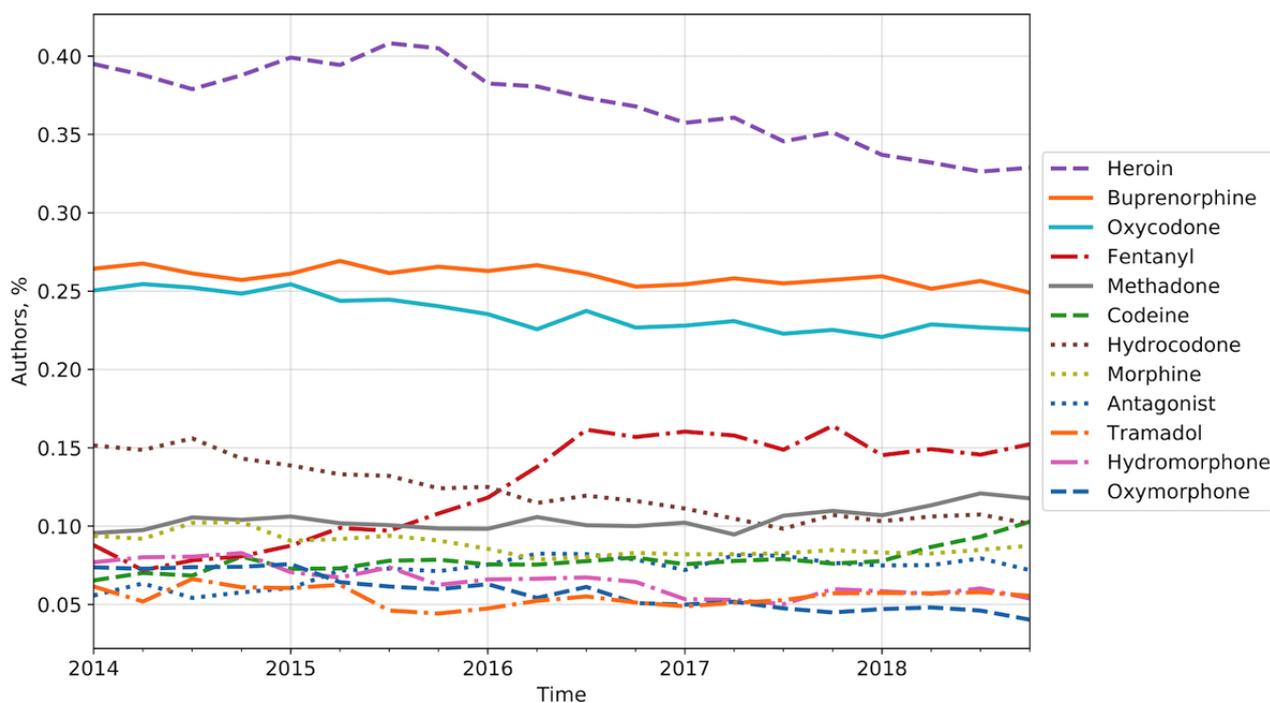
**Table 3.** Vocabulary of opioid substances. Starting from a candidate expansion set  $E^-$ , comprising 297 unique terms, the final expansion terms considered equivalent to a substance were gathered in the same class.

Substance	Terms	$\Delta$ volume, % <sup>a</sup>
Heroin	bth <sup>b</sup> , diacetylmorphine, diamorphine, dope, ecp <sup>b</sup> , goofball, goofballs, gunpowder, h, herion <sup>b</sup> , heroin <sup>b</sup> , heroine, heron, smack, speedball, speedballing, speedballs <sup>b</sup> , tar	62
Buprenorphine	bup, bupe <sup>b</sup> , buprenorphine <sup>b</sup> , butrans <sup>b</sup> , sub, suboxone <sup>b</sup> , subutex <sup>b</sup> , zub, zubsolv <sup>b</sup>	61
Hydrocodone	hydro, hydrocodone <sup>b</sup> , hydrocodones <sup>b</sup> , lortab <sup>b</sup> , lortabs <sup>b</sup> , norco <sup>b</sup> , norcos <sup>b</sup> , tuss, tussionex <sup>b</sup> , vic, vicoden, vicodin <sup>b</sup> , vicodins <sup>b</sup> , vicoprofen <sup>b</sup> , vics <sup>b</sup> , vikes, viks, zohydro <sup>b</sup>	38
Codeine	cocodamol, codein <sup>b</sup> , codeine <sup>b</sup> , codiene <sup>b</sup> , codine, dhc, dihydrocodeine <sup>b</sup> , prometh, sizzurp, syrup	28
Oxymorphone	g74, opana <sup>b</sup> , opanas, oxymorphone <sup>b</sup> , panda	25
Tramadol	desmethyltramadol, dsmt, tram, tramadol <sup>b</sup> , ultram <sup>b</sup>	22
Hydromorphone	dil, dilauded, dilaudid <sup>b</sup> , dilaudids, dillies <sup>b</sup> , dilly, dillys, diluadid <sup>b</sup> , hydromorph <sup>b</sup> , hydromorphone <sup>b</sup>	21
Oxycodone	15s, 30s, codone, contin, ms, oc, ocs, oxy <sup>b</sup> , oxycodone <sup>b</sup> , oxycontin <sup>b</sup> , oxycontins, oxycotin <sup>b</sup> , oxsy <sup>b</sup> , perc <sup>b</sup> , percocet <sup>b</sup> , percocets <sup>b</sup> , percocet, percocets, perc <sup>b</sup> , perk, roxi <sup>b</sup> , roxicodone <sup>b</sup> , roxie <sup>b</sup> , roxies <sup>b</sup> , roxis <sup>b</sup> , roxy <sup>b</sup> , roxycodone <sup>b</sup> , roxys <sup>b</sup>	14
Morphine	kadian, morph, morphine <sup>b</sup>	5
Fentanyl	acetylfentanyl <sup>b</sup> , butyr, butyrfentanyl, carf, carfent, carfentanil <sup>b</sup> , carfentanyl, duragesic <sup>b</sup> , fent <sup>b</sup> , fentanyl <sup>b</sup> , fents, fentynal, fetanyl, furanyl, sufentanil, u47700	4
Antagonist	nalaxone <sup>b</sup> , naloxone <sup>b</sup> , naltrexone, narcan <sup>b</sup> , narcon, revia, viv, vivitrol <sup>b</sup>	1
Methadone	mdone, methadone <sup>b</sup> , methodone <sup>b</sup>	1

<sup>a</sup>The increase in the volume of occurrences of a substance using the terms in the expanded vocabulary compared with only using the terms in  $\bar{Q}$ .

<sup>b</sup>Term in  $\bar{Q}$ .

**Figure 2.** Popularity of opioid substances among opioid firsthand users on Reddit. Each line represents the share of opioid firsthand users mentioning an opioid substance, measured quarterly from 2014 to 2018.



**Table 4.** Taxonomy defining the ROA categories and their corresponding terms. Primary ROA include all the expansion terms considered for the appropriate secondary ROA (original candidate expansion set comprised 199 unique terms).

Primary and secondary ROA <sup>a</sup>	Terms
<b>Ingestion</b>	
Oral	bolus, buccal, gulp, mouth, mouthful, oral <sup>b</sup> , orally, swallow <sup>b</sup>
Sublingual	sublingual <sup>b</sup> , sublingually, tongue, tounge
Drink	chug, drink, pour, pourin, sip <sup>b</sup> , sipper, sippin, swig, swish
Chew	chew <sup>b</sup> , chewy, chomp, gum
General ingestion	ingest <sup>b</sup> , ingestion
<b>Inhalation</b>	
Intranasal	intranasal, intranasally, nasal, nasally, nose, nostril, rail, sniff <sup>b</sup> , sniffer, sniffin, snoot, snooter, snort <sup>b</sup> , snorter, tooter
General inhalation	breath, breathe, dab, exhale, inhalation, inhale <sup>b</sup> , insufflate, insufflated, insufflating, insufflation, puff, toke, tokes, vap, vape, vaped, vapes, vaping, vapor, vaporise, vaporize, vaporizer, vapour
Smoking	bong, fume, hookah, pipe, smoke <sup>b</sup> , smoker, smokin, spliff
<b>Injection</b>	
Intramuscular	deltoid, imed, iming, intramuscular <sup>b</sup> , intramuscularly
Subcutaneous	subcutaneous <sup>b</sup> , subcutaneously, subq
Intravenous	arterial, bloodstream, intravenous <sup>b</sup> , intravenously, iv, ivd, ived, iving, ivs, vein, venous
General injection	bang, inject <sup>b</sup> , injectable, injection, parenteral, shoot, shot
Rectally	anal, anally, boof, boofed, boofing, bunghole, butt, pooper, rectal <sup>b</sup> , rectally
<b>Other ROA</b>	
Dermal	cutaneous, dermis, transdermal <sup>b</sup> , transdermally
Urogenital	vaginal
Intrathecal	intrathecal

