

1 **Title** Can accurate demographic information about people who use  
2 prescription medications non-medically be derived from Twitter?

3  
4 **Authors:**

5 \*Yuan-Chi Yang, PhD<sup>1</sup>

6 Mohammed Ali Al-Garadi, PhD<sup>1</sup>

7 Jennifer S. Love, MD<sup>2</sup>

8 Hannah L. F. Cooper, ScD<sup>3</sup>

9 Jeanmarie Perrone, MD<sup>4</sup>

10 Abeed Sarker, PhD<sup>1,5</sup>

11  
12 <sup>1</sup>Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta,  
13 GA, United States;

14 <sup>2</sup>Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New  
15 York, NY, United States;

16 <sup>3</sup>Department of Behavioral Sciences & Health Education, Rollins School of Public Health,  
17 Emory University, Atlanta, GA, United States;

18 <sup>4</sup>Department of Emergency Medicine, Perelman School of Medicine, University of  
19 Pennsylvania, Philadelphia, PA, United States;

20 <sup>5</sup>Department of Biomedical Engineering, Georgia Institute of Technology and Emory  
21 University, Atlanta, GA, United States;

22  
23 \*Corresponding author

24 Postal address: 101 Woodruff Circle, 4th Floor East, Atlanta, GA 30322

25 Email: [yuan-chi.yang@emory.edu](mailto:yuan-chi.yang@emory.edu)

26 Phone: 404-727-6123

27

28

29

30

31

32

33

## 1 **Abstract**

2 Traditional surveillance mechanisms for nonmedical prescription medication use (NPMU) involve  
3 substantial lags. Social media-based approaches have been proposed for conducting close-to-real-  
4 time surveillance, but such methods typically cannot provide fine-grained statistics about  
5 subpopulations. We address this gap by developing methods for automatically characterizing a  
6 large Twitter NPMU cohort (n=288,562) in terms of age-group, race, and gender. Our methods  
7 achieved 0.88 precision (95%-CI: 0.84-0.92) for age-group, 0.90 (95%-CI: 0.85-0.95) for race,  
8 and 0.94 accuracy (95%-CI: 0.92-0.97) for gender. We compared the automatically-derived  
9 statistics for the NPMU of tranquilizers, stimulants, and opioids from Twitter to statistics reported  
10 in traditional sources (eg., the National Survey on Drug Use and Health). Our estimates were  
11 mostly consistent with the traditional sources, except for age-group-related statistics, likely caused  
12 by differences in reporting tendencies and representations in the population. Our study  
13 demonstrates that subpopulation-specific estimates about NPMU may be automatically derived  
14 from Twitter to obtain early insights.

15

## 1 INTRODUCTION

2 Substance use (SU), including non-medical prescription medication use (NPMU), has  
3 been a major public health problem in the United States (US) for decades. Overdose deaths due  
4 to the SU have steadily increased over the years, regardless of prevention measures.<sup>1</sup> In 2019, the  
5 SU-related overdose death rate increased by 4.3% to 21.6 per 100,000 population,<sup>2</sup> about 20  
6 times higher than the recorded rate in 1980.<sup>1</sup> In the 12 months preceding April 2021, over  
7 100,000 SU-related deaths are expected, the highest ever recorded in a 12-month period.<sup>3</sup> Due to  
8 the enormity of the SU epidemic, the US government has announced the deployment of  
9 unprecedented resources.<sup>4</sup>

10 There are also significant disparities related to SU disorder (SUD) and the associated  
11 health outcomes. Many recent studies have highlighted the disparities depending on  
12 socioeconomic status, race/ethnicity, gender identity/biological sex, community, criminal  
13 history, and healthcare coverage.<sup>5-12</sup> For example, studies have shown that non-Hispanic Blacks  
14 and Hispanics are less likely to receive buprenorphine treatment compared to Whites, and  
15 women are less likely than men.<sup>11,13-15</sup> Moreover, non-Hispanic Blacks and American Indians  
16 and Alaska Natives (AIAN) experienced the highest increases in the drug overdose mortality  
17 rates in 2019 and 2020,<sup>16</sup> while non-Hispanic Blacks experienced much higher increase of the  
18 mortality rates due to stimulants and opioids co-ingestion compared to non-Hispanic Whites.<sup>17</sup> It  
19 has also been reported that people with lower income, living in non-metro urbanized regions, or  
20 uninsured are more likely to suffer from SUD.<sup>18</sup> Multiple disparities may co-exist, and  
21 exacerbate the likelihood of SU/SUD. Consequently, non-Hispanic Blacks, Hispanic/Latino  
22 persons, AIAN, and Native Hawaiian and Other Pacific Islanders (NHOPI), who also have low  
23 insurance coverage rates, face substantial SUD related disparities.<sup>19</sup> Distinct demographic groups

1 may also have their own unique cultural and historical contexts and norms, consequently  
2 increasing the challenges associated with surveillance and response.

3 A key to effectively tackle the epidemic and alleviate the disparities is to improve  
4 surveillance, specifically to accelerate the data curation process to provide timely, actionable  
5 insights, and improve decision making.<sup>20,21</sup> Traditional surveillance approaches and/or sources of  
6 data include surveys, such as those conducted by the National Survey on Drug Use and Health  
7 (NSDUH),<sup>22</sup> poison control centers,<sup>23</sup> hospital data about treatment admissions and discharge,<sup>24</sup>  
8 overdose-related emergency department visits (EDV),<sup>25,26</sup> and overdose death records.<sup>27</sup> Such  
9 traditional surveillance systems have considerable lags associated with the cycle of data  
10 collection, organization and release. For example, the 2020 NSDUH Annual National Report  
11 was not available until the end of October 2021. Due to such lags, trends in SU/overdose only  
12 detected and understood retrospectively, often after considerable damage has already been done  
13 and/or SU patterns have shifted. The lag is particularly problematic since the SU/overdose  
14 epidemic has been continuously evolving over the years. For example, the primary contributor of  
15 overdose-related deaths in the early 2000s was cocaine, which was later taken over by  
16 prescription opioids followed by heroin.<sup>1</sup> Also, in recent years, there have been notable increases  
17 in deaths due to synthetic opioids (*eg.*, fentanyl) and psychostimulants (*eg.*, methamphetamine),<sup>27-</sup>  
18 <sup>30</sup> but the current trajectory will only be known after months based on traditional surveillance.  
19 Therefore, there is an urgent need for close-to-real-time surveillance system for SU.

