ADDICTION

ADDICTION OPINION AND DEBATE

doi:10.1111/add.15458

Heroin use cannot be measured adequately with a general population survey

Peter Reuter[1,2], Jonathan P. Caulkins[3] & Greg Midgette[1]

University of Maryland, Department of Criminology and Criminal Justice, College Park, MD, USA,[1] University of Maryland, School of Public Policy, College Park, MD, USA[2] and Carnegie Mellon University, Heinz College of Information Systems and Public Policy, Pittsburgh, PA, USA[3]

ABSTRACT

Background Globally, heroin and other opioids account for more than half of deaths and years-of-life-lost due to drug use and comprise one of the four major markets for illegal drugs. Having sound estimates of the number of problematic heroin users is fundamental to formulating sound health and criminal justice policies. Researchers and policymakers rely heavily upon general population surveys (GPS), such as the US National Survey on Drug Use and Health (NSDUH), to estimate heroin use, without confronting their limitations. GPS-based estimates are also ubiquitous for cocaine and methamphetamine, so insights pertaining to GPS for estimating heroin use are also relevant for those drug markets. **Analysis** Four sources of potential errors in NSDUH are assessed: selective non-response, small sample size, sampling frame omissions and under-reporting. An alternative estimate drawing on a variety of sources including a survey of adult male arrestees is presented and explained. Other approaches to prevalence estimation are discussed. **Findings** Under-reporting and selective non-response in NSDUH are likely to lead to substantial underestimation. Small sample size leads to imprecise estimates and erratic year-to-year fluctuations. The alternative estimate provides credible evidence that NSDUH underestimates the number of frequent heroin users by at least three-quarters and perhaps much more. **Implications** GPS, even those as strong as NSDUH, are doomed by their nature to estimate poorly a rare and stigmatized behavior concentrated in a hard-to-track population. Although many European nations avoid reliance upon these surveys, many others follow the US model. Better estimation requires models that draw upon a variety of data sources, including GPS, to provide credible estimates. Recent methodological developments in selected countries can provide guidance. Journals should require researchers to critically assess the soundness of GPS estimates for any stigmatized drug-related behaviors with low prevalence rates.

Keywords Drug policy, general population surveys, heroin, hidden populations, opioids, prevalence estimation.

Correspondence to: Peter Reuter, University of Maryland, School of Public Policy, College Park, MD 201742, USA. E-mail: preuter@umd.edu
Submitted 21 August 2020; initial review completed 23 October 2020; final version accepted 18 February 2021

INTRODUCTION

Substance use policy has emerged as a scientific discipline with a commendable commitment to a grounding in empirical evidence, but some core quantities are almost as elusive as the Higgs boson. The juxtaposition of great demand for numbers with a limited ability to supply them can lead to uncritical acceptance of the few quantitative estimates that exist. This temptation can be particularly acute when the number enjoys the patina of rigor, such as those deriving from a representative survey. Journals and policymakers have a duty to resist that temptation.

Here we explore the important case of using general population surveys (GPS) to estimate frequent heroin use. Heroin is important in its own right, but the implications of this case study ripple out to other substances (cocaine/crack and methamphetamine), to the general practice of using confidence intervals based on sampling variability as proxies for total uncertainty and to relying upon single data sources, as opposed to triangulating with multiple data sources.

Heroin has occupied a distinctively important place in drug policy for more than a century. It and other opioids account for more than half of deaths and years-of-life-lost due to drug use [1], and are involved in 82% of fatal overdoses in Europe [2]. In the United States, the most recent national market assessment's best estimate was that its 2.3 million chronic heroin users support a criminal market with $43 billion in annual revenues; infrequent users are not uncommon, but they account for only a sliver of market demand because they consume so much less per person than do chronic users [3]. Hence, credible estimates of the number of heroin users—particularly frequent users—are invaluable for planning.

Discussions of nation-wide numbers of heroin users often cite estimates based on GPS such as the US National Survey on Drug Use and Health (NSDUH). NSDUH is among the largest and strongest GPS, with approximately 70 000 completed interviews per year. Despite the large sample size, few respondents report frequent use of heroin. Among the more than 330 000 respondents between 2011 and 2016 for whom survey data are publicly available, only 217 reported using heroin 'more than four days per week', yielding an average population-weighted estimate of 156 000 such daily or near-daily (DND) users at any given time. We believe that estimate is low by approximately an order of magnitude, as indicated by an alternative estimation exercise which takes advantage of a survey of arrestees as well as other data sources. The true number of DND users is almost certainly well above the upper end of the reported 95% confidence interval which encompasses sampling variability but not other, more fundamental, limitations of GPS estimates. In 2016, for example, the 95% confidence interval around the point estimate of 211 000 was 152 000–271 000. We believe that the true estimate is closer to 2 million DND users, and that reporting a statistical confidence interval derived from just one dimension of uncertainty (sampling variation) connotes a much better understanding of the cumulative uncertainty surrounding the estimate than is appropriate.

The problem is not confined to heroin or to the United States. The same skepticism should be applied to estimates of frequent cocaine and methamphetamine use, as these are also rare, stigmatized behaviors of difficult-to-track populations.

Most western European nations, following the guidance of the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), do not use household surveys to estimate the prevalence of either 'problem use' (i.e. 'injecting drug use or long duration/regular use of opioids, cocaine and/or amphetamines') or heroin use. However, many other countries rely upon household surveys for heroin prevalence estimates, as reflected in the United Nations Office on Drugs and Crime (UNODC)'s *World Drug Report* estimates of global and regional prevalence [4]. Indeed, one critical parameter in the global estimate is derived explicitly from the US survey and applied to other countries.

We focus upon NSDUH because it is among the best GPS. We focus on heroin because it simpler to measure than the full spectrum of diverse opioids, including prescription opioids. If this strong GPS fails at estimating that one best-known opioid, then *a fortiori* there are concerns with GPS-based estimates of problem opioid use more generally.