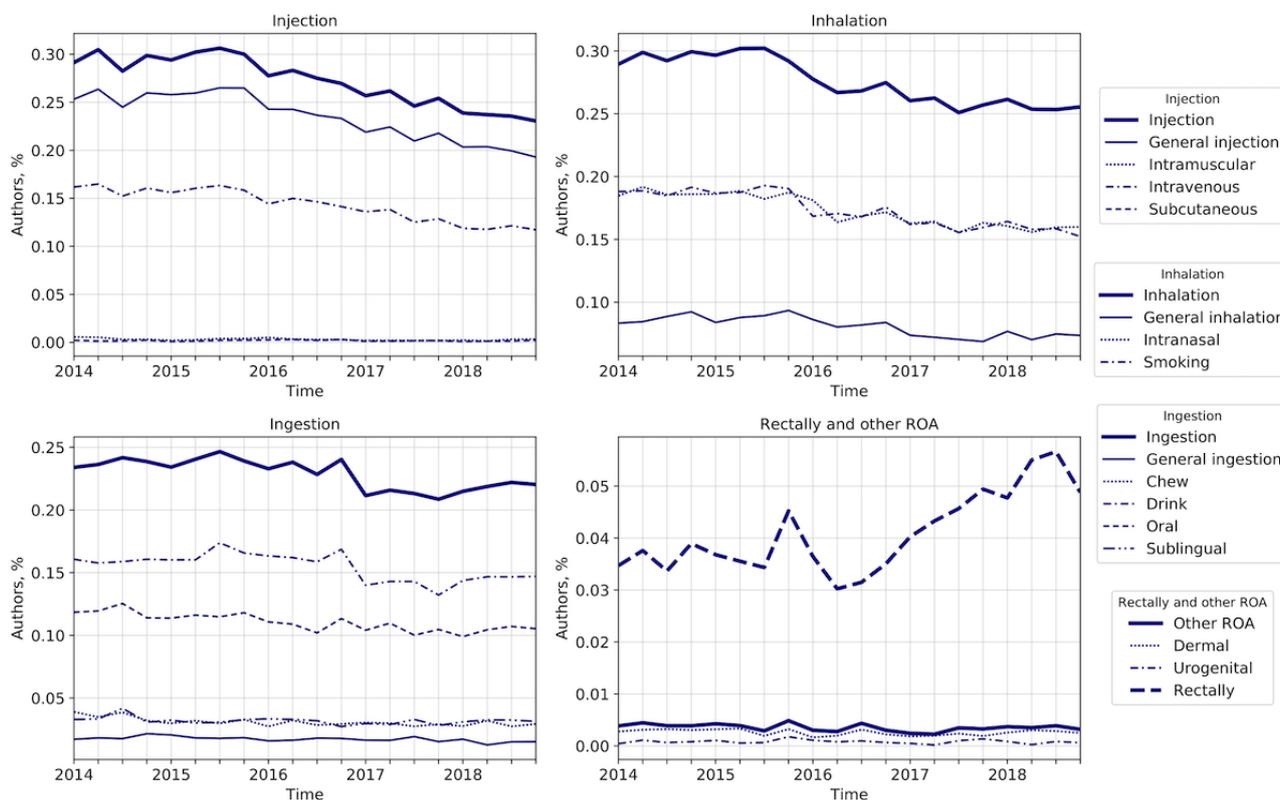
<sup>a</sup>ROA: routes of administration.

<sup>b</sup>Seed term *K*.

Figure 3 shows the estimated temporal evolution of the relative popularity of the routes of administration from 2014 to 2018, measured in quarterly snapshots. Finally, we extracted and organized the vocabulary related to drug-tampering techniques,

as shown in Table 5. In this paper, we considered the act of chewing pills a second-level route of administration under the ingestion category [8,31,32] instead of a drug-tampering method, as some research might suggest [41].

**Figure 3.** Popularity of routes of administration among opioid firsthand users on Reddit. Each line represents the fraction of opioid firsthand users mentioning an ROA-related term, measured quarterly from 2014 to 2018. Thick lines represent the share of authors mentioning primary ROA, evaluated by aggregating the contributions of all the corresponding secondary ROA. ROA: routes of administration.



**Table 5.** Vocabulary of drug-tampering methods. Expansion terms referring to the same drug-tampering method are grouped in the corresponding transformation classes (original candidate expansion set comprised 179 unique terms).

Transformation	Terms
Brew	brew <sup>a</sup> , brewer, homebrew <sup>a</sup>
Concentrate	concentrate <sup>a</sup> , concentrate, concentration, purify
Dissolve	desolve, dilute, dissolve, dissolved, dissolves, dissolve <sup>a</sup> , solute, solution, soluble, soluble
Evaporate	evap, evaporate
Extract	cwe <sup>a</sup> , extract <sup>a</sup> , extract, extraction
Grind	chop, crush <sup>a</sup> , crushable, crusher, grind <sup>a</sup> , grinded, grinder, ground, pulverize
Heat	boil, heat <sup>a</sup> , melt, microwave, overheat, simmer
Infusion	infuse, infusion <sup>a</sup> , tea, tincture
Peel	peal, peel, shave
Soak	soak <sup>a</sup> , submerge
Wash	rewash, rinse, wash

<sup>a</sup>Seed term *K*.

### Characterizing the Associations Between Opioid Substances, ROA, and Drug Tampering

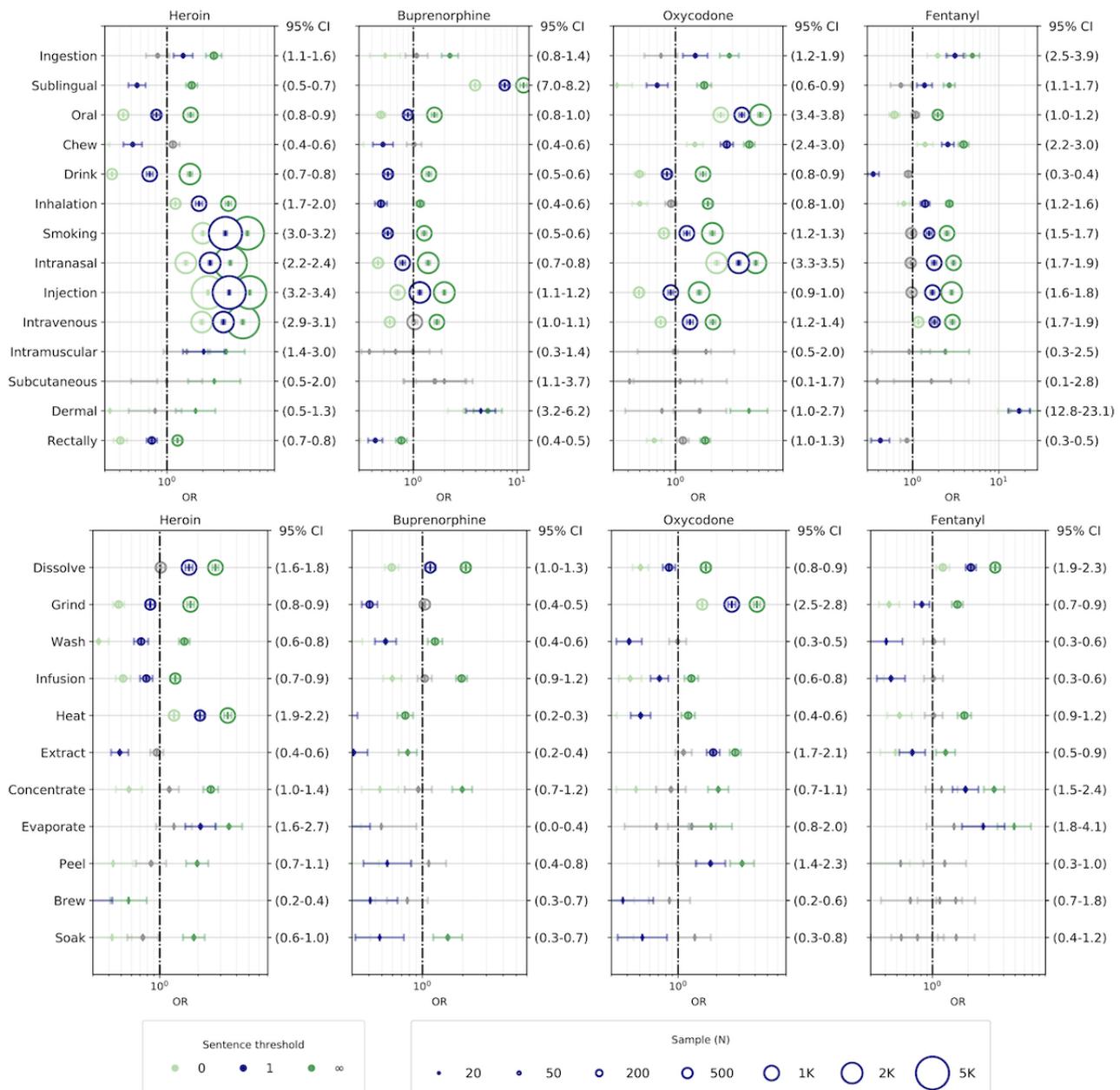
To investigate the strength of association between routes of administration, drug tampering, and opioid substances and to shed light on the interplay between the “how” and the “what” dimensions of opioid consumption, we estimated the ORs, 95%

