Published in final edited form as:

Prison J. 2013 December 1; 93(4): 375–389. doi:10.1177/0032885513500616.

Consistency in Self-Reports of Drug Use Frequency by High-Risk Offenders Over a 5-Year Interval

Michael Prendergast¹, Libo Li¹, Elizabeth Evans¹, and Elizabeth Hall¹

¹University of California, Los Angeles, USA

Abstract

Self-reports of drug use by high-risk offenders interviewed on two occasions were compared to determine the extent to which the reports were consistent at the two interview points. Self-reports of frequency of drug use over the same 12-month period were compared among parolees ($N = 380$) who had participated in prison drug treatment and who were interviewed at 1 and 5 years following prison release. The kappa coefficient was .31 ($p < .001$), generally considered a fair level of agreement. Total concordance in retrospective recall of primary drug use frequency was 54.5%.

Keywords

illicit drug use; retrospective self-report; longitudinal patterns; reliability

Introduction

Studies of longitudinal patterns of illicit drug use are usually dependent for their data on recall by respondents of events over long time periods. Depending on the design of the study, subjects may be asked to indicate which drugs they were using and how often stretching back from 12 months to 10 years or more. How likely is it that respondents can provide accurate information that requires recalling distant events or patterns of behavior? The strength of findings from longitudinal studies depends on the confidence that can be placed in the reliability of data that is based on long-term memory. Factors that may affect the reliability and accuracy of retrospective data include the length of the recall period, the degree to which respondents can accurately recall distant events, whether they have an incentive to provide accurate information, whether the behavior is undesirable or stigmatized, and the conditions under which the information is elicited (Fendrich & Vaughn, 1994; Johnson & Mott, 2001; Mensch & Kandel, 1988; O'Malley, Bachman, & Johnston, 1983).

© 2013 SAGE Publications

Corresponding Author: Michael Prendergast, Integrated Substance Abuse Programs, University of California, Los Angeles, 11075 Santa Monica Blvd. Suite 100, Los Angeles, CA 90025, USA., mlp@ucla.edu.

The content of this publication does not necessarily reflect the views or policies of the National Institute on Drug Abuse (NIDA).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Reprints and permissions: sagepub.com/journalsPermissions.nav

Although there are numerous studies of the reliability of retrospective reports of tobacco and alcohol behavior (Brigham et al., 2008; Koenig, Jacob, & Haber, 2009; Kenkel, Lillard, & Mathios, 2003; Sobell et al., 1988), findings from such studies may not generalize to long-term recall of illicit drug use, given the different cognitive effects of illicit drugs, the social marginalization of illicit drug users, and the social stigma associated with illicit drug use. But a number of studies have examined the accuracy and consistency of information on illicit drug use based on recall of events over relatively long time periods (Bailey, Flewelling, & Rachal, 1992; Fendrich & Rosenbaum, 2003; Hser, Anglin, & Chou, 1992; Johnson & Mott, 2001; Shillington, Cottler, Mager, & Compton, 1995). Overall, this literature indicates that although underreporting is common, self-report of substance use is sufficiently reliable for answering most research questions that rely on such data. In addition, research has shown that reliability can be improved by using time anchors to structure recall (Caspi et al., 1996; Morris & Slocum, 2010; Sobell & Sobell, 1996).

Of particular concern is the reliability of long-term recall, 5 years or more. Using data from the St. Louis Epidemiologic Catchment Study, Shillington et al. (1995) examined the “self-report stability” of substance use in a non-clinical, general population sample between baseline and a 10-year follow-up. Findings indicated that, depending on the specific drug, between 73% and 85% of the respondents were consistent in their reports of substance use over a 10-year period. More serious discrepancies in self-report of substance use was found by Fendrich and Vaughn (1994), who examined underreporting of lifetime marijuana and cocaine use across two administrations of the National Longitudinal Survey in 1984 and 1988. The rate of reduced use at the 1988 interview was nearly one third of marijuana users in 1984 and just over one third of cocaine users in 1984. Underreporting may be more common for “harder” drugs, as about 19% of cocaine users completely denied use at follow-up, compared with 12% of marijuana users. Underreporting was most common among those who were interviewed by telephone, among minority respondents, and among those with lower levels of education.