TROUBLING RELIANCE UPON NSDUH

Some may think that the limitations of GPS for estimating heroin use are so obvious that no one would use them, so we begin by briefly demonstrating how heavily US policy discussions and research rely upon NSDUH-based estimates. Articles in prestigious medical and public health journals as well as government reports invariably cite NSDUH estimates for the prevalence of heroin use, and usually do so without qualification; when qualifications are mentioned, their consequences are not. We conducted a literature search and found no competing estimates of frequent heroin use except in our work described below. A few sample quotations make the point.

'According to national surveillance data (NSDUH), 914,000 people reported heroin use in 2014…' [5]. No qualification is offered in this *New England Journal of Medicine* article.

In a study of the effects of prescription drug monitoring programs on heroin use that relied upon NSDUH, Ali *et al.* [6] observed that: 'This study has a few important limitations' and listed: 'First, the NSDUH does not survey the same individuals from year to year', but that was the only listed limitation that pertained to NSDUH.

In December 2017 congressional testimony, NIH Director Francis Collins stated: '[i]n 2016, over 11 million Americans misused prescription opioids, nearly 1 million used heroin, and 2.1 million had an opioid use disorder due to prescription opioids or heroin' [7]. This mirrors the 2016 Surgeon General's Report *Facing Addiction* citing NSDUH for past-year use estimates, again without qualification [8]. An NSDUH table for past-year use or misuse lists the number at 800 000 and the number of initiates at 100 000. This very lengthy report, with contributions from many agencies, does not contain a single qualification about the NSDUH prevalence data.

NSDUH also influences other important public health figures. Estimates of numbers who meet criteria for substance use disorder but nonetheless do not receive treatment (the 'treatment gap') often rely upon NSDUH or its predecessors (e.g. [9]).

NSDUH'S LIMITATIONS

According to the 2016 NSDUH public use file, the survey estimated that 941 993 people aged 12 years or older in the United States used heroin in the past year, 479 114 used in the past month, 591 971 met the DSM-IV criteria for heroin dependence and another 46 086 met the criteria for abuse but not dependence [10].

Distinguishing problem users from the larger number who have used at least once in the last year is important because problem users generate and incur disproportionate harms from use, including risk of blood-borne disease transmission and premature mortality [11]. Here we focus upon those who report using on more than some benchmark number of days in the past month because that distinction hinges on a single, relatively easy-to-answer survey question. Frequent use is a partial proxy for the number of people suffering from heroin use disorder, which is harder to measure directly. Frequent users also dominate the demand that fuels drug trafficking and its attendant impacts on public health and safety [3,12]. Reducing frequent use may be a fair surrogate target for many interventions; the US Office of National Drug Control Policy has formulated such targets in the past for other drugs [13].

How precise and accurate are NSDUH's heroin estimates? There is no ground truth against which estimates can be compared. However, for reasons discussed here, we think the answer may be 'imprecise and inaccurate'—even though NSDUH's parent agency [Substance Abuse and Mental Health Services Administration (SAMHSA)] has undertaken important methodological improvements during the survey's long history [14,15]. This judgment should not be surprising; Cook [16] suggests that GPS only record about half of US alcohol consumption—although more than half of the prevalence of alcohol use.

Many limitations are well known [17–19], but nonetheless bear listing, as follows.

Selective non-response

Response rates for most surveys have fallen substantially since approximately 1990 [20]. NSDUH rates are high by comparison but selective non-response can still be a very serious problem, as some arithmetic shows.

In 2016, NSDUH identified 173 149 eligible sampling units (meaning either households or units within group quarters). Approximately four of five (135 188) completed the screening process. Within those, 95 607 individuals were selected to be interviewed, but only 67 942 (69%) of them became respondents. Most others simply refused to participate, but some were either never at home or could not complete for other reasons, such as mental incompetence or language barriers.

Setting aside the complex sampling structure and simply multiplying the weighted screening rate (80%) by the 69% provides an approximate aggregate of 55% of eligible response units becoming respondents. In 2016, the population NSDUH intends to represent was approximately 269 million people aged 12 years and older, so NSDUH's estimate of 479 114 past-month users corresponds to a prevalence of 0.18%. If that 0.18% only applied to the 55%, and the prevalence among the other 45% were 0.56%, then the true overall prevalence of $0.55 \times 0.18\% + 0.45 \times 0.56\% = 0.36\%$ would be double that reported by NSDUH.

The prevalence of those who did not provide data is, of course, unknown. Gfroerer *et al.* [21] note that although some populations with high non-response tend to have high rates of drug use (e.g. males in urban areas) others have lower than average drug use rates (older and high-income populations). Nevertheless, one can ask whether it seems possible that those who were hard to reach or refused to cooperate might be 0.56%/0.18% = 3.1 times as likely to use heroin? It is hard to dismiss that possibility.

Furthermore, the complex sampling structure can create challenges, particularly for year-to-year trends in subnational estimates. Geographic variation over time in the inclusion (or exclusion) of populations in specific geographical areas can generate non-random shifts in all measures of use that are independent of variables used for weighting back to state or national population estimates [22].

Small sample sizes for rare events

When estimating rare behaviors, even enormous surveys rely upon the responses of relatively few people. Lipari & Hughes [23] report trends in heroin use as measured by NSDUH from 2002 to 2013. We extend their figures to 2016. It appears to show a pronounced jump in 2006 for older users, from approximately 200 000 to 376 000 people.