confidence intervals, *P* values, and volume of the co-mentions among substances, routes of administration, and drug-tampering methods. The number of sentences in Reddit posts vary greatly, but the posts are generally quite short (approximately 50% of them have 2 sentences or less, as seen in [Multimedia Appendix 4](#)). However, as about 20% of posts have more than 10 sentences, one should be cautious in adopting a bag-of-words

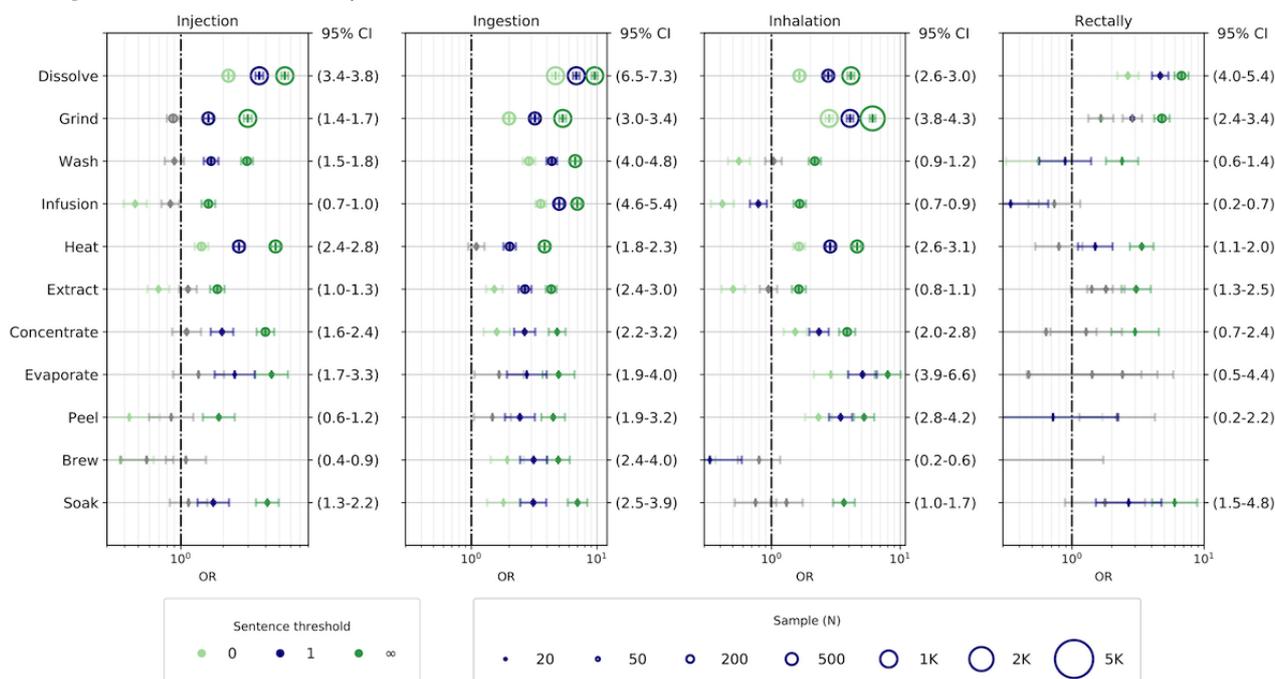
approach to measure co-occurring terms. To limit the chance of including spurious correlations due to the co-occurrences of terms far apart in the posts, we conservatively selected  $\rho=1$  (ie, considering only the co-occurrence of terms within a sentence or in the first adjacent sentences) for computing the OR. Figure 4 shows in blue the results of the analysis at  $\rho=1$ , matching 4 of the main widespread substances (ie, heroin, buprenorphine, oxycodone, and fentanyl) with the secondary ROA (upper panel) and the drug-tampering techniques (lower panel). Figure 5 shows the odds ratios of primary ROA and drug-tampering methods. For reference, the green markers represent the ORs obtained at

$\rho=0$  and  $\rho=\infty$  for the same categories. Multimedia Appendices 3-5 provide the complete set of results for all the substances identified and the secondary ROA. Due to the low representativeness of intrathecal and urogenital ROA with most of the tampering-related terms, we omitted those categories from the analysis. In the plots, the associations that are not statistically significant ( $P>.01$ ) are reported in gray, and the horizontal lines indicate the OR and the 95% confidence interval. The radius of the circle is proportional to the sample of co-mentions and the dashed vertical line corresponds to an OR of 1 for reference.

**Figure 4.** Odds ratios of the most widespread opioid substances with routes of administration (top row) and drug-tampering methods (bottom row). Labels on the right axis report the confidence interval at  $\rho=1$ . OR: odds ratio.



**Figure 5.** Odds ratios of the primary routes of administration (excluding other routes of administration) and drug-tampering methods. Labels on the right axis report the confidence interval at  $p=1$ . OR: odds ratio.



## Discussion

### Opioid Interest on Reddit

In this work, we identified over 3 million comments on 32 subreddits focused on discussing practices and implications of firsthand opioid use. We also selected a cohort of over 86,000 Reddit users interested in this topic. Such a large data set allowed us to assess the magnitude of the online interest in opioids and model its evolution during the 5 years of study, sadly verifying its rapidly increasing popularity. By the end of 2018, the opioid epidemic remained an escalating public health threat, and at the time of writing, the opioid crisis is still calling for countermeasures at scale. Hence, we believe our large data set may constitute a valid alternative source to advise decision making and a valuable starting point for future infodemiology research.

### Vocabulary Expansion

By observing the vocabularies in Tables 3-5 resulting from the expansion algorithm, we can ascertain the importance of enriching domain expertise with user-generated content and observe that some common features are captured across categories. Our method was able to detect synonyms and common short names, very specific acronyms (eg, “cwe” for cold water extraction [65]), slang expressions like “sippin” (often used when referring to the act of drinking codeine mixtures [63]), nicknames (eg, “panda” for oxymorphone), and polypharmacy instances (eg, “speedball” and “goofball” [66]). The vocabulary expansion underlines the use of prescription dosages (usually stamped on the tablets) in place of the commercial names of the substances (eg, “30s” for oxycodone). Moreover, we deduced that opioid firsthand users discussed variants of the substances (eg, “bth” and “ecp” for black tar heroin and East Coast powder), research chemical equivalents

(eg, “u47700” [67]), and formulations intended for veterinary use (eg, sufentanil, carfentanil).

ROA vocabulary included and categorized both medical terms, adding terms scarcely considered in previous studies, like “vaping,” and nonmedical or unconventional administration terms, such as “chewing,” “snorting,” “smoking,” and “boofing” [39]. Our taxonomy also enabled us to disambiguate common primary ROA, such as injection and ingestion, into specific secondary ones, like subcutaneous [39] and sublingual administrations.