Subjects in the above studies were sampled from the general population. Reliability of self-report over time might be different for a criminal population. In a series of articles, Hser et al. (1992; Anglin, Hser, & Chou, 1993; Chou, Hser, & Anglin, 1996) examined the reliability of self-report data provided by a sample of primary heroin addicts who were interviewed at two time points 10 years apart, with an overlapping period of 4 to 5 years. With reference to results for drug use, the researchers (Hser et al., 1992) found that the test-retest correlation between reports of drug use for the overlapping period was high (above .60) for frequent events such as abstinence, daily heroin use, and number of injections per month, but relatively low (less than .50) for less frequent events such as heroin use less than once per week, marijuana use, nonnarcotics use, heavy alcohol use, and drug dealing. Reported levels of drug use behavior were lower at the first interview than at the second. Baseline characteristics that were significantly associated with less consistent responses were having used alcohol heavily, having had a long history of heroin use, having had a long history of property crime, and having spent a longer time incarcerated, all suggestive of a more deviant lifestyle.

The purpose of the present analysis was to compare reports of illicit drug use by participants in a prison-based treatment program evaluation for the same 12-month period provided at two points in time: 1 year and 5 years after release from prison. Two questions are addressed. First, to what extent were reports of drug use consistent at the two interview points? Second, were selected respondent characteristics associated with different levels of consistent reporting?

Method

Data Source

The data used in this analysis were collected from a 1-year follow-up and a 5-year follow-up evaluation of the Amity prison-based therapeutic community at the R. J. Donovan Correctional Facility, a medium-security prison for men near San Diego (Prendergast, Hall, Wexler, Melnick, & Cao, 2004; Wexler, De Leon, Thomas, Kressel, & Peters, 1999; Wexler, Melnick, Lowe, & Peters, 1999). Subjects in the Amity study were randomized to the prison treatment program ($n = 425$) or to no treatment ($n = 290$). For the 1-year follow-up, 74% of the original sample was interviewed, and for the 5-year follow-up, 80% of the original sample was interviewed. The present analysis includes 380 subjects with drug use data at both interview points (see “Full sample” column in Table 2 for baseline characteristics of respondents). The focus of the analysis is on the frequency of use over the same 12-month period of the drug reported by respondents as their primary drug used at baseline, namely alcohol, marijuana, cocaine, opiates, and amphetamine. The format of the 1-year and the 5-year questionnaires differed, which required that the responses be recoded to make the frequency of use categories comparable across interview points, as described below.

Data From the 1-Year Follow-Up

At the 1-year follow-up interview, respondents reported the frequency of use of each drug (if used) for three time periods over the 12 months after their release from prison, namely, the first 3 months, the middle 6 months, and the most recent 3 months. Each of these time periods had 11 frequency of use levels, ranging from never/not used to used several times a day. For each type of drug in each time frame, we recoded the frequency of use reported by subjects into three levels: no use (if the respondent reported no use for a given drug), some use (if the respondent reported use at some level less than daily), and daily use (if the respondent reported use every day or more than once every day). Then, within each time period, the frequency of primary drug use (i.e., the drug that the respondent said was his main or primary drug at baseline) was determined by the recoded frequencies of use on all drugs belonging to the primary drug category (e.g., opiates as a primary drug consisted of heroin snorted, heroin injected, heroin and cocaine together, nonprescribed methadone, and other opiates). Specifically, if the respondent reported no use on all drugs in a primary drug category within a specific time period, the subject’s frequency of primary drug use was coded as “no use.” If the subject had a higher level of frequency of use on any of the drugs in a primary drug category, that level was used to represent the subject’s frequency of primary drug use in that time period (i.e., either “some use” or “daily use”).

Once the frequency of primary drug use in each time period was determined, the frequency of primary drug use for the entire year was determined in the same way. That is, if the subject had no use of the primary drug in all three time periods, the subject's frequency of primary drug use was coded as no use. If the subject had a higher frequency of primary drug use in any time period, that level was used to represent the subject's frequency of primary drug use in the 1st year after release (i.e., either some use or daily use). Recall of no use or daily use, even extending back many years, can be expected to be relatively accurate (assuming a desire on the part of the respondent to be accurate). As other levels of use frequency are likely to be more problematic, we chose to focus on the range of use levels under the "some use" category rather than recode to more specific time periods (e.g., monthly, weekly).

