

Viewpoint

Challenging the Continued Usefulness of Social Media Recruitment for Surveys of Hidden Populations of People Who Use Opioids

Elizabeth D Nesoff¹, MPH, PhD; Joseph J Palamar², PhD; Qingyue Li³, MSSP; Wenqian Li³, MSSP; Silvia S Martins⁴, MD, PhD

¹Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

²Department of Population Health, NYU Langone Health, New York, NY, United States

³University of Pennsylvania School of Social Policy and Practice, Philadelphia, PA, United States

⁴Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, United States

Corresponding Author:

Elizabeth D Nesoff, MPH, PhD

Department of Biostatistics, Epidemiology, and Informatics

University of Pennsylvania Perelman School of Medicine

423 Guardian Dr

Philadelphia, PA, 19104

United States

Phone: 1 215 662 2560

Fax: 1 215 573 4865

Email: elizabeth.nesoff@pennmedicine.upenn.edu

Abstract

Historically, recruiting research participants through social media facilitated access to people who use opioids, capturing a range of drug use behaviors. The current rapidly changing online landscape, however, casts doubt on social media's continued usefulness for study recruitment. In this viewpoint paper, we assessed social media recruitment for people who use opioids and described challenges and potential solutions for effective recruitment. As part of a study on barriers to harm reduction health services, we recruited people who use opioids in New York City to complete a REDCap (Research Electronic Data Capture; Vanderbilt University) internet-based survey using Meta (Facebook and Instagram), X (formerly known as Twitter), Reddit, and Discord. Eligible participants must have reported using opioids (heroin, prescription opioids, or fentanyl) for nonprescription purposes in the past 90 days and live or work in New York City. Data collection took place from August 2023 to November 2023. Including study purpose, compensation, and inclusion criteria caused Meta's social media platforms and X to flag our ads as "discriminatory" and "spreading false information." Listing incentives increased bot traffic across all platforms despite bot prevention activities (eg, reCAPTCHA and counting items in an image). We instituted a rigorous post hoc data cleaning protocol (eg, investigating duplicate IP addresses, participants reporting use of a fictitious drug, invalid ZIP codes, and improbable drug use behaviors) to identify bot submissions and repeat participants. Participants received a US \$20 gift card if still deemed eligible after post hoc data inspection. There were 2560 submissions, 93.2% (n=2387) of which were determined to be from bots or malicious responders. Of these, 23.9% (n=571) showed evidence of a duplicate IP or email address, 45.9% (n=1095) reported consuming a fictitious drug, 15.8% (n=378) provided an invalid ZIP code, and 9.4% (n=225) reported improbable drug use behaviors. The majority of responses deemed legitimate (n=173) were collected from Meta (n=79, 45.7%) and Reddit (n=48, 27.8%). X's ads were the most expensive (US \$1.96/click) and yielded the fewest participants (3 completed surveys). Social media recruitment of hidden populations is challenging but not impossible. Rigorous data collection protocols and post hoc data inspection are necessary to ensure the validity of findings. These methods may counter previous best practices for researching stigmatized behaviors.

(*J Med Internet Res* 2025;27:e63687) doi: [10.2196/63687](https://doi.org/10.2196/63687)

KEYWORDS

opioids; substance use; survey methods; social media; recruitment; survey; drug overdose; substance use disorder; online recruitment; online survey; mental health; addiction; data collection

Introduction

Opioid-involved overdoses significantly increased across the United States through 2023 [1]. Drug overdose deaths more than doubled from 2015 to 2023—increasing from 53,356 to 107,543—and 78% of drug overdose deaths in 2023 involved opioids [2,3]. Since drug use is an illegal and heavily stigmatized behavior, most research conducted on nonprescription opioid use recruited people who use opioids through settings or organizations frequented by the target population, such as syringe exchange programs or substance use disorder treatment facilities [4,5]. However, these venue-based recruitment methods may introduce selection bias as certain racial or ethnic minorities and women may feel less comfortable in and around these spaces [6-10].

Internet-based recruitment may provide an alternative method for accessing certain hard-to-reach populations and stigmatized behaviors. The primary barrier to feasibility in internet-based recruitment is access to people who use opioids, given the highly stigmatized nature of drug use [4,5]. Many popular social media platforms [11-13] provide a sophisticated suite of tools to target advertisement placement to the feeds of specific audiences defined by demographic characteristics (eg, age and location) and contextual interests such as pages or groups to which subscribers indicated interest, keyword searches, and frequently used hashtags [11-15]. Different advertisement images could also be targeted to different groups based on race and ethnicity to increase participant diversity [16]. Systematic reviews found that, compared with traditional recruitment methods, social media recruitment had lower costs, shorter recruitment periods, and improved representation of populations with mental health and substance use problems [4,14,15]. Social media platforms have been shown to be effective for recruitment in studies of populations reporting addiction, mental health problems, and other hard-to-reach populations [11,12,14,15,17-19]. In addition, while some racial or ethnic minority groups and people of lower income have had less access to internet services historically [20], studies have also shown that social media recruitment effectively engages low-socioeconomic status and racial and ethnic minority participants [16,21-23].

However, the current, rapidly changing online landscape calls into question the continued usefulness of social media platforms for recruiting and surveying people who use opioids. In January 2022, Meta, which owns Facebook and Instagram, reduced their ad-targeting options for what it calls “sensitive” topics in an effort to improve user privacy [24]. Advertisers are no longer allowed to use custom keywords or hashtags to target subscribers but must fit their ads into prescribed categories within Meta’s ad management system. Twitter, now X, is in flux after the November 2022 acquisition by Elon Musk [25]. Starting in July 2023, X began limiting the number of tweets a user can read or send in a day depending on their subscription level, leading to decreased user engagement and subscriber loss [26,27]. Reddit is also somewhat volatile. As part of the leadup to Reddit’s March 2024 IPO, Reddit announced they will charge companies such as third-party web browsing apps that want to access its content using an application programming interface (API) [28]. The API fees were unpopular among Reddit users, who saw it

as the company prioritizing profit over the community’s preferences [28]. Reddit fora, known as subreddits, are run by volunteer, unpaid moderators who enforce subreddit standards of conduct (eg, no posts about drug selling). Many subreddits shut down in protest for 2 weeks in June 2023, and in response, Reddit announced it intends to reduce the power of moderators [28,29]. These changes in user engagement and advertising options across social media platforms call into question the continued usability of social media for identifying and recruiting hard-to-reach populations for health research.

In addition, the proliferation of automated bots and malicious responders threatens internet-based survey research data integrity for incentivized surveys. Including a participation incentive increases the likelihood of attracting bots and malicious responses—a consideration that must be weighed against the effectiveness of incentives in increasing participation among the target population [30-32]. Human respondents easily create bots that bypass survey platforms’ safety protocols and data protection mechanisms or use virtual private servers to repeatedly complete surveys for financial gain [32,33]. While bots are thought to be the most prolific threat to data integrity, human threats to data validity for incentivized surveys are becoming more common and increasingly difficult to detect [31]. Individual fraudsters may falsify responses to mimic what they perceive as desirable survey answers in order to receive survey incentives [31,34]. “Click farms” groups that purposefully search for surveys with financial incentives may also contribute to invalid responses [31]. Scammers may take surveys multiple times, slightly altering their answers to appear like unique participants [31,34,35]. Even respondents who are from the target population may provide distorted or misleading survey responses if answers are rushed or careless or if they take a survey multiple times to receive a larger amount of the financial incentive [31,36].

The purpose of this Viewpoint is to share our experiences recruiting a hard-to-reach population through social media. As part of a study on barriers to engagement with harm reduction health services such as syringe exchange, we aimed to recruit people who use opioids in New York City to an internet-based survey. Eligible participants reported past-3-month use of opioids (including heroin, prescription opioids, fentanyl, and newer and emerging opioids like U-47700 and nitazenes) for nonprescription purposes (eg, to get “high”). Because the goal of our study was to investigate barriers to harm reduction engagement, we could not directly recruit from harm reduction health services organizations—a traditional point of access for studies of people who use opioids—as these people who use opioids would presumably experience lower engagement barriers. Recruiting participants through social media previously facilitated access to a group of people who use opioids traditionally not included in research and captured a greater range of drug use behaviors [4,15]. In this paper, we assess the continued usefulness of social media recruitment for this hidden population and describe challenges and potential solutions for effective recruitment across social media platforms.

