Drug and Alcohol Dependence 241 (2022) 109667

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

Most recent, modal, or median heroin purchase: Does it matter when estimating market size?

Todd A. Olmstead [a,*], Rosalie Liccardo Pacula [b], Sheila M. Alessi [c], James Scott [d]

[a] Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin, 2300 Red River Street, Austin, TX 78713, United States
[b] Sol Price School of Public Policy, University of Southern California, 635 Downey Way, Los Angeles, CA 90089, United States
[c] University of Connecticut School of Medicine, 263 Farmington Avenue, Farmington, CT 06030, United States
[d] McCombs School of Business, The University of Texas at Austin, 2110 Speedway, Austin, TX 78705, United States

ARTICLE INFO

Keywords:
Drug markets
Market size
Drug purchases
Most recent purchase
Modal purchase
Median purchase
Heroin

ABSTRACT

Background: Assessing the size of illicit drug markets is an important activity of many government agencies; however, the expenditure-based method for estimating market size relies on the relatively untested assumption that the cash value of the most recent purchase is representative of the average purchase amount. Using panel data, we test the representativeness of the most recent, modal and median purchase compared to the average purchase amount.

Methods: Data were drawn from a prior study that collected daily transaction-level purchase data from a sample of 120 people who were using heroin regularly. The same study participants completed two distinct two-week waves of data collection, separated by six months. T-tests and bootstrapping were used to detect differences within each wave between the average cash value of participant heroin purchases and the cash value of their most recent, modal and median heroin purchases.

Results: In both waves, we found (a) no evidence that the expected value of the most recent purchase differs from the expected value of the average purchase, and (b) the expected values of the modal and median purchases are smaller than the expected value of the average purchase. These results imply that estimates of total market size based on the modal or median purchase will suffer from a significant downward bias, but that estimates based on the most recent purchase will be unbiased.

Conclusions: We provide evidence in support of using the most recent (but not the modal or the median) purchase to estimate market size for heroin.

1. Introduction

Assessing the size of illegal drug markets is an important activity of many U.S. and foreign government agencies. Knowledge of the size and scale of illicit markets has been used to inform decisions regarding supply reduction strategies, law enforcement interventions, security needs, and more recently in the case of opioids, the need and availability of treatment resources (EMCDDA, 2019; INCSR 2020a, 2020b; USDEA, 2018; UNODC, 2021). It is also helpful for understanding the harms that may affect a community due to the presence and nature of the illicit market operating in or passing through them (EMCDDA, 2019; UNODC, 2021).

As discussed in great detail by Kilmer et al. (2011), there are four general approaches for estimating the size of illicit drug markets used by various government agencies. Two rely on supply-side approaches, making use of either plant production formulations (acreage of land that can be used for cultivation and expected yields for drugs derived from plants) or seizures (dividing total seizures in a given year by some assumed proportion of shipments). The other two approaches rely on demand-side approaches, using self-reported consumption information (counts of people using at particular frequencies and then multiplying these counts by typical amounts consumed for that group) or expenditure (which combine information on number of people who use drugs with information on amount spent per year rather than amount used). There are important limitations to each of these approaches, and many agencies today rely on triangulation of data from multiple approaches

* Corresponding author.
E-mail addresses: tolmstead@austin.utexas.edu (T.A. Olmstead), rmp_302@healthpolicy.usc.edu (R.L. Pacula), salessi@uchc.edu (S.M. Alessi), james.scott@mccombs.utexas.edu (J. Scott).

https://doi.org/10.1016/j.drugalcdep.2022.109667
Received 1 June 2022; Received in revised form 7 October 2022; Accepted 13 October 2022
Available online 20 October 2022

(Kilmer et al., 2014; ONDCP, 2014; UNODC, 2021), but given the growth in survey methods and synthetic drugs the past two decades, demand side estimates have been receiving significantly more attention among academics (Armstrong, 2021; Caulkins et al., 2015, 2019; Kilmer et al., 2013; Kilmer and Pacula, 2009; van Laar et al., 2013; Wilkins et al., 2002).

Ongoing work continues to evaluate and bound some of the measurement issues involved in various steps of constructing demand-side estimates, including issues related to missing populations from key population surveys (Caulkins et al., 2015; Reuter et al., 2021), underreporting of any use (Delaney-Black et al., 2010; Harrison et al., 2007; Kilmer et al., 2011), and reported number of use days (Caulkins et al., 2015; Kilmer and Pacula, 2009). Recent methods generating the size of illicit drug markets try to incorporate these important sources of uncertainty and bias (Caulkins et al., 2015; Midgette et al., 2019).

One metric that has received insufficient consideration due to limited data availability is the use of self-reported measures of average (and/or "typical") consumption or purchase amounts. The expenditure-based method, for example, was revised in recent years to include information on "last purchase" rather than "typical" or "average" purchase, because in 2000 the Arrestee Drug Abuse Monitoring (ADAM) survey, on which many US expenditure estimates were built, changed the question asked of arrestees from "how much money do you spend in an average week for your drugs" to "how much money did you spend on (drug X) the last time you purchased it" (Golub and Johnson, 2004). The motivation for the change in this question was the belief that people who use drugs (especially heavily) have an easier time remembering accurately what they did the last time they engaged in the market than they do their behavior over a full episode of time, including the past 7 days. In their evaluation of the impact of this change, Golub and Johnson (2004) show that estimates of the value of total expenditure using information on last purchase from 2000-2002 data did in fact generate larger values than those based on estimates of average purchase amounts from the 1998–1999 data, as there were many individuals reporting larger values. They were unable to speak to the accuracy of these different estimates for generating actual expenditure, however, as they did not know what actual expenditure was nor were the same questions regarding typical and last purchase asked of the same individuals in each wave.

Bond et al. (2014) revisit the question of whether the last purchase is a reasonable approximation of typical purchase among people who use heroin by examining responses to a series of questions about the last 3 heroin purchases administered to a longitudinal cohort of people who were injecting heroin in Australia. They find that, contrary to Golub and Johnston, the most recent purchase was generally representative of the other purchases reported by the respondent. However, as the authors note, the data on last three purchases were obtained through self-report on the same survey instrument, and hence may suffer from recall error when reported.