Finally, the drug-tampering vocabulary captured tampering methods that modify the physical status of the substances, like crushing and peeling, and some methods aiming at altering the chemical characteristics of the substances, like dissolving, washing, and heating [41]. We believe that even if this vocabulary might not be exhaustive of all drug-tampering methods, it offers a novel evidence-based perspective on the topic compared with the existing literature. The expanded vocabularies proved essential to fully incorporating the language complexity of online discussions and taboo behaviors [68] into at-scale analyses. Hopefully, our contribution might be useful in the future to find and understand new abusive behaviors that are discussed online, ultimately driving future research to yield more effective prevention methods.

### Adoption Popularity of Opioid Substances and ROA

Considering the share of users mentioning a term to be a proxy of firsthand involvement in opioid-related activities and including topic-specific terminology, the longitudinal views in Figures 2 and 3 can be used to rank the popularity of nonmedical usage of opioid substances and ROA and their adoption trends. Ranking the substances by average share, we can see that heroin is by far the most popular substance, mentioned on average by 1 in every 3 users. Its share of users, though, is steadily

decreasing, with a loss of 10% reported in state-specific findings by Rosenblum et al [27]. Buprenorphine and oxycodone were the most mentioned prescription opioids; they showed fairly static behavior, while hydrocodone importance decreased over time [28], possibly due to more stringent prescription regulation starting in 2014 [69]. Fentanyl showed the most abrupt behavior, dramatically increasing since 2016. Its volume of mentions in 2018 increased by almost 1.5 times compared with 2014, confirming it as one of the most recent threats [5,28]. In contrast, we did not find evidence of drastic changes in oxymorphone adoption after its partial ban in 2017 [70]. ROA adoption was led by injection and inhalation, which were the most popular ROA across the years, mentioned by 1 of every 3 authors at their peak. These were followed closely by ingestion. Rectal use and other ROA involved, on average, a significantly lower share of users, around 5% and less than 1%, respectively. Nevertheless, rectal administration has shown a sharp increase in popularity since 2016, almost doubling its share. Administration through inhalation was equally staggered by the intranasal and smoking categories of secondary ROA, strong indicators that this route of administration is indeed capturing nonmedical use of opioids.

This work on understanding which substances are currently gaining popularity may give prevention programs a strategic advantage, especially if consumption trends can be localized geographically [12,30,71], focusing the interventions needed to prevent early adoption of emerging dangerous substances like fentanyl. Moreover, tracking the evolution of interest in prescription opioids might be useful for evaluating the efficacy of ban policies, as in the case of oxymorphone. Understanding which ROA are the most adopted might eventually help address targeted campaigns informing users on safer practices, develop better tamper-resistant prescription drugs, and ultimately better inform the health system of the health risks specific to opioid adoption.

### Characterizing the Association Between Substance Consumption, ROA, and Drug-Tampering Methods

By jointly considering the results of the odds ratios in Figures 4 and 5 and Multimedia Appendices 5-7, we can outline complex preferences for the nonmedical use of opioids, triangulating substance use, ROA, and drug-tampering methods. We noticed that the majority of substances exhibited more than one high odds ratio, both with ROA and drug-tampering methods, meaning that such substances might be consumed by users in multiple nonexclusive ways. Our analysis shows that for the most part, the expected medical and nonmedical routes of administration of each substance (ie, intended ROA or known abusive administration) had high odds ratios. For prescription opioids, oral (medical) use was often confirmed (eg, oxycodone: OR 3.6, 95% CI 3.4-3.8), while intranasal administration was often the preferred nonmedical ROA, followed by injection, especially intravenous administration (eg, hydromorphone: OR 9.1, 95% CI 8.6-9.8) [32,72]. As expected, heroin appeared to be most likely consumed through injection (OR 3.3, 95% CI 3.2-3.4) or smoking, if heated up on aluminum foil (OR 3.1, 95% CI 3.0-3.2). Heroin was the only substance that showed high correlations with this administration route. It was also reported to be snorted [64].

Besides confirming and quantifying some known behaviors, our analysis can provide additional insights on the nonmedical use of intended routes of administration. In accordance with the literature [31,32,40,73], we found evidence that abuse of prescription opioids may be associated with chewing the pills (eg, oxycodone: OR 2.7, 95% CI 2.4-3.0). From the analysis of ROA and drug-tampering methods, it appears that nonmedical oral administration was correlated with dissolving (OR 9.7, 95% CI 9.0-10.4), grinding, and washing the substances. In some cases, oral and chewing-related misuse of prescription opioids simply consisted of peeling (OR 5.1, 95% CI 2.6-9.9) the external coating, which is usually hard to chew or responsible for the extended-release effect. Even though some formulations, such as Opana ER (oxymorphone hydrochloride extended-release tablets; Endo Pharmaceuticals), are known to be tamper resistant to crushing, users can peel the tablets to get rid of the extended release coating for higher recreational effects. Injection usually requires that the substance be dissolved (OR 3.5, 95% CI 3.2-3.7), while inhalation requires that the substance be ground to powder, especially for intranasal abuse (OR 6.7, 95% CI 6.3-7.1).

Our method ultimately found evidence of unconventional nonmedical administration for most of the substances. We found a high correlation between dissolving and intranasal administration (OR 4.1, 95% CI 3.8-4.4), which may indicate the adoption of “monkey water,” the practice of dissolving soluble substances, like tar heroin and fentanyl patches, and using the resulting liquid as a nasal spray [36]. Fentanyl patches were also consumed in other unforeseen ways; an unexpectedly high OR of fentanyl and chewing (OR 2.6, 95% CI 2.2-3.0) suggests that prescription patches intended for transdermal use may be chewed for nonmedical use. Our analyses revealed the high odds of abuse of codeine via drinking (OR 4.0, 95% CI 3.7-4.3) codeine syrup, made by extracting or brewing the cough suppressants (OR 14.1, 95% CI 11.5-17.2) and forming the so-called lean or purple drank [7,63,74].

Buprenorphine, usually administered sublingually in its formulations without an antagonist, such as Subutex (buprenorphine; Indivior), and orally in combination with naloxone in the form of pills, such as Suboxone (buprenorphine-naloxone; Indivior) and Zubsolv (buprenorphine-naloxone; Orexo), measured exceptionally high odds of sublingual administration (OR 7.6, 95% CI 7.0-8.2). Evidence of nonmedical use of buprenorphine was also found in the association between dissolving and sublingual use (OR 18.9, 95% CI 16.8-21.3). Opioid firsthand users know that the opioid antagonist in buprenorphine-naloxone compounds has low bioavailability if dissolved under the tongue; hence, to achieve higher opioid effects and eliminate the antagonist, these compounds are generally taken sublingually and not through other ROA, with which buprenorphine shows negative associations.

Finally, our study shows that rectal administration is a viable and unforeseen option for the nonmedical use of some opioids, resulting in higher recreational effects, especially with hydromorphone (OR 5.2, 95% CI 4.6-6.0), morphine, and oxymorphone. Rectal administration showed high correlations with the dissolving, grinding, and soaking drug-tampering

methods, possible indicators of an unconventional route of administration, largely overlooked, which involves dissolving the substances in liquid water or alcohol (ie, “butt-chugging”) [39,75]. Subcutaneous administration was only weakly associated with morphine, suggesting that the practice of “skin popping” [38], which consists of injecting the substance in the tissues under the skin, is potentially not widespread.

The complex interactions between substance use, routes of administration, and drug tampering that can be unveiled with our methodology provide a broad yet detailed perspective on the nonmedical use of opioids, evidencing abusive behaviors in which unconventional ROA and drug tampering play a key role. Knowledge about abusive behaviors could be taken into consideration by physicians during treatment programs, allowing them to favor opioid medications that are less likely to be transformed and abused. Our results should be addressed with effective health policies, driving future clinical research to better focus its efforts on understanding health-related risks and guiding the production of new tamper-resistant and abuse-deterrent opioid formulations.

### Limitations and Future Work

We acknowledge some limitations in the present research. The population sampled on Reddit might have intrinsic social media biases, and it is likely not representative of the general population (eg, for gender, age, or ethnicity). Moreover, since we enrolled the users in our cohort based on their engagement in subcommunities focusing on firsthand use of opioids, we cannot exclude the possibility that in some cases, such users might have been reporting secondhand experiences, disseminating general news, or discussing intended medical drug use for pain management. We must also consider that the selected individuals were not clinically diagnosed with opioid use disorder. Future work will be devoted to building a classifier at the user level to identify individuals with opioid use disorder. We are aware that Reddit data have some gaps [76], but since the incompleteness mostly affects the years before 2010, we consider the overall results of our work to not be significantly biased. Other limitations are related to the analytic pipeline, where we narrowed our text analysis to term counts and co-occurrences, which might have produced spillover effects in comments discussing multiple topics and could have amplified the strength of cross-associations. Future work will include n-grams and more context-based language models. Finally, it is worth stressing that the measure of association through odds ratios should not be intended by any means as an indication of causal effects. This work is an observational study focusing on the characterization of a complex and faceted social phenomenon rather than the identification of determinants or interventions, and it shares the strengths and limitations of correlational studies, especially in medical research.