Methods

Participant Recruitment

We recruited participants across 4 social media platforms: Meta (Facebook and Instagram), X (Twitter), Reddit, and Discord. Unless otherwise indicated, we tracked ad engagement (impressions and link clicks) and cost per click. Specific activities and challenges for each platform are described below, including the ad purchasing and approval processes.

Meta

Advertisements on Facebook and Instagram are distributed concurrently through Meta's ad management platform. The user sets a campaign goal—in our case, increasing website traffic measured by link clicks—a daily spending limit and campaign duration. Meta's algorithm for how cost per click is calculated lacks transparency, and it was not clear if link clicks consistently translated to completed surveys.

As previously stated, Meta no longer supports the use of custom keywords or hashtags to target subscribers. Meta's prescribed ad categories did not provide options related to drug use, and ads with images of drug paraphernalia, drug use, or smoking

were all flagged and removed. Therefore, we were limited to targeting a geographic area (New York City Designated Market Area in this study) and basic demographics (age and sex only) for the first round of survey recruitment. Our first attempt at ad placement received over 1000 automated responses in the first hour. We quickly reformatted our ads, removing all information about survey compensation. Our second attempt was flagged by Meta as "discriminatory" because we included eligibility criteria in the ad's text description. We removed this as well and successfully launched the first ad (Figure 1). We used 3 different ad images; we started with an image of a woman as one of the parent study research goals is to understand harm reduction engagement barriers for women. As the majority of Meta respondents were men, we used 2 additional images of women of varying races and ethnicities in an attempt to increase female participants, as men did not seem dissuaded by images of women. All images were obtained from stock photography websites offering copyright- and royalty-free images (eg, Unsplash).

In addition, we contacted several Facebook groups related to substance use and posted about the survey; once we received permission from moderators to join, all of our posts were removed by moderators and yielded no participants.

Figure 1. Example of an approved institutional review board (IRB) advertisement versus an approved Meta advertisement. NYC: New York City.



X (Formerly Twitter)

Purchasing ads on X is similar to Meta's platform, where the advertiser sets a campaign goal (website traffic and conversions), a daily spending limit, and a campaign duration. Ads can also be targeted to specific demographics and geographic locations. Unlike Meta, X allows ad targeting using common keywords and hashtags. Targeted keywords included "opioid," "addiction," "recovery," "opiates," "overdose," and "harm reduction," as well as options for specific opioids (eg, oxycodone). However, we found that the available keywords were in flux, with some keyword options disappearing in between ad purchases. There appeared to be even less transparency in X's algorithm regarding

how cost per click is calculated compared with Meta, as the count of link clicks appeared to update haphazardly while Meta updated in real time.

As with Meta, our ads on X did not contain any mention of compensation or inclusion criteria (Figure 2). Our first attempt at ad placement was flagged by X as "ineligible" because of an unspecified policy violation, and our account was banned from purchasing ads. We appealed this decision by emailing ad support and explaining that the campaign was for an academic study. A customer service bot replied that our ad violated "unacceptable business practices" by promoting "misleading, false, or unsubstantiated claims." We replied to the email

explaining the study purpose; ad support replied that they had manually reviewed the ad, they would not review the ad again, and we were permanently banned from purchasing ads.

Approximately 12 hours later, we received another email from ad support alerting us they had “resolved all issues,” and our ad campaign began.

Figure 2. Example of an advertisement approved by X (formerly Twitter). NYC: New York City.

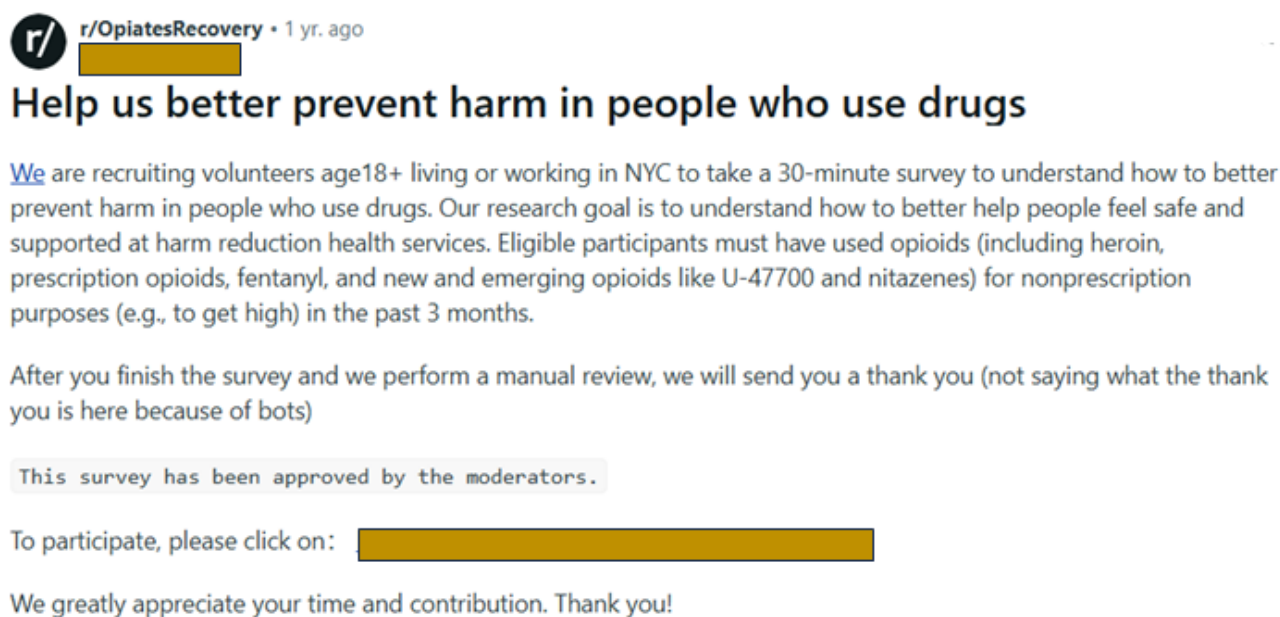


Reddit

Due to internal guidelines, our institution did not support purchasing ads on Reddit. Instead, we identified subreddits focused on topics and populations related to our study objectives (eg, [/r/OpiatesRecovery](#)) [37] and contacted moderators directly to ask for permission to recruit on the subreddit. Many subreddits explicitly banned academic study recruitment in the subreddit rules, and we did not contact those subreddits. We contacted 50 subreddits and received permission from 16 (N=4,048,000 total members; memberships are not mutually

exclusive). Reddit provides engagement metrics for all user posts, so we tracked when engagement with a post waned—on average, after a week. We posted on most subreddits weekly during the recruitment period, except for 1 subreddit that specifically requested we not post more often than monthly. Study staff posted the advertisements, and they were marked, “This survey has been approved by the moderators” (Figure 3). Many substance use–related subreddits do not allow images to be posted, so subreddit posts did not include ad images. No subreddits charged for posts.

Figure 3. Example of a Reddit post approved by subreddit moderators. NYC: New York City.