The current study makes several important contributions to the literature examining the accuracy of total expenditure estimates by assessing within a stable population of 120 people who were using heroin regularly in Northeast U.S. the daily purchases made during a full two-week period. While still based on self-report, study participants were contacted every day and reported via electronic means the heroin purchases made since the previous call. We therefore capture purchase behavior within close proximity of when it happened for the same individuals over time. Using these repeated observations over a narrow period of time, we test the representativeness of the "most recent" (last), modal and median purchase compared to the average purchase amount. We also examine the nature and magnitude of the bias generated from market-level estimates based on self-reported estimates of the most recent, modal and median purchase for this group of study participants, which can help inform researchers about the limitations of using such measures in the development of expenditure-based estimates.

2. Methods

Data were drawn from a prior study (Olmstead et al., 2015) that collected transaction-level purchase data from people who were using heroin regularly for the purpose of estimating the price-elasticity of demand for heroin. Relevant details of the elasticities study are described below, followed by the methods used to assess the representativeness of the most recent, modal and median heroin purchase. Data collection occurred between September 24, 2010 and April 2, 2012. Participants provided written informed consent approved by the University of Connecticut Health Center Institutional Review Board. Participants were informed that all information was confidential (via a Certificate of Confidentiality from the National Institute of Drug Abuse), including drug purchase data.[1]

2.1. Study sample and data

Our study sample comprised 120 people who were using heroin regularly (i.e., self-reported heroin use at least once per week during the past six months), recruited from the greater Hartford, CT area using radio and newspaper advertisements, flyers distributed at low-income housing projects and social service agencies, and our contacts with community-based substance abuse treatment programs. We excluded people who reported "dealing drugs" as one of their top 3 sources of income.[2] Other exclusion criteria included age less than 18, non-English speaking, active psychosis, suicidality, and inability to comprehend the study.

The same study participants completed two distinct two-week waves of data collection, separated by six months. During each wave of data collection, transaction-level heroin purchase data were collected via an Interactive Voice Response (IVR) telephone system that made daily calls for a period of two weeks to participants on their study-provided cellphones and asked them to report all heroin purchases made since the previous call.[3] Data collected for each heroin purchase included total purchase price, units (bags, grams, etc.), number of units, and percentage for own use.

The IVR system called participants each day at a designated time specified by the participants. Participants were allowed (and encouraged) to call back if they were unavailable when the IVR call arrived. Compliance was monitored, and if a participant did not complete two consecutive calls, a research assistant contacted the participant to re-engage them in the study. Participants received extensive training during the orientation session. Unrealistic responses were reconciled with the participant within 24 h (whenever possible), and participants were compensated to encourage compliance with study procedures.[4]

During wave 1, participants completed a total of 1490 calls during the 14-day period. However, 6 of these calls were discarded due to either missing or suspect responses that could not be reconciled with the

[1] Certificates of Confidentiality are issued by the National Institutes of Health to assure the privacy of research subjects by protecting researchers and their institutions from being compelled to disclose identifying information about the subjects (NIH, 2021).

[2] 4% of the people screened for participation in the study were excluded due to "dealing drugs".

[3] IVR telephone systems allow participants to answer recorded questions by pressing buttons on their telephone keypad. Such systems typically yield high quality data, including drug use data, compared with interviews and questionnaires (Abu-Hasaballah et al., 2007; Tucker et al., 2012).

[4] Participants received $2.50 for each completed IVR call, a $15 bonus for completing 12 or 13 out of the 14 calls or a $25 bonus for completing all 14 calls, and a $25 bonus for returning cellphones in working order.

participant within 24 h, resulting in a total of 1484 calls covering 1587 (94%) out of a possible 1680 (120 participants X 14 days) participant-days.[5] At least one heroin purchase was reported on 926 (62%) of these calls, describing purchases on 994 days. Since some days included more than one purchase, the final wave 1 sample included 1341 separate heroin purchases. Ninety-one of the 120 participants from wave 1 returned for wave 2. Wave 2 participants completed a total of 1128 calls (one wave 2 call was discarded due to missing/suspect responses) covering 1215 (95%) out of a possible 1274 (91 participants X 14 days) participant-days. At least one heroin purchase was reported on 532 (47%) of these calls, describing purchases on 560 days and resulting in a final wave 2 sample of 727 separate heroin purchases.[6]

2.2. Analysis strategy

We analyzed both waves of data with the same methods (described below). Detailed results are presented in Section 3 for the wave 1 analysis; the wave 2 findings were similar and available in supplementary materials.

To focus on the "final" demand for heroin, for each transaction we scale the total purchase price by the percentage that the participant intended for own use. In what follows, all references to cash values refer to the amount spent by participants on heroin *for their own use*.

For participant i, we calculated the following summary measures within a given wave:

1. The number (N_i), total cash value (T_i), and average cash value ($\overline{V}_i$)[7] of purchases.
2. The cash value of the participant's most recent purchase,[8] denoted R_i.
3. The modal purchase, i.e. the most common cash value among all purchases (with ties broken randomly), denoted C_i.
4. The median purchase, i.e. the middle cash value in a sorted list of all purchases, denoted Med_i.

Let S denote the total number of participants. Then the total market size, M, can be operationally defined as the total cash value of purchases across all S participants during the 14-day period.[9] Moreover, it is clear that the total market size is mathematically related to each participant's average purchase amount, as follows:

$$M = \sum_{i=1}^{S} T_i = \sum_{i=1}^{S} N_i \overline{V}_i.$$

This emphasizes that the key quantity for each participant is their average purchase value $\overline{V}_i$. If these averages can be estimated in an unbiased way, then we can also produce an unbiased estimate of the total market size. Said mathematically, suppose we have an estimator $\widehat{V}_i$ for the average purchase size for participant i, such that $E(\widehat{V}_i - \overline{V}_i) = 0$, and consider the estimator of total market size given by:

$$\widehat{M} = \sum_{i=1}^{S} N_i \widehat{V}_i.$$

If $E(\widehat{V}_i - \overline{V}_i) = 0$, then $E(\widehat{M}) = M$, and we have an unbiased estimator of market size.[10] Therefore, a natural question is whether we can plausibly consider a participant's most-recent purchase to be an unbiased estimator of their average purchase, i.e. whether we can reject the hypothesis that $E(R_i - \overline{V}_i) = 0$.