Data From the 5-Year Follow-Up

At the 5-year interview, reports on drug use were collected using the Natural History Interview (NHI). The NHI was originally developed in the mid-1970s for an evaluation of the California Civil Addict Program (McGlothlin, Anglin, & Wilson, 1977); it has since been used for many evaluations of drug treatment interventions in community and correctional settings (Brecht, Huang, Evans, & Hser, 2008; Grella & Lovinger, 2011; Prendergast et al., 2004). The NHI uses a dynamic time line follow-back method that allows collection of a continuous stream of data on behaviors of interest over extended time periods. A basic time line is constructed from available information, including official criminal justice records for offender samples. At the interview, the interviewer helps the client add to this time line by including the dates of major events to facilitate recall (e.g., marriages, childbirths, relocations) that serve as time anchors for recall of drug use activity. The interviewer then identifies time segments defined by changes in drug use patterns or by relevant events (e.g., shift from daily use to monthly use, treatment admission, incarceration) over the time period of interest. The interview then involves the repeated administration of a dynamic form for each type of drug for as many segments as necessary. These data are then converted to time series-type data for each drug. Depending on the analysis, the unit over which behavior is measured can be months, quarters, or years.

Data collected using the NHI in previous studies have been shown to be generally reliable. In an examination of retrospective self-reports using the NHI format (Anglin et al., 1993; Chou et al., 1996), test-retest correlations over the same 4- to 5-year overlapping period for interviews conducted 10 years apart with the same subjects (primary narcotics addicts; $n = 258$) were calculated for 46 selected variables. Correlations for drug use variables include abstinence from narcotics use, .71; daily narcotics use, .63; and number of injections per month, .65.

For the Amity follow-up study (Prendergast et al., 2004; Wexler, De Leon, et al., 1999; Wexler, Melnick, et al., 1999), the NHI was used to reconstruct a history of each respondent's drug use over the period from 5 years prior to entrance to the prison term associated with their participation in the Amity study up to the time of the 5-year interview. This allowed us to estimate the frequency of primary drug use over the same 12-month period as was used in the 1-year follow-up, although the structure of the NHI questionnaire

required different procedures for calculating frequency of use than was used for the 1-year follow-up data. For each month that the respondent was “on the street” (i.e., not in residential treatment or incarcerated), the frequency of primary drug use ranges from 0 to 30 days. For the present analysis, this monthly measure of frequency of use of the primary drug was then averaged across the months of the 1st year after release from prison. This yearly average frequency was coded into three use level categories (no use, some use, and daily use) to create a measure indicating the subject’s frequency of primary drug use in the 12 months following release from prison, as recalled 5 years later. Specifically, if the average yearly frequency was equal to 0, then the subject’s frequency of primary drug use was coded as “no use.” If the average frequency was greater than 0 but less than or equal to 28 days per month, it was coded as “some use.” If the average frequency was greater than 28 days per month, it was coded as “daily use.” Thus, use level categories were the same at 1 year and 5 years.

Statistical Analysis

The analysis was based on data from 380 subjects who participated in both interviews and who had sufficient information to create the two measures of the frequency of primary drug use over the 12 months postrelease. The reports of drug use frequency at the two interview points were compared to determine the consistency of reports for the same 12-month time period. Subjects were classified into three groups based on comparison of measures at the two time points: total concordance, partial concordance, and no concordance. That is, if the two measures of frequency were the same at each time point, the status of the subject was classified as “total concordance.” If the two measures differed by one level on frequency (e.g., no use by the first measure but some use by the second measure, or some use by the first measure but daily use by the second measure), the subject’s status was classified as “partial concordance.” If the two measures differed by two levels on frequency (e.g., no use by the first measure but daily use by the second measure), the status of the subject was classified as “no concordance” (see Table 1 for all combinations).

In addition to comparing the degree of concordance between the two interviews points, we also determined whether selected respondent characteristics were associated with the concordance categories. Baseline characteristics were age at study intake, race/ethnicity, number of years in prison in lifetime, general proficiency score (grade-level attainment), IQ score (Revised Beta II), mental health disorder based on the Diagnostic Interview Schedule (antisocial personality, anxiety, and major depression, coded as yes/no), and primary drug type. We created one additional variable from data reported at the 1-year follow-up: age at first use of primary drug. We selected the youngest age at first use among the drugs belonging to the subject’s primary drug category.