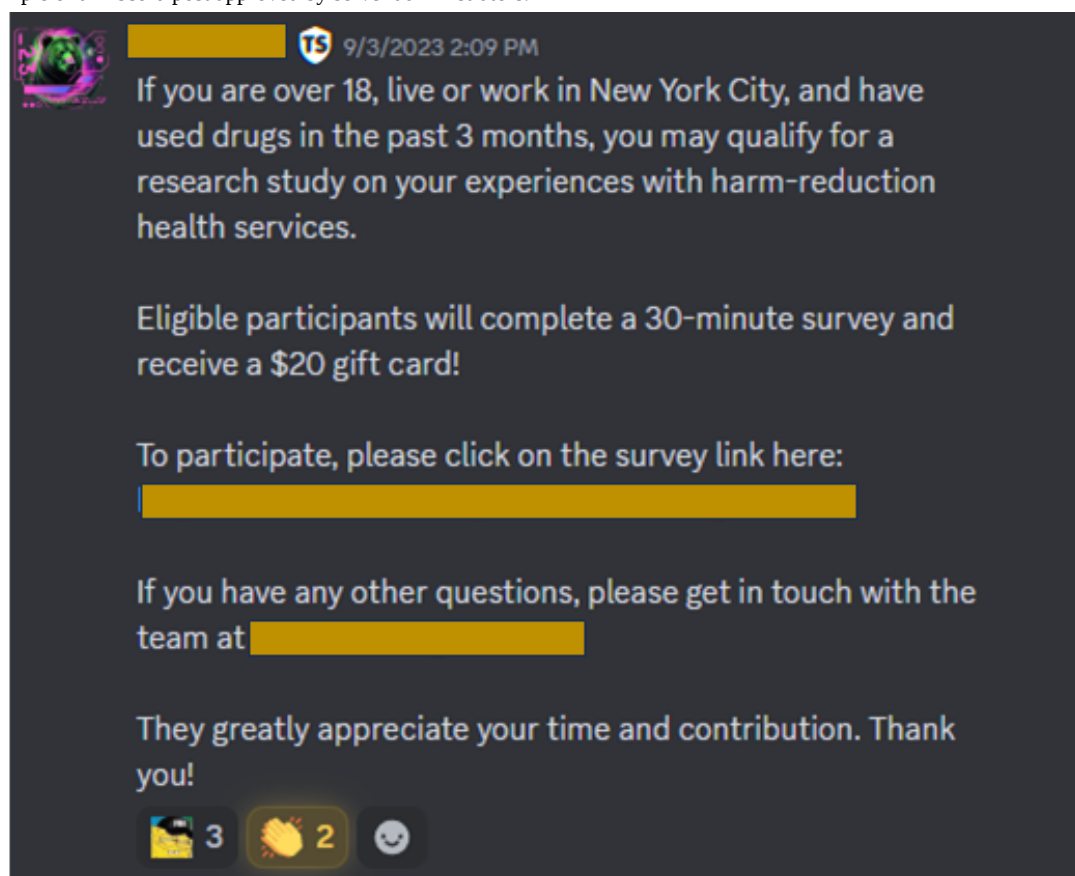
Discord

Launched in 2015, Discord allows users to create private or public communities called servers where members can chat in real time; servers can be subdivided into topic-based channels, which resemble message boards or subreddits. As with Reddit, Discord's servers are monitored by administrators who enforce the server's code of conduct. However, because Discord is not a centralized system but rather a series of servers created by individual administrators, there is less regulation from a central authority about inflammatory or illegal content such as drug selling. Individual administrators determine what is acceptable behavior on their server.

Currently, there is no mechanism to purchase ads on Discord. We used targeted keyword searches, including "recovery," "addiction," and "opioids," to identify Discord servers to contact for survey promotion. An additional server was suggested via

administrator recommendation. We avoided servers with active drug dealing. Of the 14 servers approached, administrators from 7 servers granted permission for survey dissemination. Administrators from these servers posted ads on our behalf as an announcement (Figure 4). Discord does not track user analytics for individual posts, but we recorded the number of server members for each community to estimate the potential reach of our posts (N=21,582 total members; memberships are not mutually exclusive). While some individual Discord servers may charge to approve advertisement posts, we did not encounter any.

Moderators from one of the Discord channels we approached (Bluelight.org) also invited us to post on their web forum community, a public website where members can discuss and share information on drug-related topics through posted messages.

Figure 4. Example of a Discord post approved by server administrators.

Survey Procedures

We conducted an internet-based survey on engagement with harm reduction health services (eg, syringe exchange). Participants must be ages 18 years and older, live or work in New York City (defined as the five boroughs: Manhattan, Bronx, Brooklyn, Queens, and Staten Island), and have used opioids (including heroin, nonprescription opioids, or fentanyl) for nonprescription purposes (eg, to get “high”) in the previous 90 days. Harm reduction health services availability varies widely across the United States; therefore, we chose to limit the geographic area to New York City to better understand engagement with existing services in the city. Recruiting a sample circumscribed by a geographic area may also provide locally relevant results for public health planning purposes. Respondents completed the self-directed questionnaire in REDCap (Research Electronic Data Capture; version 13.9.3; Vanderbilt University) at a location and time of their choosing. The survey was only available in English and was designed to take around 30 minutes to complete. Data collection took place from August 2023 to November 2023.

After clicking on an advertisement, participants were directed to the study homepage, which included more information on the study’s purpose, eligibility criteria, and compensation. The study homepage was hosted on TrialX iConnect, a clinical trial recruitment platform, allowing study staff to track homepage views. This site also provides unlimited tracked hyperlinks, allowing researchers to track the number of social media users engaging with a given hyperlink. However, because of bots and repeat visitors, these proved unreliable. For example, one

subreddit post that received only 533 views reported 1786 link users over the same time period.

The TrialX iConnect home page is linked to the informed consent and eligibility screening in REDCap. Participants were required to provide their age and whether they lived or worked in New York City. Each option also required the participant to select the borough where they lived or worked and provide a ZIP code unless the participant indicated they were not currently working. Questions on substances used in the past 90 days and mode of use (injecting, snorting, oral, and smoking) were adapted from the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST) [38] and expanded to include questions on newer and emerging synthetic opioids (U-47700, nitazenes) and other substances (eg, synthetic cannabinoids and xylazine). There were also several open-ended questions throughout the survey where participants were invited to provide more information about their answers (eg, “In your own words, can you describe what services you received the last time you visited a needle exchange program?”).

After completing the survey, participants were given the option to receive a US \$20 gift card by email from their choice of 4 vendors (Amazon, Target, Starbucks, or Chipotle). Participants were given the option not to provide an email address and forgo the compensation or email the principal investigator directly. After gift card options, we provided a resource guide with information on support services for substance use and mental health in New York City. We also provided a link to the TrialX iConnect homepage for participants to optionally send to others; this link was changed weekly to prevent repeat responses.

Several studies have shown that respondent-driven sampling, where people who use opioids refer other people who use opioids to participate, is particularly useful with hard-to-reach populations [39,40].

Bot Prevention

REDCap includes Google's reCAPTCHA service, a type of challenge-response test to determine whether the user is human in order to deter bot attacks and spam. However, REDCap does not allow the programmer to choose the level of CAPTCHA security; rather, Google decides the stringency of the test. In addition, we included a picture of a group of objects (eg, 4 cats) and asked respondents to enter the number of the object in a text box. This is because it is more cumbersome to program a bot to count objects in an image versus identifying an object [41]. REDCap does not allow images to randomly change in a survey, so study staff manually changed the image periodically to prevent manually programmed bots from continuing the survey. We included "honeypot" questions—items programmed to be hidden from respondents but answered by automated bots [32]—but no respondents answered these questions. We collected IP addresses to determine whether multiple surveys were submitted from the same location [42]. We also included a fictitious drug to detect participants who did not answer honestly or carefully [42]. We did not collect any personally identifiable information other than IP and email addresses for gift card distribution.

Post Hoc Data Inspection

We implemented a rigorous, multistep post hoc data cleaning protocol to detect bot submissions, repeat participants, and other mischievous or malicious responders (see Figure S2 in [Multimedia Appendix 1](#)). First, we excluded IP addresses outside the United States or Canada and duplicate IP and email addresses. We excluded respondents who endorsed the use of a fictitious drug. We also excluded participants who reported living or working in invalid ZIP codes (eg, 10048, the retired ZIP code for the World Trade Center) or if the ZIP codes did not match the borough in which they reported living or working (eg, the participant chose "Brooklyn" from the drop-down menu but wrote in a home ZIP code in Buffalo, NY). We checked that participants who listed invalid or discrepant ZIP codes did not also report experiencing homelessness, as this might explain a missing or invalid ZIP code. We next eliminated records reporting improbable or impossible drug use behaviors (eg, injecting alcohol, injecting inhalants, or injecting marijuana), as well as records reporting more than 10 lifetime overdoses. Data inspection showed some respondents listed as many as 105 lifetime overdoses. A New York City study using respondent-driven sampling for in-person data collection found that, while self-reported lifetime overdoses increased with the COVID-19 pandemic, median lifetime overdoses was 3 (IQR 2-6), and only 3.6% (n=10) of respondents reported 10 or more lifetime overdoses [43].