To address this, we calculated the difference between each participant's most recent purchase and their average purchase, and we conducted a one-sample t-test for whether $E(R_i - \overline{V}_i) = 0$, i.e. whether the mean difference was zero. One potential issue with the t-test, however, is that the participants' differences $(R_i - \overline{V}_i)$ are not homoscedastic, since each $\overline{V}_i$ is calculated using N_i observations. Therefore the simple standard-error formula used in a one-sample t-test is incorrect. Note that heteroskedasticity would tend to make the t-test anti-conservative, i.e. more likely to falsely reject a true null hypothesis; since our test failed to reject the null (see Section 3.3), anti-conservatism is clearly not a problem. Nonetheless, in the interest of robustness, we also performed a bootstrap-based test that straightforwardly accommodates heteroskedasticity in the participants' differences. This test, like a t-test, assesses whether it is plausible that $E(R_i - \overline{V}_i) = 0$, but easily accounts for the fact that N_i varies across participants. We describe this bootstrapping procedure in the technical appendix, and we note that this same procedure also yields a standard error for the estimator of total market size based on each participant's most recent purchase:

$$\widehat{M}_{Recent} = \sum_{i=1}^{S} N_i R_i.$$

Next, we asked the same question of a participant's modal and median purchases. That is, can we plausibly consider a participant's modal purchase (C_i) or median purchase (Med_i) to be an unbiased estimator of their average purchase? Mathematically, this entails testing whether $E(C_i - \overline{V}_i) = 0$ and whether $E(Med_i - \overline{V}_i) = 0$. To do so, we used the same approach outlined above, running both a one-sample t-test and our test based on the bootstrap.

3. Results

3.1. Demographics

The mean±SD age of the full sample (N = 120) was 46 ± 9 years; 58% (N = 69) were male. Approximately half of the individuals were non-Hispanic, non-African American (53%, N = 64); 42% (N = 50) were non-Hispanic, African American. Fifty-four percent (N = 65) had never been married and 28% (N = 33) were divorced. The mean±SD years of education was 12 ± 2. Forty-nine percent (N = 59) were unemployed and 38% (N = 46) were not in the labor force. The vast majority (86%, N = 102) had received treatment for substance use disorder during their lifetime, and 65% (N = 78), 38% (N = 45), and 21% (N = 25) had a lifetime diagnosis of cocaine, alcohol, or marijuana dependence, respectively. The mean±SD number of days of heroin use per week during the 30 days prior to wave 1 was 4.8 ± 2.1. The mean ±SD Mental Health Inventory (MHI-5) score was 20 ± 3 and the mean

[5] The IVR system asked participants to report all drug purchases made since the previous call. Occasionally, participants did not respond to an IVR call on a given day, and so the next call would contain purchase data covering more than one day.

[6] Combining waves, there was an average of 0.74 (i.e., (1341 + 727)/(1587 + 1215)) heroin purchases per participant-day during the study; extrapolating, this is equivalent to 270 (i.e., 0.74×365) heroin purchases per participant-year.

[7] For participant i, $\overline{V}_i$ was defined as $\frac{T_i}{N_i}$.

[8] A participant's most recent purchase was the last purchase reported on their last completed IVR call.

[9] Because the market is a population-based measure, it may be tempting to think that our definition of total market size would need to be scaled from our sample (S) to the population. However, because the purpose of our analysis is to compare three different estimators derived from the same sample, that scaling factor would drop out in the analysis of differences.

[10] Technically this holds only if there is no correlation between the total number of purchases, N_i, and the error $(\widehat{V}_i - \overline{V}_i)$. But this assumption seems correct in light of the data. For example, we find no significant evidence of correlation between N_i and $(R_i - \overline{V}_i)$ across the entire sample for either wave 1 or wave 2 (see Section 3.3 and supplementary materials).

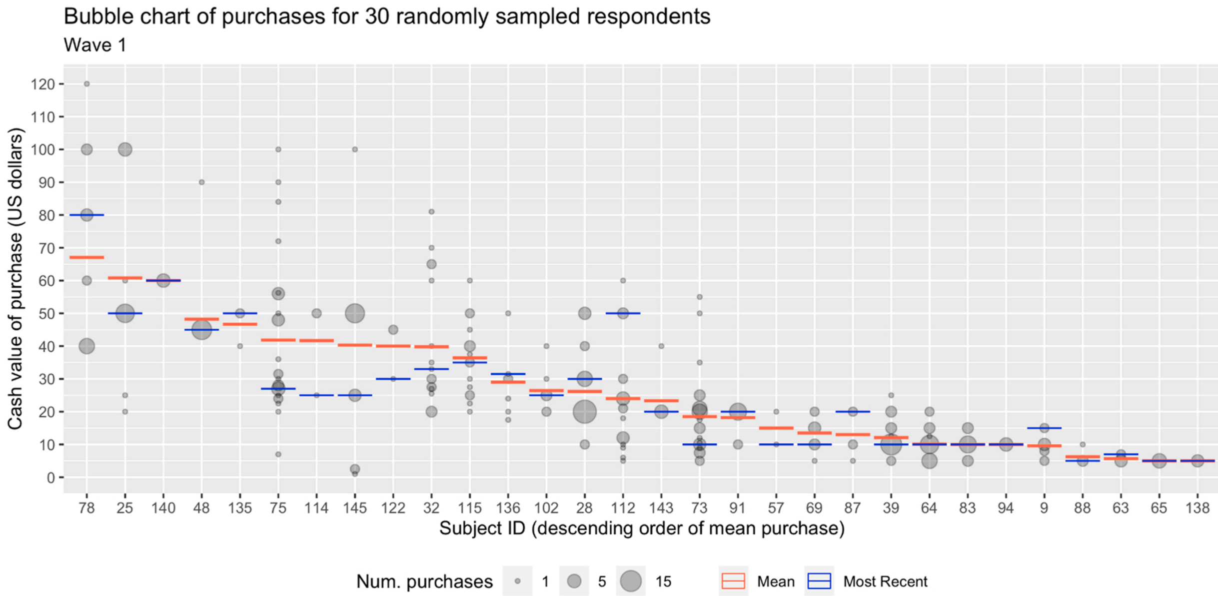

Fig. 1. Participants' individual purchases are shown in a bubble chart, with the size of the bubble proportional to the number of purchases at a specific dollar value. The mean and most recent purchase for each participant are also shown. Participants are ordered from left to right in descending order of mean purchase. Due to space constraints only 30 randomly sampled participants are included in the chart, but a version of this figure containing all participants is available as a supplementary file.