A substantial chunk of that spike comes from just two respondents. Two older Hispanic males reported past-year use. Because of their age, they were assigned sample weights of 35 867 and 36 342, respectively, approximately eight times the average sample weight, so almost half of the apparently dramatic spike in heroin use in 2006 comes from the vagaries of the survey happening to pick up two older users. When we consider the uncertainty around the estimates due to the sheer rarity of reporting heroin use in NSDUH, the spikes are less meaningful. The vertical error bars in Fig. 1 are 95% confidence intervals based on bootstrap-estimated standard errors on the point estimates. These demonstrate that even the largest year-on-year changes in point estimates are rarely statistically different from the prior year's figure. NSDUH can document large and sustained changes retrospectively, but it is hard-pressed to detect reliably how any one year differs from the previous year.

This produces substantial fluctuations from year to year in point estimates, although seldom are the changes greater than the imprecision due to sampling error, which is typically 20–50% for any given point estimate. Consider, for example, past-month days of use, which is NSDUH's

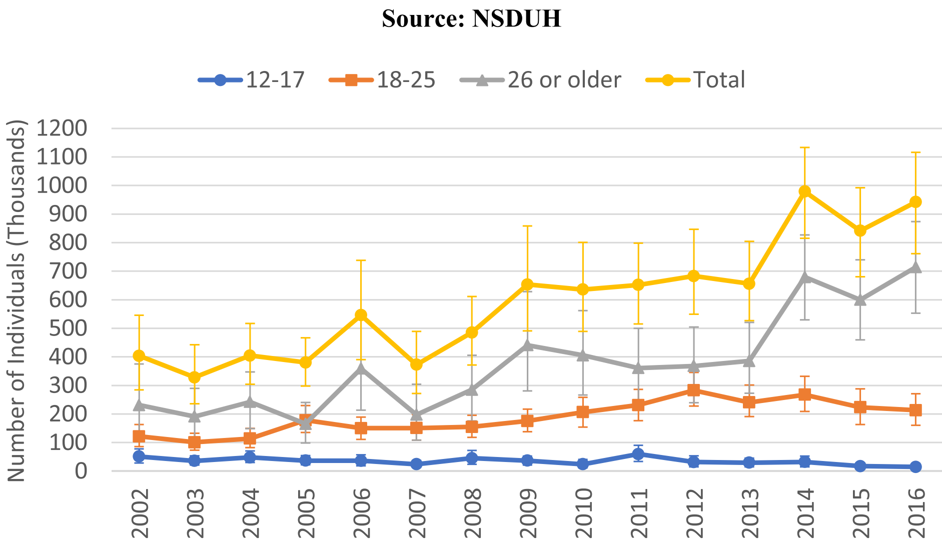

Figure 1 Annual estimates of past-year heroin use 2002–16 with 95% confidence intervals. Source: National Survey on Drug Use and Health (NSDUH)

best proxy for overall consumption. As Table 1 shows, between 2013 and 2016 the three year-on-year changes were up by 59%, down by 21% and then up by 37%. That would suggest great volatility, whereas in reality demand for heroin is distressingly stable precisely because demand is concentrated among frequent dependent users for whom substance use disorder is a chronic relapsing condition.

Sampling frame

One of NSDUH's strengths compared to GPS in some other countries is that it is no longer simply a 'household' survey; it includes adults living in noninstitutionalized group quarters (homeless shelters, rooming or boarding houses, dormitories, migratory work camps and halfway houses). In the 2010 census, this extended sampling frame included 300.8 of 308.7 million people in the United States. However, that glass is only 97% full. As NSDUH's estimate of past-month heroin use is so low (479 114 in 2016), if heroin use were common among the other 7.9 million people then NSDUH's estimate could be off substantially.

Two excluded groups may have large numbers of frequent heroin users: the unsheltered homeless and the incarcerated.

Under-reporting

Under-reporting can take two forms: (1) denying use altogether and (2) under-reporting the frequency of use, e.g. reporting use on 5 days in the past-month when the true number was 17.

One of the most rigorous studies of reporting accuracy compared self-report to NHSDA, the predecessor to NSDUH, with assays of urine and hair samples among 12–25-year-olds [18]. It did not draw conclusions about self-report of heroin for technical reasons, particularly drug tests' inability to distinguish heroin from other opioids, including medically sanctioned prescription pain relievers.

Table 1 Year-to year changes in various heroin measures (levels in thousands), 2013–2016.

	2013	*2014*	*% Change*	*2015*	*% Change*	*2016*	*% Change*
Past-year Users	656 (527 – 805)	979 (815 – 1133)	49%	841 (681 – 992)	−14%	942 (762 – 1116)	12%
Dependent	423 (324 – 540)	547 (433 – 672)	42%	559 (437 – 689)	2%	592 (474 – 718)	6%
Past-month users	264 (193 – 349)	476 (363 – 592)	80%	343 (247 – 448)	−28%	479 (358 – 609)	40%
PM days of Use	4295 (2876 – 6016)	6813 (4952 – 8861)	59%	5386 (3472 – 7359)	−21%	7371 (5144 – 10 064)	37%
Used 21+ days in PM (DND)	111 (64 – 171)	160 (100 – 229)	45%	147 (85 – 215)	−8%	211 (136 – 300)	44%

Source: National Survey on Drug Use and Health (NSDUH). PM = past month; DND = daily or near-daily. 95% confidence intervals calculated by the authors based on bootstrapped standard errors for each point estimate in brackets.

Table 2 Harrison *et al.*'s (2007) results contrasting self-report of cocaine use with urinalysis.

Self-reported use measure	*Response*	*Urinalysis result*		*Sensitivity*
		Negative	*Positive*	
Past 30 days (NSDUH core question)	No	97.9	1.1	
	Yes	0.6	0.3	0.206
Past 7 days (follow-up)	No	98.2	1.2	
	Yes	0.3	0.3	0.185
Past 3 days (follow-up)	No	98.4	1.2	
	Yes	0.1	0.2	0.163

NSUDH = National Survey on Drug Use and Health.