We used the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to identify clusters of behavior where a single individual completes the survey multiple times with the same response set. The DBSCAN algorithm is particularly useful because clusters can be of arbitrary size,

shape, and number [44]. We used the DBSCAN function in the *Scikit-learn* package for Python (version 3.10.10; Python Software Foundation) [45-47]. Not all variables could be included because the algorithm requires complete records; therefore, variables that could be left blank could not be used in the analysis. We specifically targeted demographics (eg, sex, race and ethnicity) and variables related to drug use frequency and mode of use. We limited the algorithm to a very narrow cluster radius, so records must be nearly identical to be flagged as part of a cluster.

After eliminating records flagged for repeated entries, we excluded all records completed in under 7 minutes as this was below the fastest time study staff completed the survey during pilot testing [32,48]. We also examined demographic inconsistencies (eg, Medicaid health insurance and household income over eligibility levels). We then eliminated records with suspicious responses to open-ended questions. These responses often did not correspond to the question or appeared to be AI-generated. For example, when asked to describe the types of medical services received at a needle exchange program, eliminated responses included "very friendly informative and caring," "quite encouraging and satisfying," "it's good," and "A magazine gaudiance [sic]." We also excluded records with identical or near-identical odd phrasing for these responses (eg, "General body check up [sic]" for the medical services question). As a final check, we eliminated records with email addresses comprised of random letters ending in numbers exceeding 4 digits, as these characteristics are an indication of bulk email account creators that can be built or purchased [32,49].

Data Analysis

We present basic frequencies to provide a description of the included data. We also performed a supplemental analysis of excluded records to identify any possible patterns or significant differences from the included data. We compared excluded versus included data using the Pearson chi-square test of independence for categorical variables or the Welch sample *t* test for ratio variables.

Ethical Considerations

This research was approved by the institutional review board at the University of Pennsylvania (protocol 853254). All procedures performed were in accordance with the ethical standards of the University of Pennsylvania institutional review board and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Written informed consent was obtained from all participants before they began the survey.

Results

We collected a total of 2560 complete records, 93.2% (n=2387) of which were determined to be from bots or bad actors. The majority of included responses (n=173) were collected from Meta (n=79, 45.7%) and Reddit (n=48, 27.8%; see [Table 1](#)). Meta's ads cost US \$0.68/click, or US \$14.60 per complete record. X's ads were the most expensive (US \$1.96/click) and yielded the fewest participants (3 completed surveys). This

resulted in a cost of US \$178.33 per completed record for participants recruited through X. We recruited 22.5% (n=39) of the sample through participant peer referral. Overall advertising cost was US \$9.76 per participant in the final sample.

Post hoc data investigation found that 23.9% (571/2560) of excluded records showed a duplicate IP or email address, 45.9% (1095/2560) reported consuming a fictitious drug, 15.8% (378/2560) provided an invalid ZIP code, and 9.8% (235/2560) reported improbable drug use behaviors (Table 2). The largest DBSCAN cluster contained 42 records. We cut off clusters of interest at clusters containing 6 or more records (222/2560, 8.7%) and then analyzed the individual records, noting those that appeared within a short time frame of each other. Overall, 13.7% (351/2560) of all records were in clusters and excluded. Figure S1 in Multimedia Appendix 1 presents a description of how many records were excluded at each step of post hoc data management to arrive at the final sample.

The mean age of the final sample was 29.9 years (SD 6.5; range 18-63 years; see Table 3). A plurality of the sample was male (102/173, 59%) and non-Hispanic Black (77/173, 44.5%). Approximately half of the sample reported using prescription opioids in the previous 90 days (144/173, 51.8%), while 24% (67/173) reported fentanyl use and 20% (54/173) reported heroin use. Oral consumption was the most common mode of use reported (143/173, 83.7%), followed by injection (59/173, 34.1%). A majority of the sample also reported using alcohol (133/173, 76.9%) and marijuana (90/173, 52%) in the previous 90 days. Other commonly used substances included benzodiazepines (51/173, 29.5%) and 3,4-methylenedioxymethamphetamine (MDMA) (36/173, 20.8%). One-fifth reported no lifetime experience with harm

reduction services (34/173, 19.7%), and 10.4% (18/173) reported receipt of condoms as their only lifetime harm reduction experience; 40.5% (70/173) reported no harm reduction engagement in the previous 90 days and 7.5% (13/173) reported receipt of condoms as their only engagement in the previous 90 days. Receipt of condoms was the most commonly reported type of lifetime experience with harm reduction services (79/173, 45.7%), followed by receipt of naloxone (66/173, 38.2%). Receiving sterile syringes from a pharmacy was slightly more common than from a needle exchange program (Lifetime: 49/173, 28.3% vs 41/173, 23.7%; $P=.33$; Past 90 days: 26/173, 15% vs 25/173, 14.5%; $P=.87$).

In the supplemental analysis of excluded records, 49% (n=1169) reported they were non-Hispanic white and 44.5% (n=1062) reported non-Hispanic Black, with significantly fewer reporting being Hispanic or other races or ethnicities than included records ($P<.001$; see Table S1 in Multimedia Appendix 1). A larger percentage of excluded records reported to be men (compared with women), a significantly greater difference compared with the included responses ($P=.002$). Excluded records also reported lower levels of education ($P<.001$) and higher levels of income ($P<.001$) compared with included responses. Excluded records had significantly higher reports of heroin and fentanyl use than included records ($P<.001$), as well as significantly higher use of all other substances ($P<.001$). Excluded records also showed significantly more smoking and snorting of opioids and significantly less oral consumption ($P<.001$). Excluded records reported significantly more lifetime overdoses compared with included records (included: mean 1.0, SD 1.51; excluded: mean 4.3, SD 9.43; $P<.001$). There were no significant differences in age, sexual orientation, or housing stability.

Table 1. Recruitment number and cost by social media platform, omitting records excluded during post hoc data management.

Platform	Impressions	Clicks, n	Cost/click (US \$)	Total cost (US \$)	n (%)	Cost/N (US \$)
Meta (Facebook and Instagram)	165,751	1700	0.68	1153.26	79 (45.7)	14.60
X (formerly Twitter)	161,792	270	1.96	535	3 (1.7)	178.33
Reddit	50,508	— ^a	0	0	48 (27.8)	0
Discord	—	—	0	0	3 (1.7)	0
Bluelight web forum	—	—	0	0	1 (0.6)	0
Friend referral	—	—	—	—	39 (22.5)	0

^aNot available.

Table 2. Post hoc data management violations (N=2560 total complete records).

Violation category	Total records, n (%)
IP addresses outside the United States or Canada	279 (10.9)
Duplicate IP and, or email addresses	571 (22.3)
Endorsed use of fictitious drug	1095 (42.8)
Invalid ZIP codes (eg, 10048)	400 (15.6)
ZIP codes not in the reported borough	124 (4.8)
Improbable drug use behaviors (eg, injecting alcohol, injecting marijuana)	235 (9.2)
>10 lifetime overdoses	232 (9.1)
Clustering (DBSCAN ^a)	351 (13.7)
<7-minute survey response times	76 (3)
Demographic discrepancies (eg, income and insurance)	32 (1.3)
Suspicious responses to open-ended questions (eg, “A magazine gaudiance” for medical services)	116 (4.5)
Email address check	15 (0.6)

^aDBSCAN: Density-Based Spatial Clustering of Applications with Noise.

Table 3. Description of the sample after post hoc data inspection (N=173).