To evaluate the strength of agreement between measures of primary drug use frequency between the two interview points, SAS Proc Freq was used to calculate the kappa and the weighted kappa coefficients for the concordance categories defined above. The following criteria (Altman, 1991) were used to evaluate the calculated kappa or the weighted kappa coefficients: for a kappa less than .20, agreement is poor; between .21 and .40, fair; between .41 and .60, moderate; between .61 and .80, good; and between .81 and 1.00, very

good. Analysis of the three concordance categories by participant characteristics used the chi-square test for categorical covariates and Proc General Linear Model (GLM) for continuous variables.

Results

Degree of Concordance in Retrospective Recall of Primary Drug Use Frequency

As shown in Table 1, the kappa coefficient for the degree of concordance between the 1-year and the 5-year follow-up interviews in retrospective recall of primary drug use frequency indicated a fair level of agreement. The kappa coefficient was .31 ($p < .001$) and the weighted kappa coefficient was .35 ($p < .0001$).

The main category of agreement was total concordance, followed by partial concordance and then no concordance. Total concordance in retrospective recall of primary drug use frequency was demonstrated by more than half of respondents (54.5%). No use in the overlapping 12-month period was reported at both interviews by 17.6% of respondents, another 10.5% reported some use at both time points, and 26.3% consistently reported daily use.

The reports of drug use frequency for about one third of respondents (31.6%) were categorized as partial concordance. Some inconsistency was due to reports of no use at the 1-year follow-up followed by reports of some use at the 5-year follow-up (2.6%); by contrast, 9.7% of respondents reported some use at the 1-year follow-up but no use at the 5-year follow-up. Slightly more discordance was attributable to inconsistencies in reports of some use versus daily use across interviews (12.4% and 6.8%).

For 14.0% of respondents, there was no concordance in recall of primary drug use frequency at the two interviews. Among this group, 5.8% reported no use at the 1-year follow-up but daily use at the 5-year follow-up, and another 8.2% first reported daily use and then reported no use.

Concordance Categories by Respondent Characteristics

As shown in Table 2, the only respondent characteristic examined that was statistically different across concordance categories was primary drug type. For the total concordance category, methamphetamine was reported as the main primary drug (32.4%). In the no-concordance category, the main primary drug was opiates (32.1%). The leading primary drugs for those who fell into the partial concordance category were cocaine (25.0%) and methamphetamine (24.2%). No significant differences across concordance categories were found for age, race/ethnicity, number of years incarcerated, general proficiency score, IQ, mental health disorder (antisocial personality, anxiety, depression), or age at first use of primary drug.

Discussion

Retrospective self-report is a common method to collect data on respondent behavior. Questions about such data have to do with their reliability, particularly when the recall

period is long, and whether certain respondent characteristics are associated with different levels of reliability. The current analysis examined concordance at two time points for self-reports of frequency of primary drug use over an overlapping 12-month period time among offenders who had participated in a prison-based drug treatment evaluation.

More than half of the respondents (54.5%) reported the same frequency of use at both time points (i.e., no use, some use, or daily use). Partial concordance in use level was evident among nearly a third (31.6%) of respondents, with the largest discrepancy (12.4%) being observed among those who reported some use at the 1-year follow-up and daily use at the 5-year follow-up. Finally, 14.0% of the respondents demonstrated no concordance in frequency of use at the two interviews; 5.8% reported at the 1-year interview that they had not used their primary drug over the previous 12 months, but reported daily use at the 5-year interview, whereas 8.2% reported daily use and later reported no use. Kappa coefficient was .31 (weighted .35), indicating a fair level of agreement. In examining selected respondent characteristics that might be associated with the concordance categories, only primary drug type at baseline was found to be significant.

While previous studies have used other approaches to examine the reliability of retrospective recall, none has based its findings on a high-risk offender population. Thus, it is difficult to make a direct comparison of the findings from this study with other analyses of self-report reliability over time. Nonetheless, most of these studies have tended to find a higher level of reliability for reports of drug use behavior farther in the past, although there are differences by type of drug and by respondent characteristics (Johnson & Mott, 2001; Mensch & Kandel, 1998; Shillington et al., 1995). The results from this study suggest that adult offenders with long criminal and drug use histories are less likely than other populations to consistently report on frequency of use over a long period, which could result from a number of factors, including intentional misreporting, memory issues, and stigma. More research is needed to replicate or modify this finding.