20 To address the shortcomings of traditional approaches, social media have been proposed  
21 as potential resources for timely surveillance.<sup>31-33</sup> Over 220 million Americans (~70% of the  
22 population) use social media, and many discuss health-related topics. There is thus the potential  
23 to leverage such patient-generated information for conducting timely surveillance via social

1 media. Social media data, such as that from Twitter, have been shown to contain self-reported  
2 information about SU, including NPMU, which can be detected in real-time.<sup>31,34-38</sup> While social  
3 media are promising data sources for obtaining close to real-time insights, data from them are  
4 typically massive and noisy, and extracting knowledge from them requires the development of  
5 advanced artificial intelligence methods involving natural language processing (NLP) and  
6 machine learning. However, research on social media-based SU surveillance is still in its infancy  
7 and many developments are required. Motivated lags associated with traditional surveillance  
8 systems, we attempted to equip advanced social media-based approaches with effective tools for  
9 providing timely understanding of subpopulation-level statistics, which can be used for close-to-  
10 real-time surveillance and for studying disparities.

11 Ideally, SU surveillance data needs to cover the full range of demographics (eg, race, age,  
12 gender, geographical area, socioeconomic status), and contain sufficient granularity to observe  
13 subtle differences among different demographic groups. Some studies have combined volume of  
14 social media information with meta-data such as geolocation and timestamps to derive  
15 geolocation-specific SU estimates that correlate with statistics from traditional data sources such  
16 as the NSDUH and CDC Wonder database.<sup>18,36-39</sup> However, to the best of our knowledge, no past  
17 study has attempted to estimate demographic information, such as age and gender, from a social  
18 media-based cohort. There are multiple reasons why such studies have not been conducted in the  
19 past despite their potential utility—the information is not directly available from meta-data, and  
20 the proposed estimation methods in the literature do not have adequate accuracy or granularity.<sup>40-</sup>  
21 <sup>50</sup> This poses a barrier to conducting fine-grained, subpopulation-specific research using such  
22 data—a clear disadvantage compared to the NSDUH and other traditional sources. Methods for  
23 accurately and automatically estimating the distributions of key demographic features in social

1 media subscriber cohorts can enable such fine-grained subpopulation-level analyses and  
2 comparisons.

3 In this paper, we describe the development and validation of methods for automatically  
4 estimating demographic distributions (age-group, gender, and race) in a Twitter cohort consisting  
5 of subscribers who self-reported NPMU. We integrated our developed novel methods to  
6 establish a data-centric cohort characterization pipeline, and applied the pipeline on the Twitter  
7 NPMU cohort. To validate our pipeline, we compared the distributions estimated from Twitter to  
8 those reported in traditional sources [NSDUH 2019<sup>18</sup> and Nationwide Emergency Department  
9 Sample (NEDS)<sup>51</sup>] for prescription stimulants, tranquilizers, and opioid pain relievers.

10

## 11 **RESULTS**

### 12 **Twitter nonmedical use cohort**

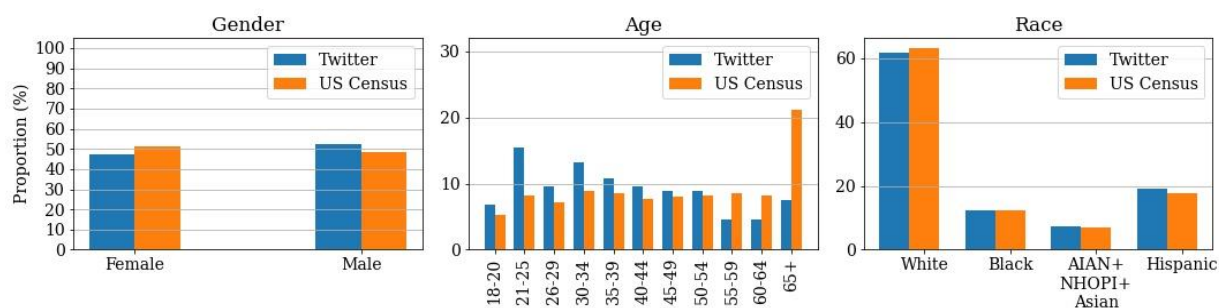
13 We collected tweets mentioning prescription medications and detected self-reported  
14 NPMU using a supervised classification system.<sup>35</sup> Posts were collected from March 6, 2018 to  
15 April 30, 2021. Our system detected 482,902 NPMU-indicating tweets and extracted their  
16 authors' meta-data, including post history, if available. In this manner, we collected 288,562  
17 Twitter subscribers' posts (NPMU cohort).

### 18 **Gender, age, and race distribution**

19 The gender, age, and race proportions for Twitter subscribers, estimated from the 2018  
20 Twitter Survey conducted by the Pew Research Center,<sup>52</sup> and those reported in the US Census<sup>18</sup>  
21 are shown in Figure 1. The estimated gender and race proportions from the two sources are  
22 comparable, while the age proportions are substantially different. Compared to the US Census

1 data, Twitter has marginally lower proportions of females (4% less) and Whites (1.5% less), and  
2 more Hispanics (1.5% more), which may be explained by the overcount of Whites and  
3 undercount of Hispanics in Census.<sup>53</sup> The closeness of the proportions from the two sources  
4 suggest that Twitter-based estimates specific to gender and race may be representative of the  
5 country's population. In contrast, in terms of age, Twitter has an overrepresentation of younger  
6 people compared to the census estimates. Specifically, the proportion of people in the 18-25  
7 group is approximately 10% higher, and the proportion for the 55+ years group is 20% lower on  
8 Twitter compared to the census estimates. The overrepresentation of younger people on Twitter,  
9 and social media in general, is a well-known phenomenon.

10 **Figure 1.** Gender, age, and race proportions estimated from Twitter and those reported in US  
11 census.



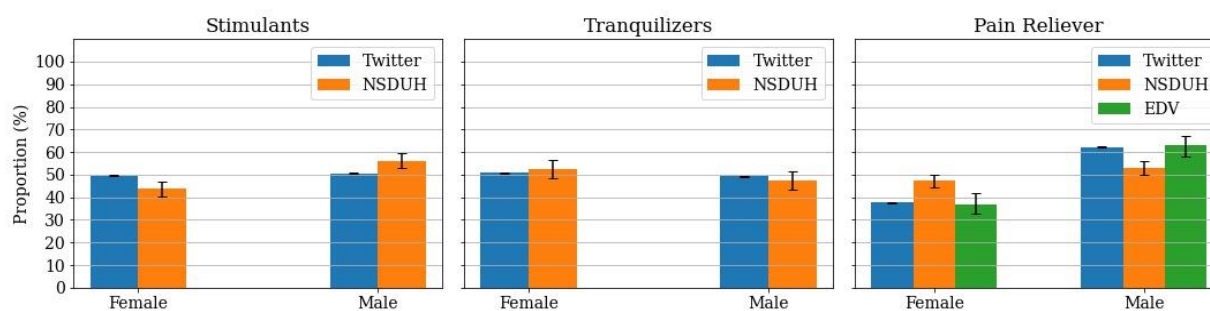
12

### 13 **Gender distribution estimates**

14 The estimated gender proportions from Twitter data and the gender proportions from  
15 traditional sources, including NSDUH and NEDS, are given in Figure 2 (details in Supplement  
16 Table S1). The three categories of medications included were opioid pain relievers, tranquilizers,  
17 and stimulants. For NPMU of tranquilizers, the estimated Twitter proportions are within the 95%  
18 confidence intervals of the proportions reported in the NSDUH. For stimulants, the Twitter  
19 proportion estimate for females is slightly higher than the NSDUH reported number (~5%). For

1 opioids, the estimated proportions are substantially different between Twitter and the NSDUH  
2 (~10%). Specifically, the proportion for females from Twitter is lower than the NSDUH  
3 estimates. Interestingly, however, we found that the numbers reported by the NSDUH also differ  
4 in terms of proportions from the opioid-related EDVs reported in NEDS, but the estimates from  
5 the latter are very close to the Twitter proportions (no significant difference). This suggests that  
6 estimates derived from Twitter may be more predictive of overdose-related events rather than  
7 nonmedical use for this category.