Variable	Total (N=173)
Age (years), mean (SD)	29.86 (6.52)
Gender, n (%)	
Man	102 (58.96)
Woman	62 (35.84)
Transgender	4 (2.31)
Nonbinary	4 (2.31)
Prefer not to say	1 (0.58)
Race and ethnicity, n (%)	
Non-Hispanic White	68 (39.31)
Non-Hispanic Black	77 (44.51)
Hispanic	13 (7.51)
Asian	7 (4.05)
Mixed race	6 (3.47)
Other	2 (1.16)
Sexual orientation, n (%)	
Heterosexual	141 (81.50)
Bisexual	18 (10.40)
Homosexual or queer	11 (6.36)
Prefer not to say	3 (1.73)
Home borough, n (%)	
Manhattan	39 (23.08)
Brooklyn	57 (33.73)
Bronx	29 (17.16)
Queens	26 (15.38)
Staten Island	12 (7.10)
Outside of New York City	6 (3.55)
History of homelessness, n (%)	
Stably housed, no past homelessness	102 (58.96)
Stably house, past homelessness	46 (26.59)
Current homelessness	15 (8.67)
Stably housed but worried about housing	10 (5.78)
Income (US \$), n (%)	
0-29,999	34 (19.65)
30,000-39,999	26 (15.03)
40,000-49,999	9 (5.2)
50,000-59,999	22 (12.72)
60,000-74,999	20 (11.56)
75,000-99,999	36 (20.81)
100,000 or more	19 (10.98)
Don't know	7 (4.05)
Education level, n (%)	
Less than a high school diploma	2 (1.16)

Variable	Total (N=173)
A high school diploma or General Educational Development (GED) test	21 (12.14)
Some college or a 2-year degree	59 (34.10)
4-year college degree	74 (42.77)
Postgraduate work	17 (9.83)
Health insurance, n (%)	
Yes	136 (78.61)
No	41 (23.70)
Children, n (%)^a	
Living with biological child <18 years old	37 (21.39)
Not living with child <18 years old (eg, child with grandparent)	19 (10.98)
No children or no children <18 years old	104 (60.12)
Opioid use, past 90 days, n (%)^a	
Prescription	144 (51.80)
Heroin	54 (19.42)
Fentanyl	67 (24.10)
Other	13 (4.68)
Mode of opioid use, n (%)^a	
Oral, total	143 (82.66)
Prescription	125 (70.22)
Heroin	3 (4.69)
Fentanyl	28 (35.44)
Other	9 (69.23)
Snort, total	34 (19.65)
Prescription	14 (7.87)
Heroin	18 (28.13)
Fentanyl	16 (20.25)
Other	1 (7.69)
Smoke, total	21 (12.14)
Prescription	6 (3.37)
Heroin	9 (14.06)
Fentanyl	9 (11.39)
Other	1 (7.69)
Inject, total	59 (34.10)
Prescription	33 (18.54)
Heroin	34 (53.13)
Fentanyl	26 (32.91)
Other	2 (15.38)
Other substances, past 90 days, n (%)^a	
Alcohol	133 (76.88)
Marijuana	90 (52.02)
Cocaine	33 (19.08)
Benzodiazepines	51 (29.48)

Variable	Total (N=173)
Methamphetamine	22 (12.72)
3,4-Methylenedioxymethamphetamine (MDMA)	36 (20.81)
Ketamine	17 (9.83)
Amphetamines (eg, Adderall)	34 (19.65)
Hallucinogens	18 (10.40)
Synthetic marijuana (K2, spice)	26 (15.03)
Xylazine	10 (5.78)
Inhalants	12 (6.94)
Harm reduction experience (lifetime), n (%)^a	
Needle exchange program for needles	41 (23.70)
Needle exchange program for health services	24 (13.87)
Needles from pharmacy	49 (28.32)
Naloxone (Narcan)	66 (38.15)
Sterile works kit	36 (20.81)
Safer smoking items (eg, pipes and stems)	30 (17.34)
Drug testing (eg, fentanyl test strips)	47 (27.17)
Pre-exposure prophylaxis (PrEP) or postexposure prophylaxis (PEP)	30 (17.34)
Condoms	79 (45.66)
Visit an overdose prevention center or supervised consumption site	41 (23.70)
Harm reduction experience (past 90 days), n (%)^a	
Needle exchange program for needles	25 (14.45)
Needle exchange program for health services	15 (8.67)
Needles from pharmacy	26 (15.03)
Naloxone (Narcan)	47 (27.17)
Sterile works kit	22 (12.72)
Safer smoking items (eg, pipes and stems)	20 (11.56)
Drug testing (eg, fentanyl test strips)	31 (17.92)
Pre-exposure prophylaxis (PrEP) or postexposure prophylaxis (PEP)	18 (10.4)
Condoms	54 (31.21)
Visit an overdose prevention center or supervised consumption site	20 (11.56)
Lifetime history of overdose, mean (SD)	
Zero	1.00 (1.51)
≥1	98 (56.65)
Depression (Center of Epidemiologic Studies Depression Scale-10), mean (SD)	
Depressed (≥10)	75 (43.35)
Not depressed (<10)	11.4 (5.4)
Anxiety (General Anxiety Disorder-7), mean (SD)	
Minimal anxiety (0-4)	106 (61.3)
Mild anxiety (5-9)	67 (38.7)
Moderate anxiety (10-14)	6.9 (5.3)
Severe anxiety (≥15)	69 (39.88)
	49 (28.32)
	45 (26.01)
	10 (5.78)

^aCategories are not mutually exclusive.

Discussion

Principal Findings

Our experience provided an opportunity to assess the continued viability of social media platforms for recruiting a hard-to-reach population of people who use opioids. While we successfully collected final survey data from a diverse sample of 173 people who use opioids, the study was not without challenges. Even the first step—setting up advertisements across platforms—proved difficult. Most institutional review boards require recruitment advertisements to contain detailed information on study purpose, compensation, and inclusion criteria, among other elements, but we found that including these details negatively affected our ability to use these platforms for recruitment. Including compensation information in advertisements resulted in an almost immediate influx of bots on Meta. Meta also flagged our ads as “discriminatory” when we included inclusion criteria, and X (formerly known as Twitter) flagged our study as “spreading false information” when we included study purposes. Therefore, we had to weigh best practices for ethical recruitment with social media platform policies and bot avoidance.

Our recruitment results are consistent with previous studies of social media recruitment, which found the most success with social media recruitment on Meta’s platforms ($n=79$, 45.7%) [50]. In our study, Meta’s ads cost US \$0.68/click, or US \$14.60 per complete record. This is an increase from a 2017 systematic review of 35 studies, which found a median cost of US \$0.51 per click and cost per participant of US \$14.41 for Facebook ads [12]. We found recruitment on X to be expensive for low enrollment despite the ability to use tailored keywords to better target ads. We only recruited 3 participants through X at a cost of US \$1.96 per click and US \$178.33 per participant. A few previous surveys used paid advertisements on Twitter to recruit participants for substance use studies; our results are comparable with previous studies, which found that recruitment through Twitter was high cost with little or no recruitment [50,51]. While Reddit and Discord recruitment posts were unpaid, the time and resources devoted to identifying and contacting moderators for individual subreddits and servers, posting ads, and monitoring comments were cumbersome for the number of yielded participants (Reddit: $n=48$, 27.8%; Discord: $n=3$, 1.7%).

With respect to post hoc data investigation, we determined that 93% (2387/2560) of collected records were invalid or suspicious despite using REDCap’s built-in reCAPTCHA for bot prevention. Excluded records reported significantly more heroin and fentanyl use than included records, as well as significantly more use of all other substances. While previous studies of people who use opioids using social media recruitment have used screening protocols to verify participants [18,42], the proliferation of bots and malicious responders has made it increasingly difficult to verify participant eligibility and data validity. IP addresses are not consistently reliable in identifying repeat participants because IP addresses change periodically at unknown intervals, and IP addresses can be reset manually. Furthermore, IP address location is not reliable for determining geographic eligibility because the IP address location may list

the headquarters of the internet provider and not necessarily where the device accessing the internet is located [52]. Virtual private networks (VPNs), which ensure a secure encrypted connection between the user and the internet, have become considerably cheaper in recent years and can be used to artificially change the geographic location of the user’s IP address. Future studies should investigate new techniques for tracking website visitors, most of which have not yet been incorporated into available internet-based survey platforms to our knowledge. Urchin Tracking Module (UTM) codes are used by Google Analytics to track the behavior of website visitors. They are often used to track which channels generate the most website traffic for commercial purposes. UTM codes have been successfully used to track participant recruitment to health studies from various online forums, including Facebook ads [53], though to our knowledge, they have not been applied to other social media platforms. They might be useful for tracking recruitment across social media platforms and to identify repeat participants. More advanced tracking systems, such as recording media access control (MAC) address—the unique number assigned to each device connected to a network—may also be useful to better identify repeat participants. MAC identifiers have been used in previous health studies but not to identify repeat survey participants to our knowledge [54,55]. In the absence of more advanced technologies to weed out fraudulent responses, researchers should continue to include data validity tests (eg, a fictitious drug and open-ended qualitative response questions). Bot-targeted screeners such as reCAPTCHA and honeypot questions are insufficient as individual human fraudsters become more common [31]. As the purpose of bots and other fraudulent responders is to closely mimic valid data for monetary compensation, researchers must remain vigilant against how stigma and other biases may impact their expectations regarding people who use opioids’ reported behavior [56].