It is instructive, however, to consider the results pertaining to cocaine (crack and powder combined, because they cannot be distinguished by urinalysis). As with heroin, regular cocaine use is stigmatized, rare and problematic, so it should present similar reporting problems. Table 2 cross-tabulates the 3761 individuals in the study who tested negative (left column) or positive (right column) against their self-reported answers to three questions in successive pairs of rows: (1) the actual NHSDA core question about use in the last 30 days, (2) the supplement's question regarding use in the last 7 days and (3) use in the last 3 days. These last two questions were asked because conventional understanding of the detection window for cocaine by urinalysis is 72 hours or a little longer, but not 30 days.

The past-month self-report rate (0.9 or 1.0%) is not much below the rate of testing positive (1.4%). However, that is because some people self-report past-month use even though they test negative, e.g. because they used in the past month but not recently.

For present purposes, the more relevant question is what proportion of those who test positive, and so almost definitely had used cocaine recently, admitted that use. That proportion, which pertains to the survey's 'sensitivity', is only one in five or lower. In other words, the great majority of recent cocaine users do not self-report that use. This is not surprising; extensive research has shown a low willingness to report use of highly stigmatized drugs [24–26].

There is additional insight available on under-reporting in NSDUH by a critical population, namely criminally active heroin users; data on their use is available through the Arrestee Drug Abuse Monitoring program (ADAM). Lattimore *et al.* [27], analyzing NSDUH respondents who self-reported arrest, find that: 'estimates of past month illicit drug use among male arrestees at ADAM sites were 1.5 to 3 times as high using the ADAM data as the NSDUH data' (no page number).

Of course, selective non-response, under-reporting and sampling frame exclusions apply to other widely used surveys. The argument here is just that they may be particularly serious for estimates of frequent use of drugs such as heroin. Under-reporting seems particularly likely to generate under-estimates. Box 1 provides a short-list of key limitations and recommendations for consideration.

Box 1. Limitations of and recommendations for GPS for rare and stigmatized behaviors that are concentrated in hard-to-track populations

Limitation	*Recommendation*
Typical general population surveys (GPS) cannot be trusted to accurately estimate the prevalence or frequency of rare and stigmatized behaviors that are concentrated in populations that are difficult to survey	Asking about acquaintance's behavior, not the respondent's behavior, can help address stigma, but not other limitations of GPS
Year-to-year changes in those estimates are not reliable	Trends should be based on smoothing estimates over multiple years
Trends in overall prevalence may not mirror and can even mask more important trends in high-frequency or problem use.	Do not treat trends in past-year prevalence as a proxy for trends in demand or drug-related problems
Confidence intervals (CIs) based only on sampling variability can be misleading; the true figure is more likely to fall outside that range than is suggested by its name (e.g. falls outside the 95% CI more than 5% of the time) because of non-sampling errors such as selective non-response and under-reporting	If such confidence intervals are reported, they should be accompanied by explicit warnings since many readers and policy makers do not understand the difference between sampling and non-sampling errors

AN ALTERNATIVE ESTIMATE

A focus upon GPS for measuring frequent heroin use is puzzling, because there are other relevant indicators. We illustrate this by describing an estimation effort that took advantage of data that addressed two of NSDUH's most conspicuous failing; namely, its weak coverage of criminally active drug users and lack of biological confirmation. Kilmer *et al*. [12] estimated the number of people using heroin (as well as cocaine and methamphetamine) in the United States on four or more days in the previous month for the period 2000–10 using a range of data sources, including ADAM, with results published, among other places, in this journal [28].

ADAM was started under the name Drug Use Forecasting (DUF) in 1987 by the National Institute of Justice [29]. DUF/ADAM collected data, including urine specimens, from a sample of arrestees in booking facilities (typically county jails), growing to include 39 counties covering one-eighth of the national population in 2003. The survey was suspended during 2004–06, reinstated in 10 counties for 2007–11, then fell to five counties in 2012 and 2013 before its termination.

ADAM unnerves the punctilious. The locations are not a random sample of the nation's jails, simply a set of large counties whose agencies were willing to provide data collection access [30]. There were barriers to complete enumeration of all potential respondents (arrestees), so response rates are overestimates. The catchment area varied over time (e.g. from city to county), and sample weight calculations were always complex and sometimes problematic.

Nonetheless, ADAM measured what counts. It gathered data—including urine samples—from a population that is rich in frequent users. The proportion of arrestees who test positive for opioids, in large proportion heroin based on self-reports, is 50 times greater than the proportion of NSDUH respondents who report past-month heroin use—approximately one in 10, not just one in 500.

Some of this gap can be attributed to sampling frames and methods. For example, among ADAM respondents testing positive for drugs in 2013, more than 20% lived outside a household at the time of their arrest. Nonetheless, most of the discrepancy between self-reported past-month prevalence among criminal justice-involved populations in NSDUH and ADAM urinalysis results arises because many people who test positive for a drug do not self-report its use. This highlights how the problems of GPS go beyond sampling frame limitations.

Kilmer *et al*. [12] modeled the proportion of adult male arrestees who test positive in ADAM sites as a function of covariates that are available for (almost) all counties, including treatment admissions, drug-related deaths, work-place urinalysis testing and standard demographic variables. Combining that model with uniform crime reports (UCR) data on numbers of arrests projected the nation-wide number of adult male arrest events that would lead to a positive test. For 2010 the estimate (884 778) was just 3% higher than what simple arithmetic (arrests times average percentage positive tests in ADAM sites) would produce. It is reassuring that the sophisticated estimate concords with the simple version. Adjustments were then made to extrapolate to juveniles and women. Detailed descriptions of the multi-sourced model are provided in appendices to Kilmer *et al*. [31] and Midgette *et al*. [32]

The resulting prevalence estimates are much larger than those produced by NSDUH, and the gap is most pronounced for the most frequent users. Caulkins *et al*. [28] contrast Kilmer *et al*.'s estimate of 1.0 million daily or near-daily users against the mere 60 000 reported in NSDUH in 2010, a ratio of 16 : 1. The subsequent increases in heroin use recorded in NSDUH did not erase the gap. NSDUH's estimate of DND users rose to 111 000 by 2013, but a recent update of Kilmer *et al*. places the figure at 1.6 million [3], preserving the 16 : 1 ratio. In subsequent years, the discrepancy varies, but the difference remains an order of magnitude (Fig. 2).