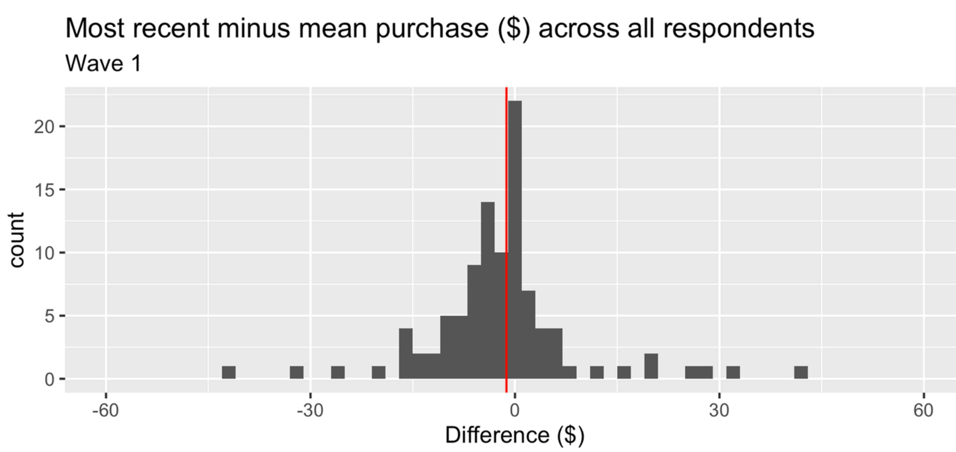

Fig. 2. For each participant, we calculated the difference between their most recent purchase and their average purchase. This histogram shows the empirical distribution of these differences across all participants. The mean of this distribution (-$1.25) is not significantly different from 0 (p-value = 0.37).

±SD delay discounting rate was 0.08 ± 0.09.[11]

3.2. Summary of individual purchases – Wave 1

Fig. 1 shows a bubble chart of participants' individual purchases during wave 1, along with their mean and most recent purchase.[12] The mean±SD cash value of the participants' mean, most recent, modal and median purchases was $25.75 ± $18.83, $24.50 ± $21.76, $22.61 ± $17.88, and $24.04 ± $$18.12, respectively.

3.3. Representativeness of most recent, modal, and median purchase – Wave 1

Fig. 2 shows a histogram of the difference between *most recent* purchase and mean purchase across all wave 1 participants. There is little evidence of a bias in this histogram: the sample mean of $R_i - \overline{V}_i$ is -$1.25; both the one-sample t-test and our bootstrap-based test *failed to reject* the null hypothesis that $E(R_i - \overline{V}_i) = 0$, returning p-values of 0.337 and 0.37, respectively.

Fig. 3 shows a histogram of the difference between *modal* purchase and mean purchase across all wave 1 participants. Despite similarities to Fig. 2, the sample mean of $C_i - \overline{V}_i$ is -$3.14, and both the one-sample t-test and our bootstrap-based test *rejected* the null hypothesis that $E(C_i - \overline{V}_i) = 0$, returning p-values of 0.004 and 0.0008, respectively,

[11] The MHI-5 is a 5-item subscale of the general health assessment SF-36 with items about depression and anxiety (Berwick et al., 1991; Ware et al., 2007); scores range from 0 (poor mental health) to 100 (good mental health). The delay discounting rate, k, was measured following Kirby et al. (1999) in which a hyperbolic function was fit to data generated by participant responses to questions offering them hypothetical choices of varying amounts of money today versus larger amounts 2–300 days in the future; higher rates indicate more impulsive choices.

[12] Due to space constraints, only 30 randomly sampled participants are included in Fig. 1, but a version containing all respondents is available in the supplementary materials (Figure S.1).

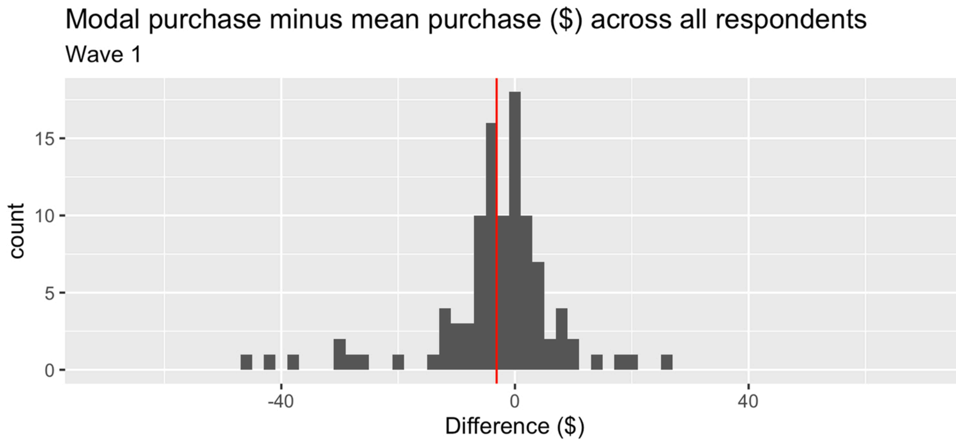

Fig. 3. For each participant, we calculated the difference between their modal and their average purchase. This histogram shows the empirical distribution of these differences across all participants. The mean of this distribution (-$3.14) is significantly different from 0 (p-value = 0.0008).

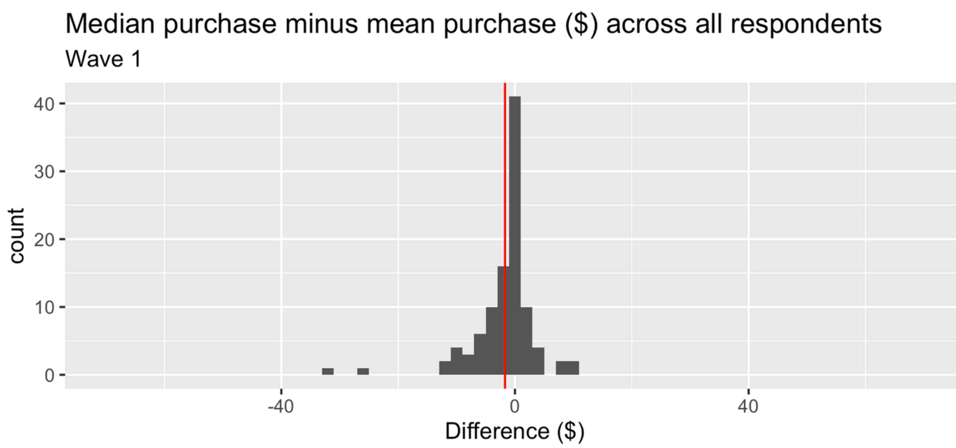

Fig. 4. For each participant, we calculated the difference between their median and their average purchase. This histogram shows the empirical distribution of these differences across all participants. The mean of this distribution (-$1.71) is significantly different from 0 (p-value = 0.00007).