Limitations

This study was a cross-sectional internet-based survey using a convenience sample of people who use opioids in one large US metropolitan area. It is possible that social media platform preference and frequency of use differ in other US cities or globally. Due to institutional regulations, we were unable to purchase ads on Reddit or TikTok. There is a possibility of selection bias resulting from which subreddits, Discord servers, and social media platforms agreed to allow recruitment. We spoke with a research advancement manager at Meta who suggested using the participant data collected to create a “lookalike” audience, or a group of participants representing the ideal survey participant. We were unable to do this because emails provided by the participants to receive the survey gift card did not match emails from registered Facebook users. This may have limited our ability to effectively target our Meta ads to people who use opioids. As this study aimed to engage people who use opioids unaffiliated with community-based organizations to study barriers to harm reduction services engagement, we were not able to partner with community-based organizations for recruitment; study recruitment materials may have benefited from input from people who use opioids with lived experiences. As the iConnect links were unable to

effectively track traffic from different social media platforms, we relied on self-reporting to identify where participants found the survey. While REDCap includes Google's reCAPTCHA service to deter bot attacks and spam, REDCap does not allow the programmer to choose the level of CAPTCHA security. Other internet-based survey platforms may include more intense protection or allow the user to decide the stringency of the CAPTCHA test, reducing the reliance on post hoc data management [32]. While we developed a rigorous post hoc data cleaning protocol, it is possible that invalid data were included in our study while valid surveys were excluded. We may have incorrectly excluded records based on IP address location if respondents used a VPN or Tor browser to hide their IP address due to concerns about being tracked. Due to the stigmatized nature of drug use, and people who use opioids' legitimate

concerns of legal implications for admitting to illegal behaviors, we limited our collection of personal identifiable information (PII) to the collection of IP and email addresses. Other internet-based health surveys have been able to compare names and home addresses against verified national databases to eliminate invalid records [57]. Future internet-based research on opioid use may need to balance the ethics of collecting PII with potentially including invalid data.

Conclusion

Social media recruitment of hidden populations is challenging but not impossible. Rigorous data collection protocols and post hoc data management are necessary to ensure the validity of findings; these methods may counter previous best practices for researching stigmatized behaviors.

Acknowledgments

This work was supported by the Penn Injury Science Center Pilot Grant Program as part of the Centers for Disease Control (CDC; grant 1R49CE003083). This work was also supported by the National Institute on Drug Abuse (grants K01DA049900, R01DA060207, R01DA057289, and R01DA060207). The authors thank the Penn Injury Science Center for their support of this research. The authors thank Mahasweta Dutt, Thomas Hohing, and Paul Sanborn for their assistance with survey coding and purchasing advertisements. Generative AI was not used in any portion of the study design, manuscript writing, or data analysis.

Authors' Contributions

EDN contributed to conceptualization, methodology, supervision, formal analysis, and writing—original and final draft. JJP contributed to writing—review and editing. QL contributed to formal analysis, and writing—original and final draft. WL contributed to formal analysis and writing—original and final draft. SSM contributed to methodology and writing—review and editing.

Conflicts of Interest

JJP has consulted for the Washington-Baltimore High Intensity Drug Trafficking Area (HIDTA) program. No conflicts of interest declared by the other authors.

Multimedia Appendix 1

Further description of excluded records by post hoc data management step and comparison of included versus excluded records. [[DOCX File, 302 KB-Multimedia Appendix 1](#)]

References

1. Post LA, Lundberg A, Moss CB, Brandt CA, Quan I, Han L, et al. Geographic trends in opioid overdoses in the US from 1999 to 2020. *JAMA Netw Open*. 2022;5(7):e2223631. [[FREE Full text](#)] [doi: [10.1001/jamanetworkopen.2022.23631](https://doi.org/10.1001/jamanetworkopen.2022.23631)] [Medline: [35900768](#)]
2. Drug overdose deaths. Centers for Disease Control and Prevention. 2023. URL: <https://www.cdc.gov/nchs/hus/topics/drug-overdose-deaths.htm> [accessed 2024-05-13]
3. Ahmad F, Cisewski J, Rossen L, Sutton P. Provisional drug overdose death counts. National Center for Health Statistics. 2024. URL: <https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm> [accessed 2024-05-13]
4. Miller PG, Sönderlund AL. Using the internet to research hidden populations of illicit drug users: a review. *Addiction*. 2010;105(9):1557-1567. [doi: [10.1111/j.1360-0443.2010.02992.x](https://doi.org/10.1111/j.1360-0443.2010.02992.x)] [Medline: [20626378](#)]
5. Nesoff ED, Martins SS, Palamar JJ. Caution Is Necessary When Estimating Treatment Need for Opioid Use Disorder Using National Surveys. *Am J Public Health*. Feb 2022;112(2):199-201. [doi: [10.2105/AJPH.2021.306624](https://doi.org/10.2105/AJPH.2021.306624)] [Medline: [35080936](#)]
6. Värmå Falk M, Strömdahl S, Ekström AM, Kåberg M, Karlsson N, Dahlborn H, et al. A qualitative study of facilitators and barriers to participate in a needle exchange program for women who inject drugs. *Harm Reduct J*. Oct 22, 2020;17(1):84. [[FREE Full text](#)] [doi: [10.1186/s12954-020-00425-9](https://doi.org/10.1186/s12954-020-00425-9)] [Medline: [33092595](#)]
7. Gibson K, Hutton F. Women Who Inject Drugs (WWID): Stigma, Gender and Barriers to Needle Exchange Programmes (NEPs). *Contemporary Drug Problems*. Jul 19, 2021;48(3):276-296. [doi: [10.1177/00914509211035242](https://doi.org/10.1177/00914509211035242)]
8. Collins AB, Boyd J, Czechaczek S, Hayashi K, McNeil R. (Re)shaping the self: An ethnographic study of the embodied and spatial practices of women who use drugs. *Health Place*. May 2020;63:102327. [[FREE Full text](#)] [doi: [10.1016/j.healthplace.2020.102327](https://doi.org/10.1016/j.healthplace.2020.102327)] [Medline: [32224291](#)]