That the estimate drawing upon multiple data sources is 16 times greater than the one based on NSDUH alone does not logically imply that NSDUH is wrong. Recognizing that it is NSDUH that errs requires common sense. For example, the Treatment Episode Data Set codebook reports 316 797 admissions to treatment in 2013 for which the primary substance of abuse was heroin [33]. While some heroin users who seek treatment cycle in and out and so may generate more than one admission per year, many others are not engaged with treatment at all, so it is difficult to imagine a population-wide average rate of three treatment admissions per year. Similarly, it is hard to square NSDUH's low estimates of DND heroin use with mortality data from the Centers for Disease Control and Prevention (CDC)'s Wide-ranging Online Data for Epidemiologic Research (WONDER) database (more than 15 000 in 2016), and WONDER's counts probably understate the true figure by 20–35% [34]. The estimates of DND users that draw upon multiple data sources are notably larger in magnitude, more consistent with other indicators and are more credibly smooth over time than the NSDUH equivalent. (Note the factor of 10 difference in the scales of the vertical axes in the left and right panels of Fig. 2.) The fact that the NSDUH DND estimates show three annual declines in the 6 years after 2010 is indicative of the implausibility of the annual estimates, given that heroin dependence is such a stable behavior.

This underestimation of frequent heroin use hinders understanding of the dynamics of the opioid epidemic. For example, it wreaks havoc with attempts to estimate what proportion of prescription opioid users escalate to problem

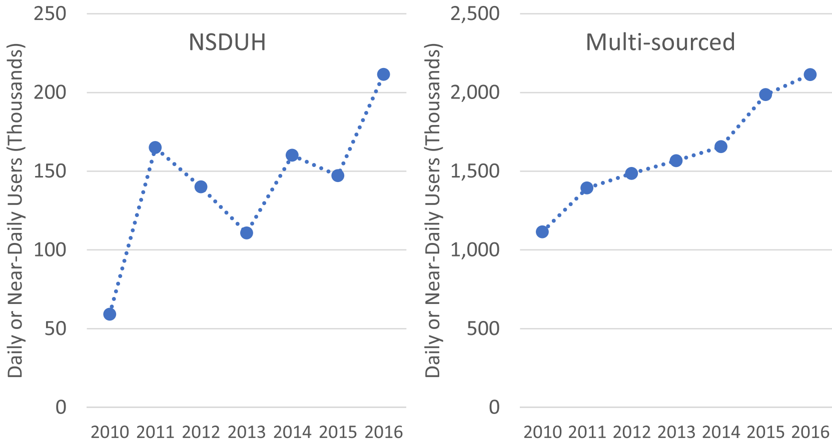

Figure 2 The multi-sourced estimates are smoother, and an order of magnitude larger, than the National Survey on Drug Use and Health (NSDUH) estimates of daily or near-daily (DND) users

heroin use [35]. If that denominator is well measured, because most people do not escalate and so are well represented in household surveys, but the numerator is off by fifteen-sixteenths, or 94%, then the actual escalation risk could be badly underestimated.

We offer the multi-sourced estimates not as the ideal, but as indicative of the limitations of any GPS. They strongly suggest that the NSDUH estimates of regular heroin use are irredeemable, even though ADAM has its own limitations. We are not advocating that blind deference to flawed NSDUH estimates be replaced with blind deference to ADAM or any other single data source. As we discuss next, the path forward may require stitching together multiple indicators.

A PATH FORWARD

General population surveys aid understanding of the epidemiology of substance use in every country. However, a GPS is doomed by its nature for estimating a rare and stigmatized behavior concentrated in a hard-to-track population. The high rate of under-reporting, the sampling frame omissions and selective non-response are fatal, rendering estimates so poor as to be misleading. They may be useful for providing bounds (e.g. Pepper [36]), although the bounds will be loose if the true prevalence is five or 10 times the reported rate. Similarly, multiplier methods can be used to adjust for under-reporting, but the potential error when scaling GPS-based estimates up by 1000% for frequent heroin use is greater than when scaling-up by 20% for overall cannabis use [37,38].

Undercounts can be so large that year-to-year changes lack credibility as a proxy for trends in the actual population, so GPS-based estimates do not permit establishing and monitoring goals for reducing frequent heroin use by a certain percentage, the way they have been used to formulate cannabis use reduction goals [39]; nor can they provide the denominator for such important parameters as the lethality of the drug on a population level [40].

Similar statements hold for regular use of cocaine and methamphetamine.

This problem is not confined to the United States. Many nations with less frequent and smaller GPS report to the UNODC cocaine, heroin and methamphetamine prevalence estimates based on household surveys. The UNODC's flagship publication (the *World Drug Report*) attempts to compensate for GPS weaknesses where it can, but for approximately half of all nations there are no alternative estimates [41]. The Global Burden of Disease estimates attempt to avoid GPS for estimating opioid, cocaine or amphetamine use, but where no other estimate is available the GBD project uses a fixed ratio of model to survey estimate, which is problematic given that the survey quality is so inconsistent across countries.