### Ethics and Privacy

Given the sensitive nature of the information shared, including users’ vulnerabilities and personal information, privacy and

ethical considerations are paramount. In this work, we followed the guidelines and directives in Eysenbach and Till [77], which describe recommendations to ethically conduct medical research with user-generated online data, and we relied on the vast experience of research works dealing with sensitive data gathered on social media [47,78-81]. The researchers had no interactions with the users and have no interest in harming any, and the analyses were performed and reported in the spirit of knowledge, prevention, and harm reduction. In this direction, it is worth noting that the subreddits under study are of public domain, are not password protected, and have thousands of active subscribers; users were fully aware of the public nature of the content they posted and of its free accessibility on the web. Moreover, Reddit offers pseudonymous accounts and strong privacy protection, making it unlikely that the true identity of a user can be recovered. Nevertheless, in order to further protect the privacy and anonymity of the users in our data set, all information about the names of the authors was anonymized before using the data for analysis. Moreover, every analysis performed was intended to provide aggregated estimates aimed at research purposes, and this work did not include any quotes or information that focused on single authors. Following the directives in Eysenbach and Till [77], our research did not require informed consent.

### Conclusions

In this work, we characterized opioid-related discussions on Reddit over 5 years, involving more than 86,000 unique users, and focused on firsthand experiences and nonmedical use. To address the complexity of the language in social media communications, especially in the presence of taboo behaviors such as drug abuse, we gathered a large set of colloquial and nonmedical terms that covered most opioid substances, routes of administration, and drug-tampering methods. We were able to characterize the temporal evolution of the discourse and identify notable trends, such as the surge in the popularity of fentanyl and the decrease in the relative interest in heroin. Focusing on routes of administration, we extended pharmacological and medical research with an in-depth characterization of how opioids substances are administered, since different practices imply different effects and potential health-related risks. We proposed a 2-layer taxonomy and corresponding vocabulary that enabled us to study both medical and recreational routes of administration. We demonstrated the presence of conventional nonmedical ROA (eg, intranasal administration and intravenous injection) and the spread of less conventional practices (eg, an increasing trend in rectal use). In particular, with reference to nonconventional ROA, we characterized for the first time at scale the phenomenon of drug tampering, which could have an impact on health outcomes, since it alters the pharmacokinetics of medications. The interplay between these dimensions was systematically characterized by quantitatively measuring the odds ratios, providing an insightful picture of the complex phenomenon of opioid consumption as discussed on Reddit.

## Acknowledgments

PB acknowledges support from the Intesa Sanpaolo Innovation Center. The funder had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript. RS was partially supported by the project Countering Online Hate Speech Through Effective On-line Monitoring, funded by Compagnia di San Paolo.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Schematic representation of the structure of Reddit. Reddit's most common access point is the front page, where the most relevant content of the moment is collected. The users can post on already-existing subreddits or they can create and moderate new ones on any topic of choice. In a subreddit, users can either create a new thread via a submission or indefinitely expand the conversation tree by commenting on an existing thread. The level of content moderation in a subreddit is solely decided by its moderators.

[[PNG File , 71 KB-Multimedia Appendix 1](#)]

## Multimedia Appendix 2

Subreddits discussing firsthand nonmedical use of opioids. An X marks the presence of a subreddit in a specific year.

[[XLSX File \(Microsoft Excel File\), 7 KB-Multimedia Appendix 2](#)]

## Multimedia Appendix 3

Relevant training parameters of the word embeddings. All the other parameters are set to default values. Two state-of-the-art word embedding models, namely word2vec, and GloVe, have been trained with all the comments and submissions in our subreddits dataset. After a-posteriori validation by a domain expert in terms of topical coherence, we choose word2vec as the reference word embedding model.

[[XLSX File \(Microsoft Excel File\), 6 KB-Multimedia Appendix 3](#)]

## Multimedia Appendix 4

Cumulative probability of finding  $n$  or fewer terms in a sentence for submissions and comments (left panel). Cumulative probability of having  $n$  or fewer sentences in a submission or a comment (right panel). Plots refer to the selected subreddit in 2018.

[[PNG File , 128 KB-Multimedia Appendix 4](#)]

## Multimedia Appendix 5

Odds Ratios of opioid substances and Secondary Routes of Administration. The central line and the bar mark the OR and the 95% confidence interval respectively, while the size of the circle is proportional to the sample of co-mentions. Measures that are not statistically significant ( $P > .01$ ) are reported in gray. Labels on the right axis report the Odds Ratio and the confidence interval.

[[PNG File , 694 KB-Multimedia Appendix 5](#)]

## Multimedia Appendix 6

Odds Ratios of opioid substances and Drug Tampering Methods. The central line and the bar mark the OR and the 95% confidence interval respectively, while the size of the circle is proportional to the sample of co-mentions. Measures that are not statistically significant ( $P > .01$ ) are reported in gray. Labels on the right axis report the Odds Ratio and the confidence interval.

[[PNG File , 577 KB-Multimedia Appendix 6](#)]

## Multimedia Appendix 7

Odds Ratios of Secondary Routes of Administration and Drug Tampering Methods. The central line and the bar mark the OR and the 95% confidence interval respectively, while the size of the circle is proportional to the sample of co-mentions. Measures that are not statistically significant ( $P > .01$ ) are reported in gray. Labels on the right axis report the Odds Ratio and the confidence interval.

[[PNG File , 598 KB-Multimedia Appendix 7](#)]