8 **Figure 2.** Gender distributions for nonmedical prescription medication use estimated from  
9 Twitter and those reported in the NSDUH. For opioid pain relievers, the gender distribution of  
10 overdose-related emergency medicine visits is also provided.



11

12

13

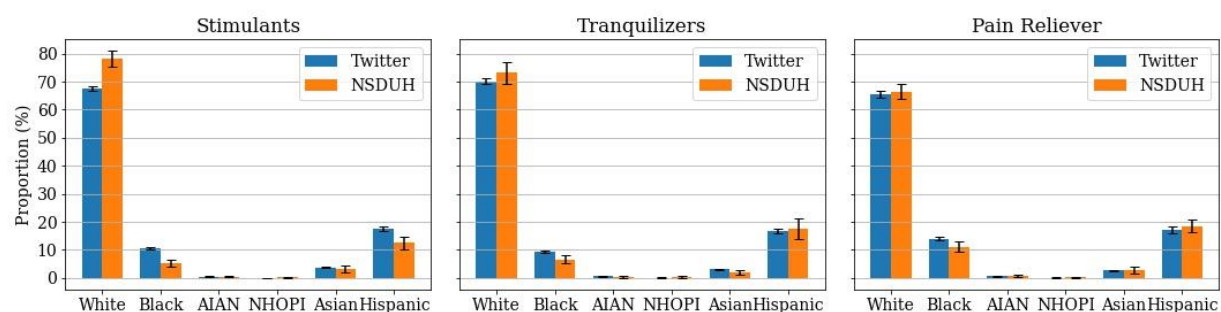
#### 14 **Race distribution estimates**

15 The estimated race proportions from Twitter and the proportions from the NSDUH are  
16 shown in Figure (details in Supplement Table S2). The estimated distributions are similar  
17 between the two data sources, and each proportion from Twitter is either within or close to the  
18 95% confidence interval of the corresponding proportion reported in the NSDUH. For all  
19 medication categories, the majority of people who reported nonmedical use are White, followed



1 by Hispanic and Black. Asians who reported nonmedical use only represent about 4% or less of  
2 the cohort; American Indian and Alaska Native (AIAN) and the Native Hawaiian or Other  
3 Pacific Islander (NHOPI) groups each represent less than 1% of the cohort. Importantly, the  
4 Twitter data had representation from all of the minority races. The most prominent differences  
5 between Twitter and the NSDUH are for the White and Hispanic stimulant groups, and for  
6 Blacks across all medications (Twitter estimates are higher than NSDUH).

7 **Figure 3.** Race distributions for nonmedical prescription medication use estimated from  
8 Twitter and those reported in the NSDUH.



9

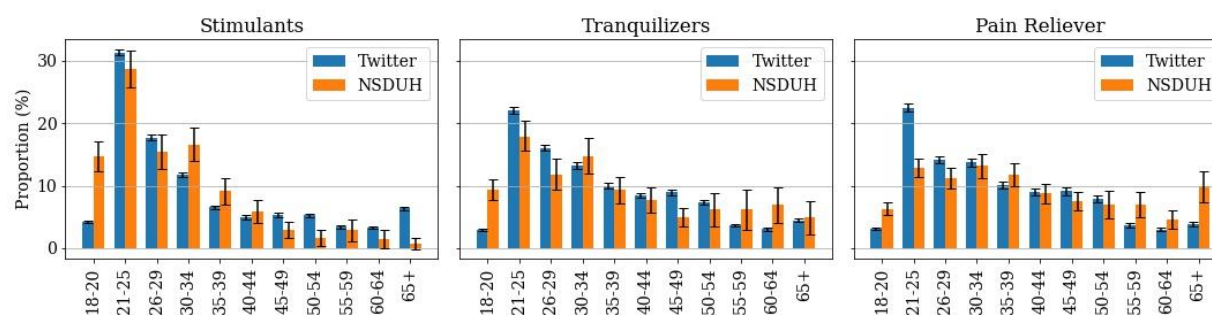
10

### 11 **Age-group distribution estimates**

12 The estimated age-group proportions from Twitter data and the proportions from the  
13 NSDUH are shown in Figure (details in Supplement Table S3). For most age-groups, the  
14 Twitter and the NSDUH estimates are similar. The most prominent differences are for young  
15 adults (18-20 and 21-25) and the elderly (65+). The estimated proportions from Twitter are  
16 consistently lower for the 18-20 group and higher for the 21-25 group. For the 65+ group, the  
17 estimated Twitter proportion is higher for stimulants, similar for tranquilizers, and lower for pain  
18 relievers compared to the NSDUH numbers. For opioid pain relievers, the estimated Twitter  
19 proportion for the 21-25 group is approximately 10% higher and for the 65+ group 6% lower

1 compared to the NSDUH. For tranquilizers, the estimated Twitter proportion for the 18-20 group  
2 is approximately 6% lower compared to the NSDUH, but the proportions for the 65+ group are  
3 not significantly different. For stimulants, the estimated Twitter proportion for the 18-20 group is  
4 approximately 10% lower and for the 65+ group 6% higher compared to the NSDUH.

5 **Figure 4.** Age-group distributions for nonmedical prescription medication use estimated from  
6 Twitter and those reported in the NSDUH.



7  
8  
9

## 10 DISCUSSION

11 To the best of our knowledge, our current work is the first to automatically estimate the  
12 distribution of demographic characteristics in a large Twitter cohort—in this case a cohort of  
13 subscribers who self-reported NPMU—and compare the automatically-obtained distributions  
14 with those reported in traditional sources. Our experiments validate that most of the estimates  
15 derived from Twitter are consistent with those reported in traditional sources, such as the  
16 NSDUH and NEDS. The major differences were in the age-group-based estimates, specifically  
17 for young adults (18-20 and 21-25) and the elderly (65+).