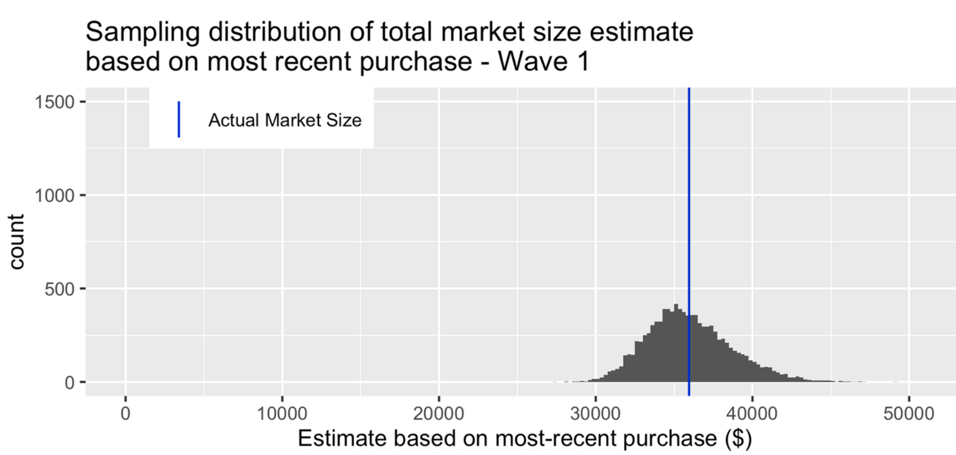

Fig. 5. The bootstrapped sampling distribution of total market size based on a single randomly sampled purchase from each participant. By construction this is an unbiased estimator of the total market size (blue line, $35,954). The estimate of total market size based on each participant's most recent purchase ($35,960, error = $6) is indistinguishable from the actual market size in this figure.

indicating that the modal purchase is downwardly biased as an estimate of a participant's average purchase. Similarly, Fig. 4 shows a histogram of the difference between *median* purchase and mean purchase across all wave 1 participants. The sample mean of $\mathrm{Med_i} - \overline{V}_i$ is -$1.71, and both the one-sample t-test and our bootstrap-based test *rejected* the null hypothesis that $E(\mathrm{Med_i} - \overline{V}_i) = 0$, returning p-values of 0.003 and 0.00007, respectively, indicating that the median purchase is also downwardly biased as an estimate of a participant's average purchase. These biases are to be expected, given that typical participant's purchase size distribution (in dollars) is right-skewed (as seen in Fig. 1).

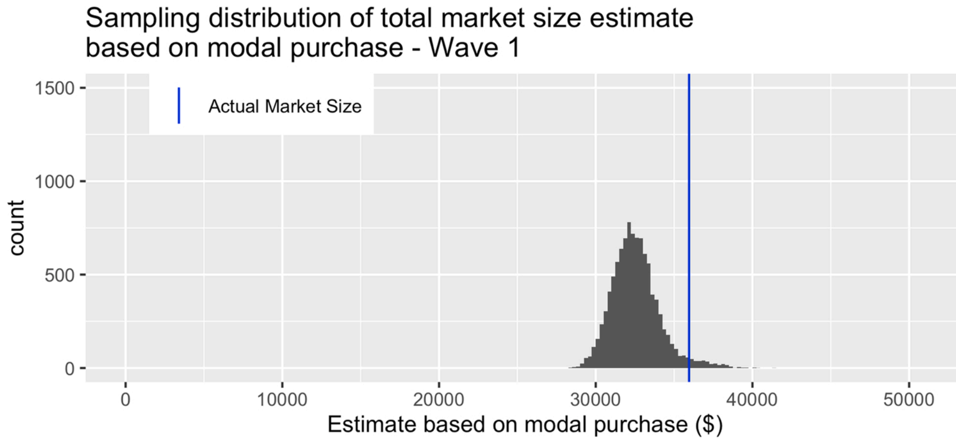

Fig. 6. The bootstrapped sampling distribution of total market size based on each participant's modal purchase (see technical appendix for details of the bootstrapping procedure). The estimate of total market size based on each participant's modal purchase is $32,162 (error = -$3792 or −10.55%).

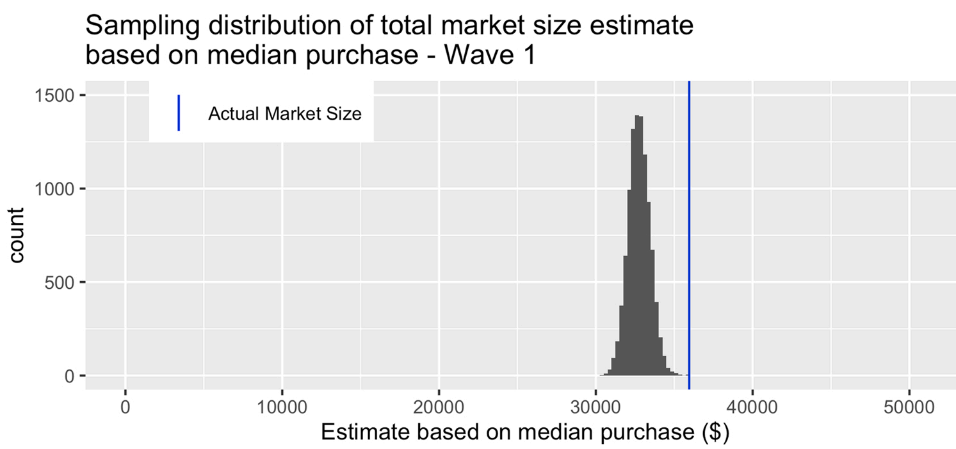

Fig. 7. The bootstrapped sampling distribution of total market size based on each participant's median purchase (see technical appendix for details of the bootstrapping procedure). The estimate of total market size based on each participant's median purchase is $32,704 (error = -$3250 or −9.03%).