## References

- Centers for Disease Control and Prevention. Drug Overdose Deaths, Centers for Disease Control and Prevention website. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/drugoverdose/data/statedeaths.html> [accessed 2020-05-27] [WebCite Cache ID <https://www.cdc.gov/drugoverdose/data/statedeaths.html>]
- Kolodny A, Courtwright DT, Hwang CS, Kreiner P, Eadie JL, Clark TW, et al. The prescription opioid and heroin crisis: a public health approach to an epidemic of addiction. *Annu Rev Public Health* 2015 Mar 18;36:559-574. [doi: [10.1146/annurev-publhealth-031914-122957](https://doi.org/10.1146/annurev-publhealth-031914-122957)] [Medline: [25581144](https://pubmed.ncbi.nlm.nih.gov/25581144/)]
- Compton WM, Jones CM, Baldwin GT. Relationship between Nonmedical Prescription-Opioid Use and Heroin Use. *N Engl J Med* 2016 Jan 14;374(2):154-163. [doi: [10.1056/nejmra1508490](https://doi.org/10.1056/nejmra1508490)]
- Rose M. Are Prescription Opioids Driving the Opioid Crisis? Assumptions vs Facts. *Pain Med* 2018 Apr 01;19(4):793-807 [FREE Full text] [doi: [10.1093/pm/pnx048](https://doi.org/10.1093/pm/pnx048)] [Medline: [28402482](https://pubmed.ncbi.nlm.nih.gov/28402482/)]
- Ciccarone D. The triple wave epidemic: Supply and demand drivers of the US opioid overdose crisis. *Int J Drug Policy* 2019 Sep;71:183-188 [FREE Full text] [doi: [10.1016/j.drugpo.2019.01.010](https://doi.org/10.1016/j.drugpo.2019.01.010)] [Medline: [30718120](https://pubmed.ncbi.nlm.nih.gov/30718120/)]
- McCabe SE, Cranford JA, Boyd CJ, Teter CJ. Motives, diversion and routes of administration associated with nonmedical use of prescription opioids. *Addict Behav* 2007 Mar;32(3):562-575 [FREE Full text] [doi: [10.1016/j.addbeh.2006.05.022](https://doi.org/10.1016/j.addbeh.2006.05.022)] [Medline: [16843611](https://pubmed.ncbi.nlm.nih.gov/16843611/)]
- Agnich LE, Stogner JM, Miller BL, Marcum CD. Purple drank prevalence and characteristics of misusers of codeine cough syrup mixtures. *Addict Behav* 2013 Sep;38(9):2445-2449. [doi: [10.1016/j.addbeh.2013.03.020](https://doi.org/10.1016/j.addbeh.2013.03.020)] [Medline: [23688907](https://pubmed.ncbi.nlm.nih.gov/23688907/)]
- Katz N, Fernandez K, Chang A, Benoit C, Butler SF. Internet-based Survey of Nonmedical Prescription Opioid Use in the United States. *Clin J Pain* 2008;24(6):528-535. [doi: [10.1097/ajp.0b013e318167a087](https://doi.org/10.1097/ajp.0b013e318167a087)]
- Butler SF, Budman SH, Licari A, Cassidy TA, Lioy K, Dickinson J, et al. National addictions vigilance intervention and prevention program (NAVIPPRO): a real-time, product-specific, public health surveillance system for monitoring prescription drug abuse. *Pharmacoepidemiol Drug Saf* 2008 Dec;17(12):1142-1154. [doi: [10.1002/pds.1659](https://doi.org/10.1002/pds.1659)] [Medline: [18932173](https://pubmed.ncbi.nlm.nih.gov/18932173/)]
- Butler SF, Black RA, Cassidy TA, Dailey TM, Budman SH. Abuse risks and routes of administration of different prescription opioid compounds and formulations. *Harm Reduct J* 2011 Oct 19;8:29 [FREE Full text] [doi: [10.1186/1477-7517-8-29](https://doi.org/10.1186/1477-7517-8-29)] [Medline: [22011626](https://pubmed.ncbi.nlm.nih.gov/22011626/)]
- Curtis HJ, Croker R, Walker AJ, Richards GC, Quinlan J, Goldacre B. Opioid prescribing trends and geographical variation in England, 1998-2018: a retrospective database study. *Lancet Psychiatry* 2019 Feb;6(2):140-150. [doi: [10.1016/S2215-0366\(18\)30471-1](https://doi.org/10.1016/S2215-0366(18)30471-1)] [Medline: [30580987](https://pubmed.ncbi.nlm.nih.gov/30580987/)]
- Schifanella R, Vedove DD, Salomone A, Bajardi P, Paolotti D. Spatial heterogeneity and socioeconomic determinants of opioid prescribing in England between 2015 and 2018. *BMC Med* 2020 May 15;18(1):127 [FREE Full text] [doi: [10.1186/s12916-020-01575-0](https://doi.org/10.1186/s12916-020-01575-0)] [Medline: [32410615](https://pubmed.ncbi.nlm.nih.gov/32410615/)]
- Richards GC, Mahtani KR, Muthee TB, DeVito NJ, Koshiaris C, Aronson JK, et al. Factors associated with the prescribing of high-dose opioids in primary care: a systematic review and meta-analysis. *BMC Med* 2020 Mar 30;18(1):68 [FREE Full text] [doi: [10.1186/s12916-020-01528-7](https://doi.org/10.1186/s12916-020-01528-7)] [Medline: [32223746](https://pubmed.ncbi.nlm.nih.gov/32223746/)]
- van Amsterdam J, van den Brink W. The Misuse of Prescription Opioids: A Threat for Europe? *Curr Drug Abuse Rev* 2015;8(1):3-14. [doi: [10.2174/187447370801150611184218](https://doi.org/10.2174/187447370801150611184218)] [Medline: [26084418](https://pubmed.ncbi.nlm.nih.gov/26084418/)]
- Brownstein JS, Freifeld CC, Madoff LC. Digital Disease Detection — Harnessing the Web for Public Health Surveillance. *N Engl J Med* 2009 May 21;360(21):2153-2157. [doi: [10.1056/nejmp0900702](https://doi.org/10.1056/nejmp0900702)]
- Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009 Mar 27;11(1):e11 [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
- Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. *PLoS Comput Biol* 2012;8(7):e1002616 [FREE Full text] [doi: [10.1371/journal.pcbi.1002616](https://doi.org/10.1371/journal.pcbi.1002616)] [Medline: [22844241](https://pubmed.ncbi.nlm.nih.gov/22844241/)]
- Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. *J Med Internet Res* 2017 Oct 31;19(10):e353 [FREE Full text] [doi: [10.2196/jmir.6426](https://doi.org/10.2196/jmir.6426)] [Medline: [29089287](https://pubmed.ncbi.nlm.nih.gov/29089287/)]
- Hu H, Phan N, Geller J, Iezzi S, Vo H, Dou D, et al. An Ensemble Deep Learning Model for Drug Abuse Detection in Sparse Twitter-Sphere. *Stud Health Technol Inform* 2019 Aug 21;264:163-167. [doi: [10.3233/SHTI190204](https://doi.org/10.3233/SHTI190204)] [Medline: [31437906](https://pubmed.ncbi.nlm.nih.gov/31437906/)]
- Prieto JT, Scott K, McEwen D, Podewils LJ, Al-Tayyib A, Robinson J, et al. The Detection of Opioid Misuse and Heroin Use From Paramedic Response Documentation: Machine Learning for Improved Surveillance. *J Med Internet Res* 2020 Jan 03;22(1):e15645 [FREE Full text] [doi: [10.2196/15645](https://doi.org/10.2196/15645)] [Medline: [31899451](https://pubmed.ncbi.nlm.nih.gov/31899451/)]
- Ertugrul A, Lin Y, Taskaya-Temizel T. CASTNet: Community-Attentive Spatio-Temporal Networks for Opioid Overdose Forecasting. arXiv. 2019. URL: [https://arxiv.org/abs/1905.04714?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%253A+arxiv%252FQ5Xk+%2528ExcitingAds%2521+cs+updates+on+arXiv.org%2529](https://arxiv.org/abs/1905.04714?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%253A+arxiv%252FQ5Xk+%2528ExcitingAds%2521+cs+updates+on+arXiv.org%2529) [accessed 2020-12-11]
- Lu J, Sridhar S, Pandey R, Hasan M, Mohler G. Redditors in Recovery: Text Mining Reddit to Investigate Transitions into Drug Addiction. arXiv. 2019. URL: <https://arxiv.org/abs/1903.04081> [accessed 2020-12-11]
- Yang Z, Nguyen L, Jin F. Predicting Opioid Relapse Using Social Media Data. arXiv. 2018. URL: <https://arxiv.org/abs/1811.12169> [accessed 2020-12-11]