In considering the results, of particular concern is the no-concordance category (14%): report of no use over the 12 months before the 1-year interview, but daily use for the same 1-month period at the 5-year interview, or reports of daily use, followed by reports of no use. Given the stark contrast between reporting no use and reporting daily use over the same 12-month period, it seems unlikely that the discrepancy is the result of forgetting. More likely, it results from deliberate misreporting, although without independent confirmation, it is not possible to know for sure which report is truthful. However, for the 5.8% of those who reported no use at 1 year but daily use at 5 years, it might be argued that the second report is the more accurate one as any motivations that might have led to denial of use at the first interview were less likely to have been present at the second interview. O'Malley et al. (1983) noted (at least for adolescents) that stigmatized behavior is more likely to be denied or underreported when it is reported closer in time to the interview. The salience of the negative associations of the behavior appears to fade as the recall period becomes greater.

The opposite behavior—reporting of daily use at Time 1, but denying any use at Time 2—is also of concern. Assuming that the report at Time 1 is truthful, analyses that relied on reports from Time 2 would result in use levels being underestimated. Research with

adolescents has found high levels of recanting of previous reports of drug use at a second interview (Fendrich & Rosenbaum, 2003), ranging from 45% for lifetime use of alcohol to 81% for lifetime use of cocaine. Given the high levels of recanting among adolescents, it is noteworthy that in the present sample of offenders recanting of previously reported use was found among only 8.2% of respondents. Additional research would be needed to confirm this difference between adolescents and offenders in the rate of recanting and to explore the reasons for the difference.

Collecting data prospectively avoids the problems with recall that are endemic to retrospective longitudinal studies. But prospective studies are often difficult to fund on an ongoing basis, and some research questions will not justify the expense of multiple waves of data collection. Moreover, questions that require answers from longitudinal data for immediate policy or clinical purposes cannot wait for the 10 or more years that a prospective study would require and instead would benefit from a well-designed retrospective study. Studies relying on longitudinal data collected retrospectively will continue to be conducted, and studies that are concerned with drug use over time will necessarily rely on self-reported information. The findings of such studies will be strengthened if researchers are able to incorporate techniques that are likely to improve accurate recall of distant events and behaviors, including the use of life history calendars that include personal and historical “anchor” events (Caspi et al., 1996; Loftus & Marburger, 1983; Morris & Slocum, 2010), the regular practice of attempting to resolve discrepancies between and within interviews (Cottler et al., 1994), procedures to reduce the tendency to provide socially desirable responses such as pairing respondents with different interviewers at each interview point (Mensch & Kandel, 1988), use of audio computer-assisted interview technology (Gribble et al., 2000), and assurances to respondents that the information provided will be kept confidential and used for research purposes only. Although perfectly accurate data are not possible, these and other methods should increase the likelihood of collecting truthful retrospective data that can be relied on to answer research questions of interest. At the same time, additional research is needed to validate techniques that would improve the accuracy and consistency of data collected in longitudinal studies.

Because longitudinal studies involve asking the same questions at multiple time points, researchers could compare responses on specific drug use items across interview points as part of preliminary analysis. Such analysis could also be periodically conducted during later interview administration to see whether the data for specific questions meet a predefined level of concordance, with attempts to introduce improved administration procedures to increase consistency. Although doing so might improve consistency across interview points (e.g., reducing the level of recanting), it cannot establish the accuracy or truthfulness of the responses without independent confirmation from objective sources, such as biological tests or information from collaterals.

The results of the present study should be considered in the context of several limitations involved in the analysis. First, the interviews were conducted with a high-risk criminal population, many of whom had antisocial personality disorder. The behavior asked about is illegal, deviant, and stigmatized—all of which are likely to lead to intentional misreporting, apart from any problems in recall. Second, we examined consistency in *self-report* of drug

use at two points in time. Although urine and hair samples were available from some respondents, they were too few for analysis and, in any case, would be of little use in confirming self-reports of use over the 12-month period observed. Third, this analysis was not a test–retest validation of a single instrument. Rather, it examined consistency in self-report on drug use frequency over the same time period using two instruments with different question formats to elicit estimates of frequency of drug use. Unlike previous studies (Hser et al., 1992), this study did not seek to determine the reliability of the Natural History Instrument. Fourth, secondary analysis of the Amity datasets is limited by the questions and coding of the original instruments, which were not intended to assess reliability issues. In particular, for the Amity evaluation, different methods were used to collect data on frequency of drug use at each interview; the 1-year interview asked about frequency of use in terms of categories (e.g., never, once a month, several times a month, daily), whereas the 5-year interview asked about the number of days used in a month. The use of broad frequency categories (no use, some use, daily use) should minimize inconsistencies in the two approaches to asking about use frequency. Finally, the analysis focused on frequency of primary drug use among high-risk male offenders. Findings may not apply to other aspects of drug use (e.g., age of first use, quantity of use), to other behaviors (e.g., crimes committed, income from crime, employment), or to other populations (e.g., low-risk offenders, women offenders, nonoffenders).