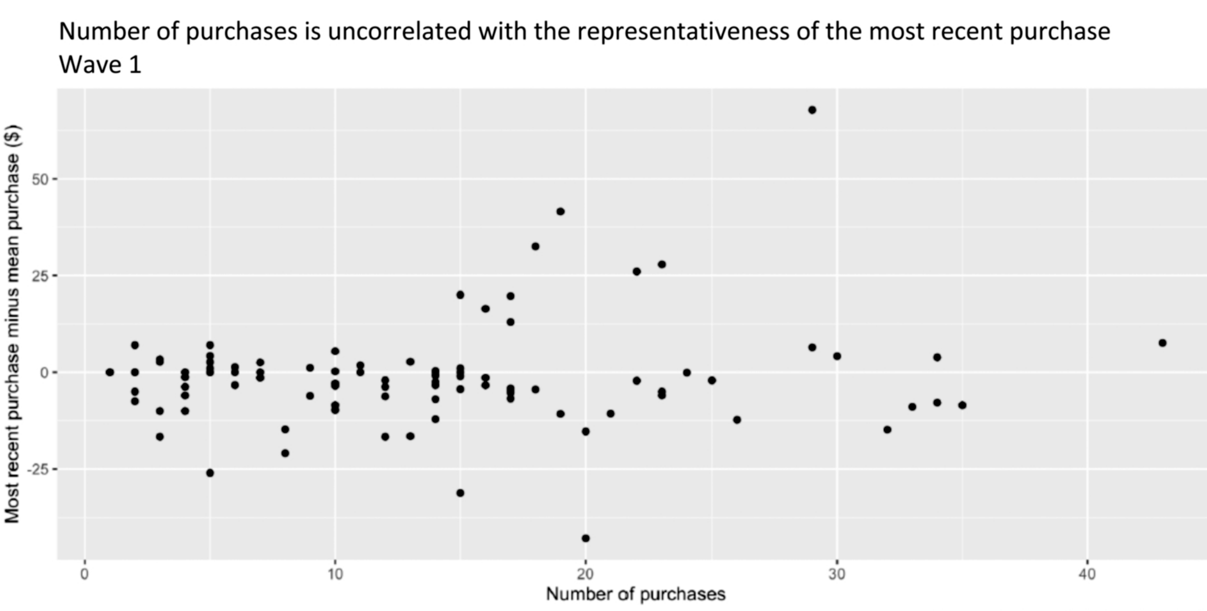

Fig. 8. A scatter plot showing, for each wave 1 participant, the discrepancy between most recent and mean purchase ($R_i - \overline{V}_i$) versus the number of purchases. The p-value under the hypothesis of zero correlation is 0.631, suggesting no evidence the number of purchases predicts the representativeness of a participant's most recent purchase value.

The actual total market size, *M*, was $35,954 in wave 1, while the estimator of total market size based on each participant's most recent purchase ($\widehat{M}_{Recent}$), modal purchase ($\widehat{M}_{Modal}$), and median purchase ($\widehat{M}_{Median}$) was $35,960 (error = $6 or 0.02%), $32,162 (error = -$3792 or −10.55%), and $32,704 (error = -$3250 or −9.03%) respectively. Figs. 5–7 show the bootstrapped sampling distributions of the three estimators of total market size: $\widehat{M}_{Recent}$, $\widehat{M}_{Modal}$, and $\widehat{M}_{Median}$, respectively (see technical appendix for details of bootstrapping procedure). As seen in Figs. 5–7, estimates of total market size based on the modal and median purchase appear to suffer from a significant downward bias, but estimates based on the most recent purchase appear to be unbiased.

Finally, Fig. 8 shows a scatter plot of the discrepancy between most recent and mean purchase ($R_i - \overline{V}_i$) versus the number of purchases for each wave 1 participant. The p-value under the hypothesis of zero correlation is 0.631, suggesting no evidence the number of purchases predicts the representativeness of a participant's most recent purchase value.

Results using our wave 2 data corroborate our main findings in Sections 3.2 and 3.3, and can be found in the supplementary materials.

4. Discussion

This study examined the representativeness of the most recent, modal and median purchase amounts (in US dollars) made by people who were using heroin regularly during two distinct two-week waves of data collection, separated by six months. In both waves, we found no evidence that the expected value of the most recent purchase differs from the expected value of the average purchase – that is, the cash value of the most recent purchase appears representative of purchasing history as summarized by the cash value of the average purchase. In contrast, in both waves the expected values of the modal and median purchases are smaller than the expected value of the average purchase. Taken together, these results imply that estimates of total market size based on the modal or median purchase will suffer from a significant downward bias, but that estimates based on the most recent purchase will be unbiased.

A decision we seek to inform is the design of surveys, specifically whether surveys that cannot ask about all purchases (as in our study) and instead must rely on a single question proxy, should ask about the "most recent" or about the "typical" purchase. Although it is an empirical question how survey respondents interpret the word "typical," if they interpret it to mean modal (most common) or median (middle point), then our findings suggest there could be downward bias in estimates of total market size; whereas if respondents interpret the word "typical" to mean "average", that would be ideal. That said, even the most recent purchase is not without potential problems as an estimator of total market size. For example, if the single question proxy is always asked on a particular day or time of day, that could introduce bias when asking about the most recent purchase that would not plague the "typical" purchase.[13] Moreover, inasmuch as the total error associated with a given estimator is a function of both its bias and variance, and our results focus exclusively on bias, future work should examine the conditions under which the most recent purchase is a better predictor of overall market size (i.e., results in lower total error) than other estimators.

Strengths of the study include two large samples of sequential heroin purchases made by people who were using heroin regularly (who represent the biggest share of the heroin market; Midgette et al., 2019). Compliance with study protocols was very high and all responses were rigorously fact-checked with potential outliers verified and/or reconciled with respondents within 24 h.