Creative integration of other data sources may yield a complementary path forward. The particulars will vary by country, but in the United States these additional data sources may include the National Survey of Substance Abuse Treatment Facilities (NSATTS) or the Healthcare Cost and Utilization Project (HCUP), both of which interact with populations of frequent users. Counts of fatal and non-fatal overdose are relevant, although the spread of fentanyl as a heroin adulterant and naloxone can greatly increase or decrease, respectively, the ratio of deaths per-person-year of use. New types of data, such as wastewater monitoring, may also be helpful [42,43]. Estimating consumption from metabolites found in municipal wastewater offers multiple advantages, including its focus upon quantity consumed, not just prevalence, and the ability to cheaply monitor changes over shorter time intervals than can annual surveys.

None of this is to suggest that GPS should be scrapped. GPS provide important information about the use of other, more common drugs and other health behaviors. For heroin and other opioids, however, they must be used more cautiously. At a minimum, interpretation should focus on trends, not levels, trust the direction of the trends more than the specific slope and, instead of focusing upon year-to-year fluctuations, reports should present something like 3-year moving averages, which are less sensitive to the very small number of heroin-using respondents in each annual survey. This is not an alien concept; NSDUH's state-specific estimates already are based upon pooling results from two survey years. Indeed, thinking in terms of 'replace data system X with data system Y' perhaps misses the most fundamental point. Estimating heroin use—particularly frequent or problematic use—is probably best performed by combining ('triangulating from') multiple data sources, and our confidence in the direction of an estimated trend should be stronger when multiple sources generally agree.

Rehm *et al*. [44] recently made the same point with respect to alcohol research, arguing that: 'we must divest ourselves of the idea that current non-probabilistic samples are always inferior to current probabilistic methods for deriving outcomes for a given population'. They argue explicitly for triangulation from traffic injuries, respondent-driven sampling studies and treatment rolls, among other sources; Robinson *et al*. [45] illustrate the feasibility of this for Scotland. Rehm *et al*. recognize that triangulation is easier when there is a gold standard anchor, such as alcohol excise tax receipts. There are no gold standard measures for frequent heroin use, but there may be multiple silver or bronze standards, including wastewater analysis, supply-side estimates and/or multiple capture/re-capture estimates, such as Bouchard *et al*. [46] recently produced for British Columbia.

In many countries, including the United States, additional means of drug use surveillance must be developed to have more to triangulate from. Collecting data on use by the criminally active population is vital, and only a handful of other countries (e.g. the Drug Use Monitoring in Australia system [47]) have invested in anything comparable. The new data collection need not come from a survey or other purpose-driven primary collection. Exploiting drug tests already administered as part of existing criminal justice processing or employment requirements are another option [3,48]. A recent *Addiction* paper [49] illustrated an innovative approach in Britain using treatment and mortality data to help to estimate the prevalence of opioid dependence. Perhaps, then, the key is not so much funding the one, right, data system but rather fostering a curious, professional and analytically sophisticated community that is dedicated to the task.

We are not the first to suggest supplementing GPS. Indeed, more than 20 years ago the office responsible for NSDUH published an article exploring what arrest and treatment data might add to the quality of estimates of 'hardcore' use [50]. More recently, local jurisdictions have also explored the possibility of supplementing NSDUH [51].

We end with a pointed query. Why has measuring frequent heroin use been given so little attention despite the extraordinary number of heroin-related overdose deaths, particularly in Canada and the United States but elsewhere as well? The HIV epidemic generated a massive improvement in monitoring systems for a variety of related behaviors, including efforts focused upon hard-to-reach, high-risk populations such as the National HIV Behavioral Surveillance (NHBS) in the United States With the welcome spread of effective treatments for AIDS and the troubling spread of fentanyl and other synthetic opioids, overdose is becoming as much of an issue as the spread of blood-borne diseases by injection use, so it is astonishing that comparable efforts have not been made to improve understanding of the number of individuals using heroin.

Declaration of interests

None.

Author contributions

Peter Reuter: Conceptualization; project administration; writing-original draft; writing-review & editing. **Jonathan Caulkins:** Conceptualization; formal analysis; visualization; writing-original draft; writing-review & editing. **Greg Midgette:** Data curation; formal analysis; methodology; visualization; writing-original draft; writing-review & editing.

References

1. Degenhardt L., Charlson F., Ferrari A., Santomauro D., Erskine H., Mantilla-Herrara A., *et al*. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet Psychiatry* 2018; **5**: 987–1012.
2. European Monitoring Centre for Drugs and Drug Addiction *European Drug Report 2020: Key Issues*. Luxembourg: Publications Office of the European Union; 2020.
3. Midgette G., Caulkins J. P., Davenport S., Kilmer B. G. *What America's Users Spend on Illegal Drugs, 2006–2016*. Report no.: RR3140. Santa Monica, CA: RAND Corporation; 2019.
4. United Nations Office on Drug and Crime. *World Drug Report 2020, Booklet 2*. Vienna: Research and Trends Analysis Branch; 2020. Available at: https://wdr.unodc.org/wdr2020/field/WDR20_Booklet_2.pdf (accessed 12 March 2021).
5. Compton W. M., Jones C. M., Baldwin G. T. Relationship between nonmedical prescription-opioid use and heroin use. *N Engl J Med* 2016; **374**: 154–63.
6. Ali M. M., Dowd W. N., Classen T., Mutter R., Novak S. P. Prescription drug monitoring programs, nonmedical use of prescription drugs, and heroin use: evidence from the

National Survey of drug use and health. *Addict Behav* 2017; **69**: 65–77.