18 The NSDUH is conducted as a survey among noninstitutionalized population in the US  
19 and thus is limited by the respondents' truthfulness and exclusion of individuals in hospitals,

1 prisons, or even treatment centers.<sup>54</sup> It is reported that the respondents tend to under-report or  
2 over-report on surveys, and this tendency is influenced by their demographics including gender,  
3 race, or age.<sup>55-59</sup> For example, among surveyed Cocaine users, African American, young adults  
4 (18-30), and female are more inclined to under-report.<sup>55</sup> Though Twitter data also relies on  
5 individual users' truthfulness and willingness to share, we suspect that the default anonymity of  
6 Twitter accounts renders the Twitter data suffer less demographic-wise under-reporting than the  
7 NSDUH and, thus, might be better suited for analyzing subpopulation differences than the  
8 NSDUH. We speculate that the closeness of the gender and race distributions for Twitter  
9 subscribers and the US population, as depicted in Figure 1, is a key reason for the Twitter-based  
10 estimates to be similar to the NSDUH, while the differences might be explained by the  
11 under/over-reporting tendency. For example, the under-reporting tendency of females on  
12 Cocaine might help explain the apparent overestimation of female stimulant users on Twitter,  
13 and the under-reporting tendency of African American might be crucial to understand the  
14 overestimation of African American users for all three medication categories on Twitter.<sup>55</sup>  
15 Similarly, the different tendencies of under/over-reporting and Twitter usage among age groups  
16 might contribute to the differences in the age-wise estimates. Also, though the Twitter data is  
17 limited to those who have internet access, it may capture a certain portion of the institutionalized  
18 or before/after their institutionalized periods. It is even possible that Twitter is less biased against  
19 the incarcerated Black population than the NSDUH.

20

## 21 **Evaluation of performance as a surveillance system**

22 Our pipeline is advantageous in its timeliness, flexibility, simplicity, and stability. The  
23 data collection can be done near-real time and continuously on a personal desktop with internet

1 connection while requiring minimal human supervision. Instead of structured answers to  
2 questionnaires, our collected data contain salient unstructured text information, allowing data  
3 mining with research questions evolving over time. For example, the collected tweets contain  
4 information regarding how and why the authors are using the medications. To collect similar  
5 information from surveys usually requires incorporating prior knowledge into question design,  
6 but that is not necessary for our pipeline. However, since the data is usually massive, the data  
7 analysis is typically done using NLP and machine-learning methods. The advantage is that, once  
8 the tailored scripts are developed, it can often be run automatically and on the fly. To expand the  
9 pipeline (eg, to collect users of illicit drugs), it would only require a university informatics team  
10 to dedicate a few months from initial exploration to operational prototypes. These advantages of  
11 our system fit nicely into the CDC's Data Modernization Initiative, Priority 2 - Accelerate Data  
12 into Action to Improve Decision-Making and Protect Health.

13 However, the data quality, acceptability, sensitivity, the predicted value positive, and  
14 representativeness for our pipeline are inevitably limited by Twitter's user base and Twitter  
15 users' willingness to share their information publicly and their truthfulness. Demographic  
16 information is often not disclosed and we have no method to validate the users' claims. Notably,  
17 our data does not rely on memory as much as surveys, as the time leap is limited to the time of  
18 non-medical use to the time of posting. Also, we speculated that the users might be more truthful  
19 in their self-disclosure because they have more power to choose what they are willing to share  
20 publicly. Therefore, though our method does have disadvantages, we believe it is an important  
21 complimentary surveillance system with high potential.

22

## 23 **Related work**

1           Our work is not the first that aims to detect demographic information from social media  
2 data, although it is the first to develop and apply such methods for characterizing a specific  
3 cohort. Past studies have proposed cohort characterization methods, including gender,<sup>40-43</sup>  
4 age,<sup>42,44-46</sup> and race.<sup>47-50</sup> Typically, these pipelines comprised supervised classification methods,  
5 and used subscribers' meta-data including names, usernames, bio, past tweets, or even images as  
6 features. Among these, the gender detection methods were reported to be the most accurate, with  
7 classifiers achieving accuracies above 94%.<sup>60</sup> In contrast, race and age estimation pipelines  
8 proposed in past research had not obtained high accuracies. They also often do not provide  
9 enough granularity. The race estimation pipelines reported in the literature usually focus only on  
10 four categories (White, Black, Asian, and Hispanic/Latino) or less, leaving out AIAN and  
11 NHOPI,<sup>47-50</sup> although AIAN has the highest overdose mortality rate among all race group.<sup>16,61</sup>  
12 For age detection, the groupings often do not match those defined in the NSDUH, making  
13 comparisons impossible.<sup>44-46</sup> Though re-grouping is possible for a few methods, they were not  
14 developed based on Twitter and thus may have limited applicability.<sup>42</sup> In our work, we  
15 developed age and race estimation pipelines with high precision and fine granularity (11 for age  
16 and six for race) based on our Twitter data. Due to the paucity of annotated datasets, we applied  
17 a search-based approach that employs text pattern matching for detecting self-disclosed age and  
18 race. Because there is no gold-standard for the negative case (*ie.*, subscribers who have not self-  
19 disclosed their age or race using the specified pattern), we focused on improving precision while  
20 maintaining acceptable retrieval rate. Precision is preferred over recall (*ie.*, some cohort  
21 members will be missed) since the number of Twitter cohort members is large and growing over  
22 time, so obtaining sufficient numbers of people from each category is not a bottleneck.  
23

## 1 **Potential applications and future work**

2           There are several aspects of our automatic pipeline that can be improved in the future.

3       First, we can improve the performance of the classifier that detects the self-disclosures of

4       NPMU. One potential route is to examine if the classifier underperforms for any demographic

5       group due to different self-disclosure behaviors, and then fine-tune the classifier accordingly.

6       Second, we may make our findings more reliable by further improving the age group and race

7       characterization methods. Three potential directions include (i) annotating more tweets matched

8       by the text patterns, (ii) enriching the set of the text patterns, and/or (iii) replacing the rule-based

9       method with a machine-learning based classifier. Third, the pipeline can be extended to illicit

10       substances, including opioids such as heroin, and stimulants such as methamphetamine.

11       Collecting a cohort of people who self-report the use of illicit substances is an important and

12       natural extension to our current work, as Literature suggests that people who report NPMU may

13       also be exposed to illicit substances.<sup>62</sup> Fourth, our pipeline provides access to targeted minority

14       groups, which is of great interest to the communities and several government agencies.<sup>63-65</sup>

15       Exacerbated by the war on drugs and inequalities in enforcement and incarceration, some of

16       these groups are extremely hard to reach by traditional means. As Twitter features the default

17       anonymity, our pipeline can reach, for example, Black males, to create an unprecedentedly large

18       cohort and to understand their unique challenges and potential effective interventions.

19       Furthermore, our pipeline allows longitudinal study, in comparison to the cross-sectional nature

20       of NSDUH.