The study also has three main limitations. First, all data were self-reported via an IVR system. However, IVR systems typically yield high quality data, including drug use data, compared with interviews and questionnaires (Tucker et al., 2012; Abu-Hasaballah et al., 2007). For example, IVR systems have been used to collect sensitive data from individuals experiencing homelessness and addiction to crack cocaine (Freedman et al., 2006), adolescents with alcohol and other substance use disorders (Kaminer et al., 2006), and people living with poverty, problem drinking, and HIV/AIDS (Barta et al., 2008). Moreover, the IVR data used in this study were part of a larger study that found an almost perfect match between the price-elasticity of demand for heroin estimated using IVR data versus data collected from the same subjects via a laboratory experiment, thereby providing strong evidence that our IVR data are valid. Second, our findings are based on a sample of people who were using heroin regularly in Northeast, US and, as such, are not generalizable to people who use other types of illicit drugs or to other parts of the world. For example, individuals in our sample may be older than the population of people who use drugs in the US or elsewhere. Related, if people who use heroin have more stable purchase habits than do people who use some other drugs, then it may be harder to estimate total market size for those other drugs from a single question about the most recent, typical, or any other specific purchase.[14] Third, our findings are limited to estimating market size (a population-level measure) and do not necessarily extend to individual-level analyses. For example, bivariate distributions of (1) R_i, (2) C_i, and (3) Med_i all vs. $\overline{V}_i$, and their associated MAD and MAPE metrics, suggest that median purchase would perform better than most recent purchase for individual-level analyses (see Figures S.9-S.14 and Table S.1 in supplementary materials). As a practical matter, however, the accuracy with which survey respondents would be able to assess their median purchase is unclear.

In conclusion, our study provides evidence in support of using the most recent (but not the modal or median) purchase to estimate market size for heroin. However, until other studies have corroborated our findings, we recommend surveys continue to ask about other recent purchases as well.

Funding

The National Institute on Drug Abuse funded the study (R21DA025859). The funding agency played no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

CRediT authorship contribution statement

Olmstead, Pacula and Scott designed the study and wrote the first draft of the manuscript. Alessi oversaw collection of the data. Scott conducted the statistical analyses. All authors contributed to and have approved the final manuscript.

Conflict of Interest

The authors declare that they do not have a conflict of interest.

Acknowledgements

The authors wish to thank Rachel Atwater, Jacquelyn Booker, Ellen

[13] For example, Krebs et al. (2016) report an association between increased illicit drug use and monthly disbursement of social assistance payments (i.e., a "check effect"), while Goedel et al. (2019) report increased rates of overdose fatalities at the beginning of a month relative to the end of the preceding month.

[14] In waves 1 and 2, the mean±SD of the individual-level coefficients of variation in purchase sizes was 0.41 ± 0.24 and 0.38 ± 0.26, respectively (see Figures S.15-S.16 in supplementary materials for histograms of the individual-level coefficients of variation for both waves).

Ciesielski, Sean Sierra, and Venkataramana Yanamandra for outstanding assistance with this research. We also wish to thank Brian Pacula and two anonymous referees for helpful comments.

Technical Appendix

Bootstrapped tests and standard errors

Our bootstrapped test for the representativeness of the most recent, modal and median purchase was conducted as follows. Let $V_i = V_1, \ldots, V_{N_i}$ be the set of purchases for participant i in a given wave. To form an overall bootstrap sample, we bootstrapped each participant's purchases independently. That is, we resampled each V_i with replacement separately, thereby ensuring that each participant has the same number of purchases as in the original sample. We then randomly assigned one of these bootstrapped purchases to be the notional "most recent" purchase for participant i; this reflected the null hypothesis that $E(\mathrm{R_i} - \overline{V}_i) = 0$, i.e. that the most recent purchase is as good as a randomly sampled purchase for the purpose of estimating the mean purchase $\overline{V}_i$. For each bootstrapped sample we also calculated:

1. The mode of each participant's purchases.
2. The median of each participant's purchases.
3. The difference between the mean purchase and the modal purchase.
4. The difference between the mean purchase and the median purchase.
5. The difference between the notionally most recent purchase and the mean purchase.

We then repeated this sampling process 2500 times, building up a bootstrapped sampling distribution of the error made when using a participant's most recent, modal, and median purchase to estimate their average purchase. These sampling distributions were visually confirmed as being very nearly Gaussian in shape. Therefore the bootstrapped sampling distributions were used to calculate standard errors, which were subsequently used in Wald tests for whether $E(\mathrm{R_i} - \overline{V}_i) = 0$, $E(\mathrm{C_i} - \overline{V}_i) = 0$, and $E(\mathrm{Med_i} - \overline{V}_i) = 0$.

For each wave of data, the same bootstrapping procedure was used to form a sampling distribution for the estimate of total market size.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.drugalcdep.2022.109667.

References

Abu-Hasaballah, K., James, A., Aseltine, R., 2007. Lessons and pitfalls of interactive voice response in medical research. Contemp. Clin. Trials 28, 593–602.

Armstrong, M.J., 2021. Legal cannabis market shares during Canada's first year of recreational legalisation. Int. J. Drug Policy 88, 103028.

Barta, W., Portnoy, D., Kiene, S., Tennen, H., Abu-Hasaballah, K., Ferrer, R., 2008. A daily process investigation of alcohol-involved sexual risk behavior among economically disadvantaged problem drinkers living with HIV/AIDS. AIDS Behav. 12, 729–740.

Berwick, D.M., Murphy, J.M., Goldman, P.A., Ware, J.E., Barksy, A.J., Weinstein, M.C., 1991. Performance of a five-item mental health screening test. Med. Care 29, 169–176.

Bond, B., Caulkins, J.P., Scott, N., Kilmer, B., Dietze, P., 2014. Are users' most recent drug purchases representative? Drug Alcohol Depend. 142, 133–138.

Caulkins, J.P., Kilmer, B., Reuter, P.H., Midgette, G., 2015. Cocaine's fall and marijuana's rise: questions and insights based on new estimates of consumption and expenditures in US drug markets. Addiction 110 (5), 728–736.

Caulkins, J.P., Davenport, S., Doanvo, A., Furlong, K., Siddique, A., Turner, M., Kilmer, B., 2019. Triangulating web & general population surveys: do results match legal cannabis market sales? Int. J. Drug Policy 73, 293–300.

Delaney-Black, V., Chiodo, L.M., Hannigan, J.H., Greenwald, M.K., Janisse, J., Patterson, G., Huestis, M.A., Ager, J., Sokol, R.J., 2010. Just say "I Don't": lack of concordance between teen report and biological measures of drug use. Pediatrics 126 (5), 887–893.