9. Davis CS, Burris S, Kraut-Becher J, Lynch KG, Metzger D. Effects of an intensive street-level police intervention on syringe exchange program use in Philadelphia, PA. *Am J Public Health*. Feb 2005;95(2):233-236. [doi: [10.2105/AJPH.2003.033563](https://doi.org/10.2105/AJPH.2003.033563)] [Medline: [15671455](https://pubmed.ncbi.nlm.nih.gov/15671455/)]
10. Scott MC, Wahl OF. Substance abuse stigma and discrimination among African American male substance users. *SRA*. 2011;1(1):60-66. [doi: [10.5463/sra.v1i1.3](https://doi.org/10.5463/sra.v1i1.3)]
11. Guillory J, Wiant KF, Farrelly M, Fiacco L, Alam I, Hoffman L, et al. Recruiting hard-to-reach populations for survey research: using Facebook and Instagram advertisements and in-person intercept in LGBT bars and nightclubs to recruit LGBT young adults. *J Med Internet Res*. 2018;20(6):e197. [FREE Full text] [doi: [10.2196/jmir.9461](https://doi.org/10.2196/jmir.9461)] [Medline: [29914861](https://pubmed.ncbi.nlm.nih.gov/29914861/)]
12. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. *J Med Internet Res*. 2017;19(8):e290. [FREE Full text] [doi: [10.2196/jmir.7071](https://doi.org/10.2196/jmir.7071)] [Medline: [28851679](https://pubmed.ncbi.nlm.nih.gov/28851679/)]
13. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. Washington, DC.; 2021. URL: <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/> [accessed 2025-04-01]
14. Sanchez C, Grzenda A, Varias A, Widge AS, Carpenter LL, McDonald WM, et al. Social media recruitment for mental health research: A systematic review. *Compr Psychiatry*. 2020;103:152197. [FREE Full text] [doi: [10.1016/j.comppsych.2020.152197](https://doi.org/10.1016/j.comppsych.2020.152197)] [Medline: [32992073](https://pubmed.ncbi.nlm.nih.gov/32992073/)]
15. Thornton LK, Harris K, Baker AL, Johnson M, Kay-Lambkin FJ. Recruiting for addiction research via Facebook. *Drug Alcohol Rev*. 2016;35(4):494-502. [doi: [10.1111/dar.12305](https://doi.org/10.1111/dar.12305)] [Medline: [26179224](https://pubmed.ncbi.nlm.nih.gov/26179224/)]
16. Watson B, Robinson DHZ, Harker L, Arriola KRJ. The inclusion of African-American study participants in web-based research studies: viewpoint. *J Med Internet Res*. 2016;18(6):e168. [FREE Full text] [doi: [10.2196/jmir.5486](https://doi.org/10.2196/jmir.5486)] [Medline: [27334683](https://pubmed.ncbi.nlm.nih.gov/27334683/)]
17. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. *Internet Interv*. 2016;4:1-10. [FREE Full text] [doi: [10.1016/j.invent.2016.01.001](https://doi.org/10.1016/j.invent.2016.01.001)] [Medline: [30135786](https://pubmed.ncbi.nlm.nih.gov/30135786/)]
18. Wilkerson JM, Shenk JE, Grey JA, Simon Rosser BR, Noor SW. Recruitment strategies of methamphetamine-using men who have sex with men into an online survey. *J Subst Use*. 2015;20(1):33-37. [FREE Full text] [doi: [10.3109/14659891.2013.868936](https://doi.org/10.3109/14659891.2013.868936)] [Medline: [25642143](https://pubmed.ncbi.nlm.nih.gov/25642143/)]
19. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. *J Med Internet Res*. 2014;16(5):e117. [FREE Full text] [doi: [10.2196/jmir.3229](https://doi.org/10.2196/jmir.3229)] [Medline: [24784982](https://pubmed.ncbi.nlm.nih.gov/24784982/)]
20. Internet, broadband fact sheet. Pew Research Center. 2024. URL: <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/?tabId=480dace1-fd73-4f03-ad88-eae66e1f4217> [accessed 2024-05-12]
21. Blumenberg C, Menezes AMB, Gonçalves H, Assunção MCF, Wehrmeister FC, Barros AJD. How different online recruitment methods impact on recruitment rates for the web-based coortesnaweb project: a randomised trial. *BMC Med Res Methodol*. 2019;19(1):127. [FREE Full text] [doi: [10.1186/s12874-019-0767-z](https://doi.org/10.1186/s12874-019-0767-z)] [Medline: [31217008](https://pubmed.ncbi.nlm.nih.gov/31217008/)]
22. Sullivan PS, Khosropour CM, Luisi N, Amsden M, Coggia T, Wingood GM, et al. Bias in online recruitment and retention of racial and ethnic minority men who have sex with men. *J Med Internet Res*. 2011;13(2):e38. [FREE Full text] [doi: [10.2196/jmir.1797](https://doi.org/10.2196/jmir.1797)] [Medline: [21571632](https://pubmed.ncbi.nlm.nih.gov/21571632/)]
23. Folk JB, Harrison A, Rodriguez C, Wallace A, Tolou-Shams M. Feasibility of social media-based recruitment and perceived acceptability of digital health interventions for caregivers of justice-involved youth: mixed methods study. *J Med Internet Res*. 2020;22(4):e16370. [FREE Full text] [doi: [10.2196/16370](https://doi.org/10.2196/16370)] [Medline: [32352388](https://pubmed.ncbi.nlm.nih.gov/32352388/)]
24. Tahir D. Facebook ad ban may squelch medical research recruitment. *Politico*. Jan 2, 2022. URL: <https://www.politico.com/news/2022/01/02/facebook-ad-ban-medical-research-recruitment-526275> [accessed 2022-07-20]
25. Hsu T. Twitter's advertisers pull back as layoffs sweep through company. *New York Times*. Nov 2, 2022. URL: <https://www.nytimes.com/2022/11/04/technology/twitter-advertisers.html> [accessed 2022-11-03]
26. Bryant M. Twitter applies reading limit after users report issues with platform. *The Guardian*. Jul 1, 2023. URL: <https://www.theguardian.com/technology/2023/jul/01/twitter-applies-reading-limit-after-users-report-issues-with-platform> [accessed 2023-08-27]
27. Lorenz T. How Twitter lost its place as the global town square. *Washington Post*. Jul 7, 2023. URL: <https://www.washingtonpost.com/technology/2023/07/07/twitter-dead-musk-tiktok-public-square/> [accessed 2023-08-27]
28. Ingram D. Reddit CEO slams protest leaders, saying he'll change rules that favor 'landed gentry'. *NBC News*. Jun 15, 2023. URL: <https://www.nbcnews.com/tech/tech-news/reddit-protest-blackout-ceo-steve-huffman-moderators-rcna89544> [accessed 2023-08-27]
29. Kircher MM. What's going on with Reddit? *New York Times*. Jun 16, 2023. URL: <https://www.nytimes.com/2023/06/16/style/whats-going-on-with-reddit.html> [accessed 2023-08-09]
30. Abdelazeem B, Hamdallah A, Rizk MA, Abbas KS, El-Shahat NA, Manasrah N, et al. Does usage of monetary incentive impact the involvement in surveys? A systematic review and meta-analysis of 46 randomized controlled trials. *PLoS One*. 2023;18(1):e0279128. [FREE Full text] [doi: [10.1371/journal.pone.0279128](https://doi.org/10.1371/journal.pone.0279128)] [Medline: [36649255](https://pubmed.ncbi.nlm.nih.gov/36649255/)]
31. Johnson MS, Adams VM, Byrne J. Addressing fraudulent responses in online surveys: insights from a web - based participatory mapping study. *People and Nature*. 2023;6(1):147-164. [doi: [10.1002/pan3.10557](https://doi.org/10.1002/pan3.10557)]