21

## 22 **Limitations**

1           Our work is largely limited by the data source. The demographic distribution of Twitter  
2 subscribers is different than the US population, especially for age groups. Though we adjust the  
3 estimates according to the 2018 Twitter survey, it only partially solves the issue. For people who  
4 do not use Twitter or do not discuss their non-medical use, we have no other means to collect  
5 their data and, thus, there may be groups of people that are not represented in our analyses. Also,  
6 social media data is noisy and may contain false information (*eg.*, fake races or genders), which  
7 we have no alternative approach to verify. Additionally, tweets are short (limited by 280  
8 characters) and consist of colloquial words, posing significant limitations on the classifier  
9 development, and thus the performance of the overall pipeline.

10           Our proposed methods *cannot* and *should not* be used for identifying the age group, race  
11 or gender identity of individual Twitter subscribers. Their performances are not perfect so there  
12 is no guarantee that they will not incorrectly characterize a single subscriber. Our methods  
13 essentially *estimate* the distributions of demographic information within a given *cohort*. Due to  
14 the large size of the cohort, we anticipate that the small numbers of incorrect characterizations  
15 are eclipsed.

16           Finally, an important limitation of our current work is that it excludes certain segments of  
17 the population who have been underrepresented in past research—such as those with non-binary  
18 gender identities. Our current work could not achieve such level of granularity due to lack of  
19 available data. The promising findings from our current work is the first step and our planned  
20 future work includes improving the inclusive of our analysis (*eg.*, by inclusion of non-binary  
21 population, the uninsured, and those unreachable via traditional means).

22

## 23 **METHODS**

1

## 2 **Twitter nonmedical use cohort**

3           Twitter data was collected through a data processing pipeline that we have described in  
4 past research.<sup>34,35</sup> The components of the pipeline include collecting publicly available streaming  
5 data about prescription medications, classifying the data, and then retrospectively collecting the  
6 metadata of subscribers who are detected to self-report NPMU. We collected English tweets  
7 mentioning at least one of over 20 PMs (including generic and trade names, street names, and  
8 common misspellings)<sup>66</sup> that have the potential for nonmedical use (see Supplement Table S4).<sup>34</sup>  
9 We developed annotation guidelines with our domain expert (JP), and annotated a subset  
10 consisting of 16,443 tweets into four categories: *non-medical use*, *consumption*, *information*, or  
11 *non-relevant* (see Supplement Table S5 and S6 for annotation details).<sup>34</sup> We used the annotated  
12 data to train machine-learning classifiers, and the best performing classifier (based-on  
13 RoBERTa-Large, with an accuracy of 82.3%) was deployed in our data collection pipeline. Our  
14 cohort consists of tweets that were signaled as *non-medical use* by the machine-learning  
15 classifier.

16

## 17 **Gender distribution estimates**

18           The genders of the Twitter subscribers were estimated based on the meta-1 classifier in  
19 our prior work.<sup>60</sup> The genders are treated in a binary framework (excluding those with non-  
20 binary gender identities due to lack of available data) and should be interpreted as how users  
21 represent themselves online and thus are closer to the user's gender identities than biological  
22 sexes. We developed the classifier based on gender-labeled datasets made available by Liu &  
23 Ruths<sup>41</sup> and Volkova et al.<sup>67</sup> In total, we were able to retrieve the meta-data of 67,181



1 subscribers, consisting of 35,812 (53.3%) females and 31,369 (46.7%) males, which we used to  
2 develop the pipeline. We validated the performance on people who use prescription medications  
3 nonmedically from Twitter on a set of 412 subscribers whose genders were identified using their  
4 linked, public Facebook profiles. The classifier achieved accuracy of 94.4% (95%-Confidence  
5 Intervals: 92.0%-96.6%).

6

### 7 **Age distribution estimates**

8 A rule-based approach, which searches for text patterns that are self-disclosures about the  
9 subscribers' ages, was employed to detect age-groups. Further rules are applied to resolve  
10 inconsistencies among detected information. The pattern matching is done using regular  
11 expressions. Sample text patterns include “(\d\d) birthday to me” or “i'm (\d\d)” where “\d”  
12 denotes digits (0-9). We also constructed a filter to remove irrelevant statements that are not  
13 associated with ages, such as “I'm 20 weeks pregnant.” The pipeline was developed based on a  
14 set of 2,000 subscribers, among which 1,540 tweets from 609 subscribers matched with the text  
15 patterns and were annotated. The annotation agreement based on an overlapping set of 346  
16 subscribers (952 tweets) is 89.3% with Cohen's  $K=0.89$  (95.8% with Cohen's  $K=0.96$  on tweets).  
17 The test accuracy is 0.88 (95% CI: 0.84-0.92) [0.90 (95% CI: 0.86-0.94) if allowed 1 year age  
18 discrepancy] on the subscribers who have matched tweets (referred as precision in the text) and  
19 0.93 (95% CI: 0.90-0.95) on the matched tweets.

20

### 21 **Race distribution estimates**

22 The race estimation module is similar to the age estimation one and applies rules and  
23 patterns. Relevant expressions indicating race are searched using regular expressions. Example

1 text patterns include “*i’m (black)*” or “*i’m (white)*”. We also constructed a filter to remove  
2 irrelevant statements such as “*I’m black salmon.*” The pipeline was developed based on a set of  
3 4,000 tweets, among which 1,124 tweets from 578 subscribers matched with the text patterns and  
4 were annotated. The annotation agreement based on an overlapping set of 293 subscribers (533  
5 tweets) is 87.7% with Cohen’s  $K=0.78$  (94.0% with Cohen’s  $K=0.88$  on tweets). The test  
6 accuracy of the pipeline is 0.90 (95% CI: 0.85-0.95) on the subscribers who have matched tweets  
7 (referred as precision in the text) [0.94 (95% CI: 0.91-0.97) on the matched tweets]. Since the  
8 pipeline is not designed for subscribers who have more than 1 race (“more” in the 2018 Twitter  
9 Survey<sup>68</sup> and the NSDUH<sup>18</sup>), we did not include those subscribers when reporting the results and  
10 comparing with the references.

#### 11 **Baseline for Twitter subscribers’ demographics**

12 We established the baseline for the US Twitter subscribers based on the 2018 Twitter  
13 Survey conducted by Pew Research Center.<sup>68</sup> We focused on Twitter subscribers who have at  
14 least used Twitter once a week and calculated the proportions of the targeted demographics (eg.,  
15 age, gender, race) among these subscribers. We noted that the survey was conducted only among  
16 the adults (aged 18+) and the Asian, AIAN, and NHOPI races were grouped into “others.”

#### 17 **Baseline for Twitter age and race characterization**

18 We established the baseline for Twitter Age and Race characterization based on a data set  
19 of 156,368 general Twitter subscribers. We first collected streaming tweets with stopwords in  
20 nltk package as keywords on Aug 27, 2021. Data collection using stopwords is an attempt to  
21 collect a random set of Twitter subscribers. We then collected the subscribers’ metadata and  
22 applied the age and race pipelines. Our objective was to calibrate our pipeline by estimating how  
23 many of the subscribers within certain age or race groups actually self-disclosed their age and

1 race on Twitter and were captured by our pipeline. The estimated rates (of detection/self-  
2 disclosure) were then used as weights to normalize the age/race proportions obtained from the  
3 NPMU dataset.