EMCDDA, 2019 . European Monitoring Centre for Drugs and Drug Addiction and Europol, EU Drug Markets Report 2019. Publications Office of the European Union, Luxembourg. Also available online at: ⟨https://www.emcdda.europa.eu/publications/joint-publications/eu-drug-markets-report-2019_en⟩, last accessed December 6, 2021.

Freedman, M., Lester, K., McNamara, C., Milby, J., Schumacher, J., 2006. Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. J. Subst. Abus. Treat. 30, 105–111.

Goedel, W.C., Green, T.C., Viner-Brown, S., Rich, J.D., Marshall, B.D.L., 2019. Increased overdose mortality during the first week of the month: Revisiting the "check effect" through a spatial lens. Drug Alcohol Depend. 197, 49–55.

Golub, A., Johnson, B.D., 2004. How much do Manhattan-arrestees spend on drugs? Drug Alcohol Depend. 76 (3), 235–246.

Harrison, L.D., Martin, S.S., Enev, T., Harrington, D., 2007. Comparing drug testing and self-report of drug use among youths and young adults in the general population (DHHS Publication No. SMA 07-4249, Methodology Series M-7). Substance Abuse and Mental Health Services Administration, Office of Applied Studies, Rockville, MD.

Kaminer, Y., Litt, M., Burke, R., Burleson, J., 2006. An interactive voice response (IVR) system for adolescents with alcohol use disorders: a pilot study. Am. J. Addict. 15, 122–125.

Kilmer, B., Pacula, R.L., 2009. Estimating the size of the global drug market: a demand-side approach. RAND Technical Report TR-711-EC. RAND Europe, Cambridge, UK.

Kilmer, B., Caulkins, J.P., Pacula, R.L., Reuter, P.H., 2011. Bringing perspective to illicit markets: estimating the size of the US marijuana market. Drug Alcohol Depend. 119, 153–160.

Kilmer B., Caulkins J.P., Midgette G., Dahlkemper L., MacCoun R.J., and R.L. Pacula. (2013). Before the Grand Opening: Measuring Washington State's Marijuana Market in the Last Year Before Legalized Commercial Sales" RAND Report RR-466, December 2013.

Kilmer, B., Everingham, S.S., Caulkins, J.P., Midgette, G., Pacula, R.L., Reuter, P., Burns, R.M., Han, B., Lundberg, R., 2014. How Big Is the U.S. Market for Illegal Drugs? RAND Corporation, Santa Monica, CA. ⟨https://www.rand.org/pubs/research_briefs/RB9770.html⟩.

Kirby, K.N., Petry, N.M., Bickel, W.K., 1999. Heroin addicts have higher discount rates for delayed rewards than non-drug using controls. J. Exp. Psychol.: Gen. 128, 78–87.

Krebs, E., Wang, L., Olding, M., DeBeck, K., Hayashi, K., Milloy, M.J., Wood, E., Nosyk, B., Richardson, L., 2016. Increased drug use and the timing of social assistance receipt among people who use illicit drugs. Soc. Sci. Med. 171, 94–102.

van Laar, M., Frijns, T., Trautmann, F., Lombi, L., 2013. Sizing the cannabis market: a demand-side and user-specific approach in seven European countries. Curr. Drug Abus. Rev. 6 (2), 152–164.

Midgette, G., Davenport, S., Caulkins, J.P., Kilmer, B., 2019. What America's Users Spend on Illegal Drugs, 2006-2016. RAND Corporation, Santa Monica, CA.

National Institutes of Health. Available at: ⟨https://grants.nih.gov/policy/humans subjects/coc.htm⟩. Last referenced October 14, 2021.

Office of National Drug Control Policy, 2014. What American Users Spend on Illicit Drugs, 2000-2010. Executive Office of the President,, Washington DC.

Olmstead, T.A., Alessi, S.M., Kline, B., Pacula, R.L., Petry, N.M., 2015. The price elasticity of demand for heroin: Matched longitudinal and experimental evidence. J. Health Econ. 41, 59–71.

Reuter, P., Caulkins, J.P., Midgette, G., 2021. Heroin use cannot be measured adequately with a general population survey. Addiction.

Tucker, J., Roth, D., Huang, J., Crawford, M., Simpson, C., 2012. Effects of interactive voice response self-monitoring on natural resolution of drinking problems: utilization and behavioral economic factors. J. Stud. Alcohol Drugs 73, 686–698.

INCSR, 2020a. U.S. Department of State 2020 – International Narcotics Control Strategy Report (INCSR) Volume I: Drug and Chemical Control. U.S. Department of State, Bureau of International Narcotics Matters. Bureau of International Narcotics and

Law Enforcement Affairs, Washington, D.C. Available at: ⟨https://www.state.gov/2020-incsr-volume-i-drug-and-chemical-control-as⟩- submitted for publication-to-congress/ Last accessed December 6, 2021.

INCSR, 2020b . U.S. Department of State, 2020 – International Narcotics Control Strategy Report (INCSR) Volume II: Money Laundering and Financial Crimes report. U.S. Department of State, Bureau of International Narcotics Matters. Bureau of International Narcotics and Law Enforcement Affairs, Washington, D.C. Available at: ⟨https://www.state.gov/2020-incsr-volume-ii-money-laundering⟩- and-financial-crimes-as-submitted for publication-to-congress/ Last accessed December 6, 2021.

United Nations Office on Drugs and Crime. World Drug Report 2021. United Nations publication, Sales No. E.21.XI.8. Also available at ⟨https://www.unodc.org/unodc/en/data-and-analysis/wdr2021.html⟩.

US Drug Enforcement Administration. National Drug Threat Assessment, Springfield, Va., DEA-DCT-DIR-032–18, October 2018.

Ware Jr, J.E., Kosinski, M., Bjorner, J.B., Turner-Bowker, D.M., Maruish, M., 2007. SF-36 Health Survey. Manual and Interpretation Guide. 2. QualityMetric Incorporated, Lincoln, RI.

Wilkins, C., Bhatta, K., Casswell, S., 2002. A ‘demand side’ estimate of the dollar value of the cannabis black market in New Zealand. Drug Alcohol Rev. 21 (2), 145–151.