24. Chancellor S, Nitzburg G, Hu A, Zampieri F, De CM. Discovering alternative treatments for opioid use recovery using social media. 2019 May 04 Presented at: 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 124A. [doi: [10.1145/3290605.3300354](https://doi.org/10.1145/3290605.3300354)]
25. Phalen P, Ray B, Watson DP, Huynh P, Greene MS. Fentanyl related overdose in Indianapolis: Estimating trends using multilevel Bayesian models. *Addict Behav* 2018 Nov;86:4-10. [doi: [10.1016/j.addbeh.2018.03.010](https://doi.org/10.1016/j.addbeh.2018.03.010)] [Medline: [29631798](https://pubmed.ncbi.nlm.nih.gov/29631798/)]
26. Zhu W, Chernew ME, Sherry TB, Maestas N. Initial Opioid Prescriptions among U.S. Commercially Insured Patients, 2012-2017. *N Engl J Med* 2019 Mar 14;380(11):1043-1052 [FREE Full text] [doi: [10.1056/NEJMsa1807069](https://doi.org/10.1056/NEJMsa1807069)] [Medline: [30865798](https://pubmed.ncbi.nlm.nih.gov/30865798/)]
27. Rosenblum D, Unick J, Ciccarone D. The Rapidly Changing US Illicit Drug Market and the Potential for an Improved Early Warning System: Evidence from Ohio Drug Crime Labs. *Drug Alcohol Depend* 2020 Mar 01;208:107779. [doi: [10.1016/j.drugalcdep.2019.107779](https://doi.org/10.1016/j.drugalcdep.2019.107779)] [Medline: [31931266](https://pubmed.ncbi.nlm.nih.gov/31931266/)]
28. Black JC, Margolin ZR, Olson RA, Dart RC. Online Conversation Monitoring to Understand the Opioid Epidemic: Epidemiological Surveillance Study. *JMIR Public Health Surveill* 2020 Jun 29;6(2):e17073 [FREE Full text] [doi: [10.2196/17073](https://doi.org/10.2196/17073)] [Medline: [32597786](https://pubmed.ncbi.nlm.nih.gov/32597786/)]
29. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz J, et al. Social Media Based Analysis of Opioid Epidemic Using Reddit. *AMIA Annu Symp Proc* 2018;2018:867-876 [FREE Full text] [Medline: [30815129](https://pubmed.ncbi.nlm.nih.gov/30815129/)]
30. Balsamo D, Bajardi P, Panisson A. Firsthand Opiates Abuse on Social Media: Monitoring Geospatial Patterns of Interest Through a Digital Cohort. 2019 Presented at: WWW '19: The World Wide Web Conference; May 13-17, 2019; San Francisco, CA. [doi: [10.1145/3308558.3313634](https://doi.org/10.1145/3308558.3313634)]
31. Kirsh K, Peppin J, Coleman J. Characterization of prescription opioid abuse in the United States: focus on route of administration. *J Pain Palliat Care Pharmacother* 2012 Dec;26(4):348-361. [doi: [10.3109/15360288.2012.734905](https://doi.org/10.3109/15360288.2012.734905)] [Medline: [23675595](https://pubmed.ncbi.nlm.nih.gov/23675595/)]
32. Gasior M, Bond M, Malamut R. Routes of abuse of prescription opioid analgesics: a review and assessment of the potential impact of abuse-deterrent formulations. *Postgrad Med* 2016 Jan;128(1):85-96. [doi: [10.1080/00325481.2016.1120642](https://doi.org/10.1080/00325481.2016.1120642)] [Medline: [26566680](https://pubmed.ncbi.nlm.nih.gov/26566680/)]
33. Strang J, Bearn J, Farrell M, Finch E, Gossop M, Griffiths P, et al. Route of drug use and its implications for drug effect, risk of dependence and health consequences. *Drug Alcohol Rev* 1998 Jun;17(2):197-211. [doi: [10.1080/09595239800187001](https://doi.org/10.1080/09595239800187001)] [Medline: [16203485](https://pubmed.ncbi.nlm.nih.gov/16203485/)]
34. Young AM, Havens JR, Leukefeld CG. Route of administration for illicit prescription opioids: a comparison of rural and urban drug users. *Harm Reduct J* 2010 Oct 15;7:24 [FREE Full text] [doi: [10.1186/1477-7517-7-24](https://doi.org/10.1186/1477-7517-7-24)] [Medline: [20950455](https://pubmed.ncbi.nlm.nih.gov/20950455/)]
35. Katz N, Dart RC, Bailey E, Trudeau J, Osgood E, Paillard F. Tampering with prescription opioids: nature and extent of the problem, health consequences, and solutions. *Am J Drug Alcohol Abuse* 2011 Jul;37(4):205-217. [doi: [10.3109/00952990.2011.569623](https://doi.org/10.3109/00952990.2011.569623)] [Medline: [21517709](https://pubmed.ncbi.nlm.nih.gov/21517709/)]
36. Ciccarone D. Heroin in brown, black and white: structural factors and medical consequences in the US heroin market. *Int J Drug Policy* 2009 May;20(3):277-282 [FREE Full text] [doi: [10.1016/j.drugpo.2008.08.003](https://doi.org/10.1016/j.drugpo.2008.08.003)] [Medline: [18945606](https://pubmed.ncbi.nlm.nih.gov/18945606/)]
37. Carlson RG, Nahhas RW, Martins SS, Daniulaityte R. Predictors of transition to heroin use among initially non-opioid dependent illicit pharmaceutical opioid users: A natural history study. *Drug Alcohol Depend* 2016 Mar 01;160:127-134 [FREE Full text] [doi: [10.1016/j.drugalcdep.2015.12.026](https://doi.org/10.1016/j.drugalcdep.2015.12.026)] [Medline: [26785634](https://pubmed.ncbi.nlm.nih.gov/26785634/)]
38. Coon TP, Miller M, Kaylor D, Jones-Spangle K. Rectal insertion of fentanyl patches: a new route of toxicity. *Ann Emerg Med* 2005 Nov;46(5):473. [doi: [10.1016/j.annemergmed.2005.06.450](https://doi.org/10.1016/j.annemergmed.2005.06.450)] [Medline: [16271683](https://pubmed.ncbi.nlm.nih.gov/16271683/)]
39. Rivers Allen J, Bridge W. Strange Routes of Administration for Substances of Abuse. *Am J Psychiatry Residents J* 2017 Dec;12(12):7-11. [doi: [10.1176/appi.ajp-rj.2017.121203](https://doi.org/10.1176/appi.ajp-rj.2017.121203)]
40. McCaffrey S, Manser KA, Trudeau KJ, Niebler G, Brown C, Zarycranski D, et al. The natural history of prescription opioid abuse: A pilot study exploring change in routes of administration and motivation for changes. *J Opioid Manag* 2018;14(6):397-405. [doi: [10.5055/jom.2018.0472](https://doi.org/10.5055/jom.2018.0472)] [Medline: [30629276](https://pubmed.ncbi.nlm.nih.gov/30629276/)]
41. Mastropietro DJ. Drug Tampering and Abuse Deterrence. *J Dev Drugs* 2013;03(1):119. [doi: [10.4172/2329-6631.1000119](https://doi.org/10.4172/2329-6631.1000119)]
42. Manikonda L, Beigi G, Liu H, Kambhampati S. Twitter for sparking a movement, reddit for sharing the moment: #metoo through the lens of social media. arXiv. 2018. URL: <https://arxiv.org/abs/1803.08022> [accessed 2020-12-14] [WebCite Cache ID <https://arxiv.org/abs/1803.08022>]
43. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The Pushshift Reddit Dataset. arXiv. 2020. URL: <https://arxiv.org/abs/2001.08435> [accessed 2020-12-14]
44. Medvedev A, Lambiotte R, Delvenne J. The anatomy of Reddit: An overview of academic research. arXiv. 2018. URL: <https://arxiv.org/abs/2001.08435> [accessed 2020-12-14]
45. De Choudhury M, De S. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI.
46. Enes KB, Valadares Brum PP, Oliveira Cunha T, Murai F, Couto da Silva AP, Lobo Pappa G. Reddit weight loss communities: do they have what it takes for effective health interventions? 2018 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '18); December 3-6, 2018; Santiago, Chile.