#### 4 **Calculation of the proportions for PWUS based on Twitter data**

5 For each subscriber's characteristics (gender, age, and race), we used the number of  
6 Twitter subscribers in each category, inferred by the corresponding pipelines, to estimate the  
7 proportion. For age and race, this calculation was limited on the Twitter subscribers whose age  
8 or race can be inferred. The proportions were further normalized via the rates of detection as:

$$9 \quad \textit{normalized proportion} = \frac{1}{\textit{normalization factor}} \times \frac{(\textit{raw proportion})}{(\textit{rate of detection})}$$

10 For example, if 10% of the Twitter subscribers are black and only 1% of the random Twitter  
11 subscribers disclose that they are black, then we can estimate that roughly 1 in 10 black Twitter  
12 subscribers self-disclose their race (rate of detection). Then if we captured 100 people who  
13 reported the nonmedical use of stimulants and disclosed that they are black, we infer that roughly  
14 1000 people who use stimulants captured in our pipeline are black and use this number to  
15 calculate the normalized proportion. For the race proportions, since AIAN, NHOPI, and Asian  
16 were combined into "others" category in the 2018 Twitter Survey, we only calibrated for the  
17 "others" category as a whole and assumed that their relative proportions are the same as obtained  
18 using the race characterization pipeline.

19

#### 20 **Calculation of proportions for the NSDUH data**

21 We established the baseline for the NSDUH/US Census based on Table 12.1A (age) and  
22 Table 12.2A (gender and race) in the 2019 NSDUH. We calculated the gender, age, and race  
23 proportion through the estimated "Numbers in Thousands." For age, we used the "Total (2019)"

1 column. For the age and gender, we used the “Age 12+ (2019)” column. For the gender and race  
2 for people who report NPMU, we again used the “Age 12+ (2019)” column on Table 1.47A  
3 (stimulants), Table 1.53A (tranquilizers), and table 1.44A (pain relievers). For age, we used the  
4 “Misuse in the past Year (2019)” column on Table 1.14A (stimulants), Table 1.16A  
5 (tranquilizers), and table 1.13A (pain relievers).

6

### 7 **Calculation of proportions for the NEDS data**

8 We calculated the gender proportion of the EDV by using the “No.” column for all opioid  
9 poisoning on the Supplemental Table 2C in the Annual Surveillance Report of Drug-related  
10 Risks and Outcomes.<sup>51</sup> The weighted estimates provided in the table are from the NEDS 2016.

11

### 12 **Estimation of 95% confidence intervals**

13 For the Twitter data and the test performance of the pipelines, the confidence intervals  
14 are estimated via bootstrapping. For the NSDUH and NEDS data, the 95% confidence intervals  
15 are estimated using simulation. For each category, we approximate the distribution as a normal  
16 distribution with the reported number as mean and the standard error as standard deviation. We  
17 then repeatedly sampled the joint distribution for all the categories in the targeted demographics  
18 and calculated the proportions, assuming each category is independent of each other. The 95%  
19 confidence intervals were then constructed using the 0.025 quantile and the 0.975 quantile within  
20 the list of proportions of the given category. For the NEDS data, the standard errors for estimated  
21 numbers (“No.” column) were estimated through the standard errors (“SE” column) for rates

22 (“Rate” column) as  $(\text{No.}) \times \frac{(\text{SE})}{(\text{Rate})}$ .

23

1 **FUNDING**

2 Research reported in this publication was supported by the National Institute on Drug  
3 Abuse (NIDA) of the National Institutes of Health (NIH) under award number  
4 R01DA046619. The content is solely the responsibility of the authors and does not  
5 necessarily represent the official views of the NIH.

6 **AUTHOR CONTRIBUTIONS**

7 YY conducted and directed the machine learning experiments, evaluations and data  
8 analyses, with assistance from MAA and AS. AS provided supervision for full study. JSL,  
9 HLFC, and JP provided toxicology and public health domain expertise for interpreting  
10 the results. YY drafted the manuscript and all authors contributed to the final  
11 manuscript.

12 **CONFLICT OF INTEREST**

13 None declared

14 **ACKNOWLEDGEMENTS**

15 The authors thank the support from the National Institute of Health and National  
16 Institute of Drug Abuse.

17 **DATA AVAILABILITY**

18 The datasets generated during and/or analysed during the current study are not publicly  
19 available due to Twitter API's use terms and privacy concern but are available from the  
20 corresponding author on reasonable request.