32. Griffin M, Martino RJ, LoSchiavo C, Comer-Carruthers C, Krause KD, Stults CB, et al. Ensuring survey research data integrity in the era of internet bots. *Qual Quant*. 2022;56(4):2841-2852. [FREE Full text] [doi: [10.1007/s11135-021-01252-1](https://doi.org/10.1007/s11135-021-01252-1)] [Medline: [34629553](https://pubmed.ncbi.nlm.nih.gov/34629553/)]
33. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsy JA, Hong F, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. *J Med Internet Res*. 2020;22(10):e23021. [FREE Full text] [doi: [10.2196/23021](https://doi.org/10.2196/23021)] [Medline: [33026360](https://pubmed.ncbi.nlm.nih.gov/33026360/)]
34. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying multiple submissions in Internet research: preserving data integrity. *AIDS Behav*. 2008;12(6):964-973. [FREE Full text] [doi: [10.1007/s10461-007-9352-2](https://doi.org/10.1007/s10461-007-9352-2)] [Medline: [18240015](https://pubmed.ncbi.nlm.nih.gov/18240015/)]
35. Kuriakose N, Robbins M. Don't get duped: fraud through duplication in public opinion surveys. *Statistical Journal of the IAOS*. 2018;21(1):195-214. [FREE Full text]
36. Arthur W, Hagen E, George F. The lazy or dishonest respondent: detection and prevention. *Annu Rev Organ Psychol Organ Behav*. 2021;8(1):105-137. [doi: [10.1146/annurev-orgpsych-012420-055324](https://doi.org/10.1146/annurev-orgpsych-012420-055324)]
37. List of subreddits about drugs, addiction, drug discussions, drug music, etc. cyrilio. 2019. URL: https://www.reddit.com/r/ListOfSubreddits/comments/abmkx6/list_of_subreddits_about_drugs_addiction_drug/?rdt=57427 [accessed 2023-08-13]
38. WHO ASSIST Working Group. The alcohol, smoking and substance involvement screening test (ASSIST): development, reliability and feasibility. *Addiction*. 2002;97(9):1183-1194. [FREE Full text] [doi: [10.1046/j.1360-0443.2002.00185.x](https://doi.org/10.1046/j.1360-0443.2002.00185.x)] [Medline: [12199834](https://pubmed.ncbi.nlm.nih.gov/12199834/)]
39. Heckathorn DD. Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*. 1997;44(2):174-199. [doi: [10.2307/3096941](https://doi.org/10.2307/3096941)]
40. Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems*. 2002;49(1):11-34. [doi: [10.1525/sp.2002.49.1.11](https://doi.org/10.1525/sp.2002.49.1.11)]
41. Chen W, Luo J, Zhang F, Tian Z. A review of object detection: datasets, performance evaluation, architecture, applications and current trends. *Multimed Tools Appl*. 2024;83(24):65603-65661. [doi: [10.1007/s11042-023-17949-4](https://doi.org/10.1007/s11042-023-17949-4)]
42. Palamar JJ, Acosta P. On the efficacy of online drug surveys during the time of COVID-19. *Subst Abus*. 2020;41(3):283-285. [FREE Full text] [doi: [10.1080/08897077.2020.1784362](https://doi.org/10.1080/08897077.2020.1784362)] [Medline: [32697173](https://pubmed.ncbi.nlm.nih.gov/32697173/)]
43. Des Jarlais D, Weng CA, Feelemyer J, McKnight C. Non-fatal drug overdose among persons who inject drugs during first two years of the COVID-19 pandemic in New York City: Prevalence, risk factors. *Drug Alcohol Depend Rep*. 2023;8:100171. [FREE Full text] [doi: [10.1016/j.dadr.2023.100171](https://doi.org/10.1016/j.dadr.2023.100171)] [Medline: [37323939](https://pubmed.ncbi.nlm.nih.gov/37323939/)]
44. Langhnoja SG, Barot MP, Mehta DB. Web usage mining to discover visitor group with common behavior using DBSCAN clustering algorithm. *IJEIT*. 2013;2(7):173. [FREE Full text]
45. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *Journal of Machine Learning Research*. 2011;12(85):2825-2830. [FREE Full text]
46. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. AAAI Press; 1996. Presented at: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96); April 01, 2025:226-231; Portland, Oregon. URL: <https://dl.acm.org/doi/10.5555/3001460.3001507> [doi: [10.5555/3001460.3001507](https://doi.org/10.5555/3001460.3001507)]
47. Schubert E, Sander J, Ester M, Kriegel HP, Xu X. DBSCAN Revisited, Revisited: why and how you should (still) use DBSCAN. *ACM Trans Database Syst*. 2017;42(3):1-21. [doi: [10.1145/3068335](https://doi.org/10.1145/3068335)]
48. Leiner DJ. Too fast, too straight, too weird: non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*. 2019;13(3):229-248. [doi: [10.18148/srm/2019.v13i3.7403](https://doi.org/10.18148/srm/2019.v13i3.7403)]
49. Wang Z, Qin M, Chen M, Jia C. Hiding fast flux botnet in plain email sight. In: Lin X, Ghorbani A, Ren K, Zhu S, Zhang A, editors. *Security and Privacy in Communication Networks. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Cham. Springer; 2018.
50. Parker JN, Hunter AS, Bauermeister JA, Bonar EE, Carrico A, Stephenson R. Comparing social media and in-person recruitment: lessons learned from recruiting substance-using, sexual and gender minority adolescents and young adults for a randomized control trial. *JMIR Public Health Surveill*. 2021;7(12):e31657. [FREE Full text] [doi: [10.2196/31657](https://doi.org/10.2196/31657)] [Medline: [34855613](https://pubmed.ncbi.nlm.nih.gov/34855613/)]
51. Mackesy-Amiti M, Johnson T, Boodram B. Comparison of methods for recruiting suburban opioid users for research. UIC University Library. Preprint posted online on December 03, 2021. [doi: [10.25417/uic.17078276.v1](https://doi.org/10.25417/uic.17078276.v1)]
52. Nur AY. Accuracy and coverage analysis of IP geolocation databases. 2023. Presented at: International Balkan Conference on Communications and Networking Istanbul; 2023 Jun 08; Turkey. URL: <https://ieeexplore.ieee.org/document/10167899>
53. Walter S, Clanton TB, Langford OG, Rafii MS, Shaffer EJ, Grill JD, et al. Recruitment into the alzheimer prevention trials (APT) webstudy for a trial-ready cohort for preclinical and prodromal alzheimer's disease (TRC-PAD). *J Prev Alzheimers Dis*. 2020;7(4):219-225. [FREE Full text] [doi: [10.14283/jpad.2020.46](https://doi.org/10.14283/jpad.2020.46)] [Medline: [32920623](https://pubmed.ncbi.nlm.nih.gov/32920623/)]
54. Aiello AE, Simanek AM, Eisenberg MC, Walsh AR, Davis B, Volz E, et al. Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial. *Epidemics*. 2016;15:38-55. [FREE Full text] [doi: [10.1016/j.epidem.2016.01.001](https://doi.org/10.1016/j.epidem.2016.01.001)] [Medline: [27266848](https://pubmed.ncbi.nlm.nih.gov/27266848/)]
55. Abedi N, Bhaskar A, Chung E. Tracking spatio-temporal movement of human in terms of space utilization using media-access-control address data. *Applied Geography*. 2014;51:72-81. [doi: [10.1016/j.apgeog.2014.04.001](https://doi.org/10.1016/j.apgeog.2014.04.001)]

56. DeSimone JA, Harms PD. Dirty data: The effects of screening respondents who provide low-quality data in survey research. *J Bus Psychol*. 2017;33(5):559-577. [doi: [10.1007/s10869-017-9514-9](https://doi.org/10.1007/s10869-017-9514-9)]
57. Glazer JV, MacDonnell K, Frederick C, Ingersoll K, Ritterband LM. Liar! Liar! Identifying eligibility fraud by applicants in digital health research. *Internet Interv*. Sep 2021;25:100401. [FREE Full text] [doi: [10.1016/j.invent.2021.100401](https://doi.org/10.1016/j.invent.2021.100401)] [Medline: [34094883](https://pubmed.ncbi.nlm.nih.gov/34094883/)]

Abbreviations

API: application programming interface
ASSIST: Alcohol, Smoking, and Substance Involvement Screening Test
DBSCAN: Density-Based Spatial Clustering of Applications with Noise
MAC: media access control
MDMA: 3,4-methylenedioxymethamphetamine
NYC: New York City
PII: personal identifiable information
REDCap: Research Electronic Data Capture
UTM: Urchin Tracking Module
VPN: virtual private network

Edited by N Cahill; submitted 26.06.24; peer-reviewed by M Barratt; comments to author 14.11.24; revised version received 14.02.25; accepted 27.02.25; published 30.04.25

Please cite as:

Nesoff ED, Palamar JJ, Li Q, Li W, Martins SS

Challenging the Continued Usefulness of Social Media Recruitment for Surveys of Hidden Populations of People Who Use Opioids
J Med Internet Res 2025;27:e63687

URL: <https://www.jmir.org/2025/1/e63687>

doi: [10.2196/63687](https://doi.org/10.2196/63687)

PMID:

©Elizabeth D Nesoff, Joseph J Palamar, Qingyue Li, Wenqian Li, Silvia S Martins. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.04.2025. 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 (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.