7. McCance-Katz E., Houry D., Collins F. *Testimony on Addressing the Opioid Crisis in America: Prevention, Treatment, and Recovery before the Senate Subcommittee*. National Institutes of Health (NIH). 2017 [cited 2018 Aug 30]. Available at: https://www.nih.gov/about-nih/who-we-are/nih-director/testimony-addressing-opioid-crisis-america-prevention-treatment-recovery-before-senate-subcommittee (accessed 12 March 2021).
8. Substance Abuse and Mental Health Services Administration (US), Office of the Surgeon General (US). *Facing Addiction in America: The Surgeon General's Report on Alcohol, Drugs, and Health*. Washington (DC): US Department of Health and Human Services; 2016 [cited 2018 Aug 30]. Reports of the Surgeon General. Available at: http://www.ncbi.nlm.nih.gov/books/NBK424857/ (accessed 12 March 2021).
9. Hickman M., Larney S., Peacock A., Jones H., Grebely J., Degenhardt L. Competing global statistics on prevalence of injecting drug use: why does it matter and what can be done? *Addiction* 2018; **113**: 1768.
10. Center for Behavioral Health Statistics and Quality. *2016 National Survey on Drug Use and Health Public Use File*. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2017. Available at: https://www.datafiles.samhsa.gov/study/national-survey-drug-use-and-health-nsduh-2016-nid17184 (accessed 12 March 2021).
11. Degenhardt L., Grebely J., Stone J., Hickman M., Vickerman P., Marshall B. D. L. *et al*. Global patterns of opioid use and dependence: harms to populations, interventions, and future action. *Lancet* 2019; **394**: 1560–79.
12. Kilmer B., Sohler Everingham S. S., Caulkins J. P., Midgette G., Pacula R. L., Reuter P. H. *et al*. *What America's Users Spend on Illegal Drugs, 2000–2010*. Washington DC: Office of National Drug Control Policy; 2014 [cited 2018 Aug 30]. Available at: https://obamawhitehouse.archives.gov/sites/default/files/ondcp/policy-and-research/wausid_results_report.pdf (accessed 12 March 2021).
13. *National Drug Control Strategy [internet]*. Washington, DC: Office of National Drug Control Policy; 2016. Available at: https://obamawhitehouse.archives.gov/ondcp/policy-and-research/ndcs (accessed 12 March 2021).
14. Gfroerer J. C., Eyerman J., Chromy J. R. *Redesigning an Ongoing National Household Survey: Methodological Issues*. DHHS Publication no. SMA 03-3768. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies; 2002.
15. Gfroerer J. *War Stories from the Drug Survey: How Culture, Politics, and Statistics Shaped the National Survey on Drug Use and Health*. Cambridge: Cambridge University Press; 2018. Available at: https://www.cambridge.org/core/books/war-stories-from-the-drug-survey/66061FA755EF6B7BCFD1E5D8FAA4A44A (accessed 12 March 2021).
16. Cook P. J. *Paying the Tab: The Costs and Benefits of Alcohol Control*. Princeton, NJ: Princeton University Press; 2011.
17. Harrison L. D. The validity of self-reported data on drug use. *J Drug Issues* 1995; **25**: 91–111.
18. Harrison L. D., Martin S. S., Enev T., Harrington D. *Comparing Drug Testing and Self-Report of Drug Use Among Youths and Young Adults in the General Population*. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies; 2007.
19. Johnson T. P. Sources of error in substance use prevalence surveys. *Int Sch Res Notices* 2014; **2014**: 1–21.
20. National Research Council. *Nonresponse in Social Science Surveys: A Research Agenda*. Washington, DC: The National Academies Press; 2013. Available at: https://www.nap.edu/catalog/18293/nonresponse-in-social-science-surveys-a-research-agenda (accessed 12 March 2021).
21. Gfroerer J., Lessler J., Parsley T. Studies of nonresponse and measurement error in the National Household Survey on drug abuse. *Nat Inst Drug Abuse Res Monogr* 1997; **167**: 273–95.
22. Midgette G., Reuter P. Has cannabis use among youth increased after changes in its legal status? A commentary on use of monitoring the future for analyses of changes in state cannabis Laws. *Prev Sci* 2020; **21**: 137–45.
23. Lipari R. N., Hughes A. Trends in Heroin Use in the United States: 2002 to 2013. In: *The CBHSQ Report*. Rockville, MD: Substance Abuse and Mental Health Services Administration (US); 2013 [cited 2018 Aug 30]. Available at: http://www.ncbi.nlm.nih.gov/books/NBK343534/ (accessed 12 March 2021).
24. Fendrich M., Johnson T. P., Sudman S., Wislar J. S., Spiehler V. Validity of drug use reporting in a high-risk community sample: a comparison of cocaine and heroin survey reports with hair tests. *Am J Epidemiol* 1999; **149**: 955–62.
25. Magura S. Validating self-reports of illegal drug use to evaluate National Drug Control Policy: a reanalysis and critique. *Eval Program Plann* 2010; **33**: 234–7.
26. Hunt D. E., Kling R., Almozlino Y., Jalbert S., Chapman M. T., Rhodes W. Telling the truth about drug use: how much does it matter? *J Drug Issues* 2015; **45**: 314–29.
27. Lattimore P. K., Steffey D. M., Gfroerer J., Bose J., Pemberton M. R., Penne M. A. Arrestee Substance Use: Comparison of Estimates from the National Survey on Drug Use and Health and the Arrestee Drug Abuse Monitoring Program. In: *CBHSQ Data Review*. Rockville, MD: Substance Abuse and Mental Health Services Administration (US); 2012 [cited 2018 Aug 30]. Available at: http://www.ncbi.nlm.nih.gov/books/NBK390289/ (accessed 12 March 2021).
28. Caulkins J. P., Kilmer B., Reuter P. H., Midgette G. Cocaine's fall and marijuana's rise: questions and insights based on new estimates of consumption and expenditures in US drug markets. *Addiction* 2015; **110**: 728–36.
29. DeJong C., Wish E. D. Is it advisable to urine test arrestees to assess risk of rearrest? A comparison of self-report and urinalysis-based measures of drug use. *J Drug Issues* 2000; **30**: 133–46.
30. Golub A., Johnson B. D., Taylor A., Liberty H. J. The validity of arrestees' self-reports: variations across questions and persons. *Justice Q* 2002; **19**: 477–502.
31. Kilmer B., Sohler Everingham S. S, Caulkins J. P., Midgette G., Pacula R. L., Reuter P. H. *et al*. *What America's Users Spend on Illegal Drugs, 2000–2010, Technical Report*. Washington, DC: Office of National Drug Control Policy; 2014 [cited 2018 Aug 30]. Available at: https://obamawhitehouse.archives.gov/sites/default/files/ondcp/policy-and-research/wausid_technical_report.pdf (accessed 12 March 2021).
32. Midgette G., Caulkins J. P., Davenport S., Kilmer B. G. *What America's Users Spend on Illegal Drugs, 2006–2016, Appendix B*. Santa Monica, CA: RAND Corporation; 2019. Report No.: RR3140. Available at: https://www.rand.org/pubs/research_reports/RR3140.html (accessed 12 March 2021).