21

22 **REFERENCES**

- 1 1 Jalal, H. *et al.* Changing dynamics of the drug overdose epidemic in the United States from 1979  
2 through 2016. *Science* **361** (2018).
- 3 2 Mattson, C. L. *et al.* Trends and geographic patterns in drug and synthetic opioid overdose  
4 deaths—United States, 2013–2019. *Morbidity and Mortality Weekly Report* **70**, 202 (2021).
- 5 3 Vital Statistics Rapid Release. *Provisional Drug Overdose Death Counts*,  
6 <<https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm>> (2022).
- 7 4 *Biden-Harris Administration Calls for Historic Levels of Funding to Prevent and Treat Addiction*  
8 *and Overdose | The White House*, <[https://www.whitehouse.gov/ondcp/briefing-](https://www.whitehouse.gov/ondcp/briefing-room/2021/05/28/biden-harris-administration-calls-for-historic-levels-of-funding-to-prevent-and-treat-addiction-and-overdose/)  
9 [room/2021/05/28/biden-harris-administration-calls-for-historic-levels-of-funding-to-prevent-](https://www.whitehouse.gov/ondcp/briefing-room/2021/05/28/biden-harris-administration-calls-for-historic-levels-of-funding-to-prevent-and-treat-addiction-and-overdose/)  
10 [and-treat-addiction-and-overdose/](https://www.whitehouse.gov/ondcp/briefing-room/2021/05/28/biden-harris-administration-calls-for-historic-levels-of-funding-to-prevent-and-treat-addiction-and-overdose/)> (
- 11 5 Benjamin Lê Cook & Alegría, M. Racial-Ethnic Disparities in Substance Abuse Treatment: The  
12 Role of Criminal History and Socioeconomic Status. *Psychiatric Services* **62**, 1273-1281,  
13 doi:10.1176/ps.62.11.pss6211\_1273 (2011).
- 14 6 Acevedo, A. *et al.* Disparities in the treatment of substance use disorders: does where you live  
15 matter? *The journal of behavioral health services & research* **45**, 533-549 (2018).
- 16 7 Mennis, J. & Stahler, G. J. Racial and ethnic disparities in outpatient substance use disorder  
17 treatment episode completion for different substances. *Journal of substance abuse treatment*  
18 **63**, 25-33 (2016).
- 19 8 Burlew, K., McCuistian, C. & Szapocznik, J. Racial/ethnic equity in substance use treatment  
20 research: the way forward. *Addiction Science & Clinical Practice* **16**, 1-6 (2021).
- 21 9 National Academies of Sciences, Engineering & Medicine. *Communities in Action: Pathways to*  
22 *Health Equity*. (The National Academies Press, 2017).
- 23 10 Nan Marie Astone, Steven Martin & Aron, L. Death Rates for US Women Ages 15 to 54. (2015).
- 24 11 Lagisetty, P. A., Ross, R., Bohnert, A., Clay, M. & Maust, D. T. Buprenorphine treatment divide by  
25 race/ethnicity and payment. *JAMA psychiatry* **76**, 979-981 (2019).
- 26 12 Hansen, H., Siegel, C., Wanderling, J. & DiRocco, D. Buprenorphine and methadone treatment  
27 for opioid dependence by income, ethnicity and race of neighborhoods in New York City. *Drug*  
28 *and alcohol dependence* **164**, 14-21 (2016).
- 29 13 Substance Abuse and Mental Health Services Administration. The Opioid Crisis and the  
30 Black/African American Population: An Urgent Issue. (Office of Behavioral Health Equity.  
31 Substance Abuse and Mental Health Services Administration, 2020).
- 32 14 Substance Abuse and Mental Health Services Administration. The Opioid Crisis and the  
33 Hispanic/Latino Population: an Urgent Issue. (Office of Behavioral Health Equity. Substance  
34 Abuse and Mental Health Services Administration, 2020).
- 35 15 Hadland, S. E. *et al.* Trends in receipt of buprenorphine and naltrexone for opioid use disorder  
36 among adolescents and young adults, 2001-2014. *JAMA pediatrics* **171**, 747-755 (2017).
- 37 16 Hedegaard, H., Miniño, A. M., Spencer, M. R. & Warner, M. Drug Overdose Deaths in the United  
38 States, 1999–2020. *NCHS Data Brief* (2021).
- 39 17 Townsend, T. *et al.* Racial/ethnic and geographic trends in combined stimulant/opioid  
40 overdoses, 2007-2019. *American Journal of Epidemiology* (2022).
- 41 18 Substance Abuse and Mental Health Services Administration. Results from the 2019 National  
42 Survey on Drug Use and Health: Detailed tables. (2020).
- 43 19 U.S. Department of Health and Human Services Office of Minority Health. *Minority Population*  
44 *Profiles*, <<https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=2&lvlID=26>> (
- 45 20 Kolodny, A. & Frieden, T. R. Ten steps the federal government should take now to reverse the  
46 opioid addiction epidemic. *JAMA* **318**, 1537-1538 (2017).
- 47 21 Centers for Disease Control and Prevention. Data Modernization Initiative Strategic  
48 Implementation Plan. (2021).

- 1 22 Substance Abuse and Mental Health Services Administration. Key substance use and mental  
2 health indicators in the United States: results from the 2018 National Survey on Drug Use and  
3 Health. (2019).
- 4 23 Gummin, D. D. *et al.* 2019 Annual report of the American Association of poison control centers'  
5 National Poison Data System (NPDS): 37th annual report. *Clinical Toxicology* **58**, 1360-1541  
6 (2020).
- 7 24 Slavova, S., Bunn, T. L. & Talbert, J. Drug overdose surveillance using hospital discharge data.  
8 *Public Health Reports* **129**, 437-445 (2014).
- 9 25 Slavova, S., Rock, P., Bush, H. M., Quesinberry, D. & Walsh, S. L. Signal of increased opioid  
10 overdose during COVID-19 from emergency medical services data. *Drug and alcohol dependence*  
11 **214**, 108176 (2020).
- 12 26 Clifford, C., Sethi, M., Cox, D. & Manini, A. F. First-Line Vasopressor and Mortality Rates in ED  
13 Patients with Acute Drug Overdose. *Journal of Medical Toxicology* **17**, 1-9 (2021).
- 14 27 Hedegaard H, M. A., Warner M Drug Overdose Deaths in the United States, 1999-2018., (Centers  
15 for Disease Control and Prevention, 2020).
- 16 28 Volkow, N. D. Collision of the COVID-19 and Addiction Epidemics. *Annals of Internal Medicine*  
17 **173**, 61-62, doi:10.7326/M20-1212 (2020).
- 18 29 Lopez, A. M. *et al.* Co-use of methamphetamine and opioids among people in treatment in  
19 Oregon: A qualitative examination of interrelated structural, community, and individual-level  
20 factors. *International Journal of Drug Policy* **91**, 103098 (2021).
- 21 30 Ciccarone, D. The rise of illicit fentanyls, stimulants and the fourth wave of the opioid overdose  
22 crisis. *Current Opinion in Psychiatry* **34**, 344-350 (2021).
- 23 31 Sarker, A. *et al.* Social media mining for toxicovigilance: automatic monitoring of prescription  
24 medication abuse from Twitter. *Drug safety* **39**, 231-240 (2016).
- 25 32 O'Connor, K. *et al.* Pharmacovigilance on twitter? Mining tweets for adverse drug reactions.  
26 *AMIA annual symposium proceedings* **2014**, 924 (2014).
- 27 33 Amir, S., Dredze, M. & Ayers, J. W. in *Proceedings of the Sixth Workshop on Computational*  
28 *Linguistics and Clinical Psychology*. 114-120.
- 29 34 O'Connor, K., Sarker, A., Perrone, J. & Hernandez, G. G. Promoting Reproducible Research for  
30 Characterizing Nonmedical Use of Medications Through Data Annotation: Description of a  
31 Twitter Corpus and Guidelines. *Journal of Medical Internet Research* **22**, e15861 (2020).
- 32 35 Al-Garadi, M. A. *et al.* Text classification models for the automatic detection of nonmedical  
33 prescription medication use from social media. *BMC Medical Informatics and Decision Making*  
34 **21**, 27, doi:10.1186/s12911-021-01394-0 (2021).
- 35 36 Chary, M. *et al.* Epidemiology from tweets: estimating misuse of prescription opioids in the USA  
36 from social media. *Journal of Medical Toxicology* **13**, 278-286 (2017).
- 37 37 Graves, R. L. *et al.* Opioid discussion in the Twittersphere. *Substance use & misuse* **53**, 2132-  
38 2139 (2018).
- 39 38 Sarker, A., Gonzalez-Hernandez, G., Ruan, Y. & Perrone, J. Machine Learning and Natural  
40 Language Processing for Geolocation-Centric Monitoring and Characterization of Opioid-Related  
41 Social Media Chatter. *JAMA Netw Open* **2**, e1914672,  
42 doi:10.1001/jamanetworkopen.2019.14672 (2019).
- 43 39 Centers for Disease Control and Prevention. *CDC Wonder online databases*,  
44 <<https://wonder.cdc.gov/>> (  
45 40 Burger, J. D., Henderson, J., Kim, G. & Zarrella, G. in *Proceedings of the conference on empirical*  
46 *methods in natural language processing*. 1301-1309 (Association for Computational Linguistics).
- 47 41 Liu, W. & Ruths, D. in *2013 AAAI Spring Symposium Series*.