47. Saha K, Kim SC, Reddy MD, Carter AJ, Sharma E, Haimson OL, et al. The Language of LGBTQ+ Minority Stress Experiences on Social Media. *Proc ACM Hum Comput Interact* 2019 Nov;3(CSCW):89 [FREE Full text] [doi: [10.1145/3361108](https://doi.org/10.1145/3361108)] [Medline: [32935081](https://pubmed.ncbi.nlm.nih.gov/32935081/)]
48. Baumgartner J. Pushshift Reddit. Pushshift. URL: <https://files.pushshift.io/reddit/> [accessed 2020-05-27]
49. Baumgartner J. r/datasets - I have every publicly available Reddit comment for research. Reddit. URL: <https://www.reddit.com/r/datasets/comments/3bxlg7/> [accessed 2020-05-27]
50. SpaCy. Spacy industrial-strength Natural Language Processing in Python. Spacy. URL: <https://spacy.io/> [accessed 2020-05-27]
51. Barabási AL. The origin of bursts and heavy tails in human dynamics. *Nature* 2005 May 12;435(7039):207-211. [doi: [10.1038/nature03459](https://doi.org/10.1038/nature03459)] [Medline: [15889093](https://pubmed.ncbi.nlm.nih.gov/15889093/)]
52. Malmgren RD, Stouffer DB, Campanharo ASLO, Amaral LAN. On universality in human correspondence activity. *Science* 2009 Sep 25;325(5948):1696-1700 [FREE Full text] [doi: [10.1126/science.1174562](https://doi.org/10.1126/science.1174562)] [Medline: [19779200](https://pubmed.ncbi.nlm.nih.gov/19779200/)]
53. Muchnik L, Pei S, Parra LC, Reis SDS, Andrade JS, Havlin S, et al. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Sci Rep* 2013;3:1783 [FREE Full text] [doi: [10.1038/srep01783](https://doi.org/10.1038/srep01783)] [Medline: [23648793](https://pubmed.ncbi.nlm.nih.gov/23648793/)]
54. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977 Mar;33(1):159. [doi: [10.2307/2529310](https://doi.org/10.2307/2529310)]
55. langdetect. Python Software Foundation. URL: <https://pypi.org/project/langdetect/> [accessed 2020-12-14] [WebCite Cache ID <https://pypi.org/project/langdetect/>]
56. pyclld2. Python Software Foundation. URL: <https://pypi.org/project/pyclld2/> [accessed 2020-10-12]
57. pyclld3. Python Software Foundation. URL: <https://pypi.org/project/pyclld3/> [accessed 2020-10-12]
58. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: NIPS'13: 26th International Conference on Neural Information Processing Systems; Dec 5-8, 2013; Lake Tahoe, NV.
59. Urban Dictionary. Urban Dictionary. URL: <https://www.urbandictionary.com> [accessed 2020-05-27]
60. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *JOSS* 2018 Sep;3(29):861. [doi: [10.21105/joss.00861](https://doi.org/10.21105/joss.00861)]
61. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 25-29, 2014; Doha, Qatar.
62. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. 2010 Presented at: 9th Python in Science Conference (SciPy 2010); June 28-July 3, 2010; Austin, TX.
63. Hart M, Agnich LE, Stogner J, Miller BL. 'Me and My Drank:' Exploring the Relationship Between Musical Preferences and Purple Drank Experimentation. *Am J Crim Just* 2013 Jun 1;39(1):172-186. [doi: [10.1007/s12103-013-9213-7](https://doi.org/10.1007/s12103-013-9213-7)]
64. Surratt HL, Kurtz SP, Buttram M, Levi-Minzi MA, Pagano ME, Cicero TJ. Heroin use onset among nonmedical prescription opioid users in the club scene. *Drug Alcohol Depend* 2017 Oct 01;179:131-138 [FREE Full text] [doi: [10.1016/j.drugalcdep.2017.06.034](https://doi.org/10.1016/j.drugalcdep.2017.06.034)] [Medline: [28772173](https://pubmed.ncbi.nlm.nih.gov/28772173/)]
65. Bausch J, Kershman A, Shear J, Lewis L. Tamper resistant lipid-based oral dosage form for opioid agonists. US Patent 8,273,798. 2012. URL: <https://patentimages.storage.googleapis.com/f6/a0/b2/087f721ed668c8/US8273798.pdf> [accessed 2020-12-11]
66. Ellis MS, Kasper ZA, Cicero TJ. Twin epidemics: The surging rise of methamphetamine use in chronic opioid users. *Drug Alcohol Depend* 2018 Dec 01;193:14-20. [doi: [10.1016/j.drugalcdep.2018.08.029](https://doi.org/10.1016/j.drugalcdep.2018.08.029)] [Medline: [30326396](https://pubmed.ncbi.nlm.nih.gov/30326396/)]
67. Prekupec MP, Mansky PA, Baumann MH. Misuse of Novel Synthetic Opioids. *J Addict Med* 2017;11(4):256-265. [doi: [10.1097/adm.0000000000000324](https://doi.org/10.1097/adm.0000000000000324)]
68. Allan K, Burrige K. Forbidden words: Taboo and the censoring of language. Cambridge, United Kingdom: Cambridge University Press; 2006.
69. 2014 - Final Rule: Rescheduling of Hydrocodone Combination Products From Schedule III to Schedule II. Drug Enforcement Administration, Department of Justice. URL: [https://www.deadiversion.usdoj.gov/fed\\_regs/rules/2014/fr0822.htm](https://www.deadiversion.usdoj.gov/fed_regs/rules/2014/fr0822.htm) [accessed 2020-05-27] [WebCite Cache ID [https://www.deadiversion.usdoj.gov/fed\\_regs/rules/2014/fr0822.htm](https://www.deadiversion.usdoj.gov/fed_regs/rules/2014/fr0822.htm)]
70. Oxymorphone (marketed as Opana ER) Information. US Food and Drug Administration. 2018. URL: <https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/oxymorphone-marketed-opana-er-information> [accessed 2020-05-27]
71. Basak A, Cadena J, Marathe A, Vullikanti A. Detection of Spatiotemporal Prescription Opioid Hot Spots With Network Scan Statistics: Multistate Analysis. *JMIR Public Health Surveill* 2019 Jun 17;5(2):e12110 [FREE Full text] [doi: [10.2196/12110](https://doi.org/10.2196/12110)] [Medline: [31210142](https://pubmed.ncbi.nlm.nih.gov/31210142/)]
72. Omidian A, Mastropietro D, Omidian H. Routes of Opioid Abuse and its Novel Deterrent Formulations. *J Develop Drugs* 2015;4(5):e1. [doi: [10.4172/2329-6631.1000141](https://doi.org/10.4172/2329-6631.1000141)]
73. Butler SF, Cassidy TA, Chilcoat H, Black RA, Landau C, Budman SH, et al. Abuse rates and routes of administration of reformulated extended-release oxycodone: initial findings from a sentinel surveillance sample of individuals assessed for substance abuse treatment. *J Pain* 2013 Apr;14(4):351-358. [doi: [10.1016/j.jpain.2012.08.008](https://doi.org/10.1016/j.jpain.2012.08.008)] [Medline: [23127293](https://pubmed.ncbi.nlm.nih.gov/23127293/)]

74. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of Codeine Misuse on Instagram: Content Analysis. *JMIR Public Health Surveill* 2018 Mar 20;4(1):e22 [FREE Full text] [doi: [10.2196/publichealth.8144](https://doi.org/10.2196/publichealth.8144)] [Medline: [29559422](https://pubmed.ncbi.nlm.nih.gov/29559422/)]
75. El Mazloum R, Snenghi R, Barbieri S, Feltracco P, Omizzolo L, Vettore G, et al. 'Butt-chugging' a new way of alcohol assumption in young people. *Eur J Public Health* 2015;25:1. [doi: [10.1093/eurpub/ckv170.089](https://doi.org/10.1093/eurpub/ckv170.089)]
76. Gaffney D, Matias JN. Caveat emptor, computational social science: Large-scale missing data in a widely-published Reddit corpus. *PLoS One* 2018;13(7):e0200162 [FREE Full text] [doi: [10.1371/journal.pone.0200162](https://doi.org/10.1371/journal.pone.0200162)] [Medline: [29979741](https://pubmed.ncbi.nlm.nih.gov/29979741/)]
77. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. *BMJ* 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [doi: [10.1136/bmj.323.7321.1103](https://doi.org/10.1136/bmj.323.7321.1103)] [Medline: [11701577](https://pubmed.ncbi.nlm.nih.gov/11701577/)]
78. Moreno MA, Goniou N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. *Cyberpsychol Behav Soc Netw* 2013 Sep;16(9):708-713 [FREE Full text] [doi: [10.1089/cyber.2012.0334](https://doi.org/10.1089/cyber.2012.0334)] [Medline: [23679571](https://pubmed.ncbi.nlm.nih.gov/23679571/)]
79. Chancellor S, Birnbaum ML, Caine ED, Silenzio VMB, De Choudhury M. A taxonomy of ethical tensions in inferring mental health states from social media. 2019 Presented at: 2019 ACM Conference on Fairness, Accountability, and Transparency; Jan 29-31, 2019; Atlanta, GA.
80. Ramírez-Cifuentes D, Freire A, Baeza-Yates R, Puntí J, Medina-Bravo P, Velazquez DA, et al. Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis. *J Med Internet Res* 2020 Jul 07;22(7):e17758 [FREE Full text] [doi: [10.2196/17758](https://doi.org/10.2196/17758)] [Medline: [32673256](https://pubmed.ncbi.nlm.nih.gov/32673256/)]
81. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring Online Discussions About Suicide Among Twitter Users With Schizophrenia: Exploratory Study. *JMIR Ment Health* 2018 Dec 13;5(4):e11483. [doi: [10.2196/11483](https://doi.org/10.2196/11483)]

## Abbreviations

**OR:** odds ratio

**ROA:** routes of administration

**UMAP:** uniform manifold approximation and projection

*Edited by G Eysenbach; submitted 08.06.20; peer-reviewed by A Roundtree, AM Auvinen; comments to author 01.08.20; revised version received 15.10.20; accepted 28.10.20; published 04.01.21*

*Please cite as:*

*Balsamo D, Bajardi P, Salomone A, Schifanella R*

*Patterns of Routes of Administration and Drug Tampering for Nonmedical Opioid Consumption: Data Mining and Content Analysis of Reddit Discussions*

*J Med Internet Res* 2021;23(1):e21212

URL: <https://www.jmir.org/2021/1/e21212>

doi: [10.2196/21212](https://doi.org/10.2196/21212)

PMID: [33393910](https://pubmed.ncbi.nlm.nih.gov/33393910/)

©Duilio Balsamo, Paolo Bajardi, Alberto Salomone, Rossano Schifanella. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 04.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.