33. Center for Behavioral Health Statistics and Quality. *Treatment Episode Data Set—Admissions (TEDS-A), 2013 Codebook*. United States Department of Health and Human Services, Substance Abuse and Mental Health Services Administration; 2015. Available at: https://www.datafiles.samhsa.gov/sites/default/files/field-uploads-protected/studies/TEDS-A-2013/TEDS-A-2013-datasets/TEDS-A-2013-DS0001/TEDS-A-2013-DS0001-info/TEDS-A-2013-DS0001-info-codebook.pdf (accessed 12 March 2021).
34. Ruhm C. J. Corrected US opioid-involved drug poisoning deaths and mortality rates, 1999–2015. *Addiction* 2018; **113**: 1339–44.
35. Cerdá M., Santaella J., Marshall B. D. L., Kim J. H., Martins S. S. Nonmedical prescription opioid use in childhood and early adolescence predicts transitions to heroin use in young adulthood: a National Study. *J Pediatr* 2015; **167**: 605–12 e2.
36. Pepper J. V. How do response problems affect survey measurement of trends in drug use? In: Manski C. F., Pepper J. V., Petrie C. V., editors. *Informing America's Policy on Illegal Drugs: What We Don't Know Keeps Hurting Us*. National Academies Press; 2001, pp. 321–48.
37. Cuellar M. Trends in self-reporting of marijuana consumption in the United States. *Stat Public Policy* 2018; **5**: 1–10.
38. Kilmer B., Caulkins J. P., Pacula R. L., Reuter P. Bringing perspective to illicit markets: estimating the size of the US marijuana market. *Drug Alcohol Depend* 2011; **119**: 153–60.
39. National Drug Control Strategy. Washington, DC: Office of National Drug Control Policy; 1989. Available at: https://www.ncjrs.gov/pdffiles1/ondcp/119466.pdf (accessed 12 March 2021).
40. Pardo B., Reuter P. Narcotics and drug abuse. *Criminol Public Policy* 2018; **17**: 419–36.
41. United Nations Office on Drug and Crime. *World Drug Report 2019 Methodology Report*. Vienna: Research and Trends Analysis Branch; 2019. Available at: https://wdr.unodc.org/wdr2019/prelaunch/WDR-2019-Methodology-FINAL.pdf (accessed 12 March 2021).
42. *European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Wastewater Analysis and Drugs: a European Multi-City Study*. Luxembourg: Publications Office of the European Union: EMCDDA; 2018.
43. Zuccato E., Castiglioni S., Senta I., Borsotti A., Genetti B., Andreotti A., *et al.* Population surveys compared with wastewater analysis for monitoring illicit drug consumption in Italy in 2010–2014. *Drug Alcohol Depend* 2016; **161**: 178–88.
44. Rehm J., Kilian C., Rovira P., Shield K. D., Manthey J. The elusiveness of representativeness in general population surveys for alcohol. *Drug Alcohol Rev* 2020; **40**: 161–5. https://doi.org/10.1111/dar.13148
45. Robinson M., Kibuchi E., Gray L., McCartney G. Approaches to triangulation of alcohol data in Scotland: commentary on Rehm *et al*. *Drug Alcohol Rev* 2020; **40**: 173–5. https://doi.org/10.1111/dar.13164
46. Bouchard M., MacDonald M., Ponce C., Milloy M.-J., Hayashi K., DeBeck K. *Estimating the size of the fentanyl market in British Columbia. Report to the Cullen Commission*. Vancouver, BC: Cullen Commission; 2020.
47. Makkai T. *Drug Use Monitoring in Australia (DUMA): a Brief Description*, Vol. **21**. Canberra: Australian Institute of Criminology; 1999.
48. Wish E. D., Billing A., Artigiani E. E. *Community Drug Early Warning System: The CDEWS-2 Replication Study*. Washington, DC: Office of National Drug Control Policy; 2015. Available at: https://cesar.umd.edu/sites/cesar.umd.edu/files/pubs/cdews2_finalreport_4_8_15v3_replication.pdf (accessed 12 March 2021).
49. Jones H. E., Harris R. J., Downing B. C., Pierce M., Millar T., Ades A. E., *et al.* Estimating the prevalence of problem drug use from drug-related mortality data. *Addiction* 2020; **115**: 2393–404. https://doi.org/10.1111/add.15111
50. Wright D., Gfroerer J., Epstein J. The use of external data sources and ratio estimation to improve estimates of hardcore drug use from the NHSDA. *NIDA Res Monogr* 1997; **167**: 477–97.
51. McNeely J., Gourevitch M. N., Paone D., Shah S., Wright S., Heller D. Estimating the prevalence of illicit opioid use in New York City using multiple data sources. *BMC Public Health* 2012; **12**: 443.