- 1 42 Sap, M. *et al.* in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language*  
2 *Processing (EMNLP)*. 1146-1151.
- 3 43 Wang, Z. *et al.* Demographic inference and representative population estimates from  
4 multilingual social media data. *The World Wide Web Conference*, 2056-2067 (2019).
- 5 44 Morgan-Lopez, A. A., Kim, A. E., Chew, R. F. & Ruddle, P. Predicting age groups of Twitter users  
6 based on language and metadata features. *PloS one* **12**, e0183537 (2017).
- 7 45 Garcia-Guzman, R. *et al.* Trend-based categories recommendations and age-gender prediction  
8 for pinterest and twitter users. *Applied Sciences* **10**, 5957 (2020).
- 9 46 Pandya, A., Oussalah, M., Monachesi, P. & Kostakos, P. On the use of distributed semantics of  
10 tweet metadata for user age prediction. *Future Generation Computer Systems* **102**, 437-452  
11 (2020).
- 12 47 Preoțiu-Pietro, D. & Ungar, L. User-level race and ethnicity predictors from twitter text.  
13 *Proceedings of the 27th International Conference on Computational Linguistics*, 1534-1545  
14 (2018).
- 15 48 Ye, J. & Skiena, S. The secret lives of names? name embeddings from social media. *Proceedings*  
16 *of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*,  
17 3000-3008 (2019).
- 18 49 Xu, S. *et al.* Leveraging social media to promote public health knowledge: example of cancer  
19 awareness via Twitter. *JMIR public health and surveillance* **2**, e17 (2016).
- 20 50 Jung, S.-G., An, J., Kwak, H., Salminen, J. & Jansen, B. J. Assessing the accuracy of four popular  
21 face recognition tools for inferring gender, age, and race. *Twelfth international AAAI conference*  
22 *on web and social media* (2018).
- 23 51 Centers for Disease Control and Prevention. Annual Surveillance Report of Drug-Related Risks  
24 and Outcomes — United States Surveillance Special Report. (2019).
- 25 52 Wojcik, S. & Hughes, A. *Sizing Up Twitter Users*,  
26 <<https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>> (2019).
- 27 53 United States Census Bureau. *Census Bureau Releases Estimates of Undercount and Overcount in*  
28 *the 2020 Census*, <[https://www.census.gov/newsroom/press-releases/2022/2020-census-](https://www.census.gov/newsroom/press-releases/2022/2020-census-estimates-of-undercount-and-overcount.html)  
29 [estimates-of-undercount-and-overcount.html](https://www.census.gov/newsroom/press-releases/2022/2020-census-estimates-of-undercount-and-overcount.html)> (2022).
- 30 54 Substance Abuse and Mental Health Services Administration. *National Survey of Drug Use and*  
31 *Health (NSDUH)-Study Notes*, <[https://www.datafiles.samhsa.gov/dataset/national-survey-drug-](https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2020-nsduh-2020-ds0001)  
32 [use-and-health-2020-nsduh-2020-ds0001](https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2020-nsduh-2020-ds0001)> (2021).
- 33 55 Fendrich, M., Johnson, T. P., Wislar, J. S., Hubbell, A. & Spiehler, V. The utility of drug testing in  
34 epidemiological research: results from a general population survey. *Addiction* **99**, 197-208,  
35 doi:<https://doi.org/10.1111/j.1360-0443.2003.00632.x> (2004).
- 36 56 Fendrich, M. & Johnson, T. P. Race/ethnicity differences in the validity of self-reported drug use:  
37 Results from a household survey. *Journal of Urban Health* **82**, iii67-iii81,  
38 doi:10.1093/jurban/jti065 (2005).
- 39 57 Johnson, T. & Fendrich, M. Modeling sources of self-report bias in a survey of drug use  
40 epidemiology. *Ann Epidemiol* **15**, 381-389, doi:10.1016/j.annepidem.2004.09.004 (2005).
- 41 58 Bromberg, J. R. *et al.* Methodology and Demographics of a Brief Adolescent Alcohol Screen  
42 Validation Study. *Pediatr Emerg Care* **35**, 737-744, doi:10.1097/pec.0000000000001221 (2019).
- 43 59 Percy, A., McAlister, S., Higgins, K., McCrystal, P. & Thornton, M. Response consistency in young  
44 adolescents' drug use self-reports: a recanting rate analysis. *Addiction* **100**, 189-196,  
45 doi:<https://doi.org/10.1111/j.1360-0443.2004.00943.x> (2005).
- 46 60 Yang, Y.-C., Al-Garadi, M. A., Love, J. S., Perrone, J. & Sarker, A. Automatic gender detection in  
47 Twitter profiles for health-related cohort studies. *JAMIA Open* **4**, ooab042 (2021).



- 1 61 Friedman, J. R. & Hansen, H. Evaluation of Increases in Drug Overdose Mortality Rates in the US  
2 by Race and Ethnicity Before and During the COVID-19 Pandemic. *JAMA Psychiatry*,  
3 doi:10.1001/jamapsychiatry.2022.0004 (2022).
- 4 62 Papazisis, G., Tsakiridis, I. & Sifis, S. Nonmedical use of prescription drugs among medical  
5 students and the relationship with illicit drug, tobacco, and alcohol use. *Substance abuse:  
6 research and treatment* **12**, 1178221818802298 (2018).
- 7 63 Substance Abuse and Mental Health Services Administration. *Behavioral Health Equity*,  
8 <<https://www.samhsa.gov/behavioral-health-equity/policy>> (2022).
- 9 64 National Institute of Health. *Inclusion of Women and Minorities as Participants in Research  
10 Involving Human Subjects*, <<https://grants.nih.gov/policy/inclusion/women-and-minorities.htm>>  
11 (2022).
- 12 65 National Institute on Minority Health and Health Disparities. *NIH Minority Health and Health  
13 Disparities Strategic Plan 2021-2025*, <[https://www.nimhd.nih.gov/about/strategic-plan/nih-  
14 strategic-plan-directors-foreword.html](https://www.nimhd.nih.gov/about/strategic-plan/nih-strategic-plan-directors-foreword.html)> (2021).
- 15 66 Sarker, A. & Gonzalez-Hernandez, G. An unsupervised and customizable misspelling generator  
16 for mining noisy health-related text sources. *J Biomed Inform* **88**, 98-107,  
17 doi:10.1016/j.jbi.2018.11.007 (2018).
- 18 67 Volkova, S., Wilson, T. & Yarowsky, D. in *Proceedings of the 2013 Conference on Empirical  
19 Methods in Natural Language Processing*. 1815-1827 (Association for Computational  
20 Linguistics).
- 21 68 *2018 Twitter Survey*, <<https://www.pewresearch.org/politics/dataset/2018-twitter-survey/>> (  
22