Conclusion

Many research questions can only be answered from data collected retrospectively, often over long time periods and based on self-reports that cannot be independently confirmed. Thus, the reliability of the recalled information is important in assessing the strengths and validity of findings based on such data. Findings from this study indicated that reports by high-risk offenders of frequency of drug use over 12 months recalled at two time points showed fair correspondence, with more than half of respondents reporting the same level of use at both interviews. Although this finding suggests that analysis of longitudinal data already collected, particularly from offenders, should be interpreted with caution, it is likely that future studies that rely on retrospective recall can improve reliability through the use of life history calendars and other approaches that have been shown to improve recall. Additional research is needed to determine the long-term reliability of self-reports of drug use behaviors among different populations and different time periods and using different recall techniques.

Acknowledgments

We wish to thank Sara Captain and Jason Grossman for editorial assistance.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported in part by the UCLA Center for Advancing Longitudinal Drug Abuse Research (CALDAR) under Grant P30DA016383 from the National Institute on Drug Abuse (NIDA).

References

- Altman, DG. Practical statistics for medical research. New York, NY: Chapman & Hall; 1991.
- Anglin MD, Hser YI, Chou CP. Reliability and validity of retrospective behavioral self-report by narcotics addicts. *Evaluation Review*. 1993; 17:91–108.
- Bailey SL, Flewelling RL, Rachal JV. The characterization of inconsistencies in self-reports of alcohol and marijuana use in a longitudinal study of adolescents. *Journal of Studies on Alcohol*. 1992; 53:636–647. [PubMed: 1434637]
- Brecht ML, Huang D, Evans E, Hser YI. Polydrug use and implications for longitudinal research: Ten-year trajectories for heroin, cocaine, and methamphetamine users. *Drug and Alcohol Dependence*. 2008; 96:193–201. [PubMed: 18329825]
- Brigham J, Lessov-Schlagger CN, Javitz HS, McElroy J, Krasnow R, Swan GE. Reliability of adult retrospective recall of lifetime tobacco use. *Nicotine & Tobacco Research*. 2008; 10:287–299. [PubMed: 18236293]
- Caspi A, Moffitt TE, Thorton A, Freedman D, Amell JW, Smeijers J, Silva PA. The life history calendar: A research and clinical assessment method for collecting retrospective event-history data. *International Journal of Methods in Psychiatric Research*. 1996; 6:101–114.
- Chou CP, Hser YI, Anglin MD. Pattern reliability of narcotics addicts' self-reported data: A confirmatory assessment of construct validity and consistency. *Substance Use & Misuse*. 1996; 31:1189–1216. [PubMed: 8853237]
- Cottler LB, Compton WM, Brown L, Shell A, Keating S, Shillington A, Hummel R. The discrepancy interview protocol: A method for evaluating and interpreting discordant survey responses. *International Journal of Methods in Psychiatric Research*. 1994; 4:1–10.
- Fendrich M, Rosenbaum DP. Recanting of substance use reports in a longitudinal prevention study. *Drug and Alcohol Dependence*. 2003; 70:241–253. [PubMed: 12757962]
- Fendrich M, Vaughn CM. Diminished lifetime substance use over time: An inquiry into differential underreporting. *Public Opinion Quarterly*. 1994; 58:96–123.
- Grella CE, Lovinger K. 30-year trajectories of heroin and other drug use among men and women sampled from methadone treatment in California. *Drug and Alcohol Dependence*. 2011; 118:251–258. [PubMed: 21549528]
- Gribble JN, Miller HG, Cooley PC, Catania JA, Pollack L, Turner CF. The impact of T-ACASI interviewing on reported drug use among men who have sex with men. *Substance Use & Misuse*. 2000; 35:869–890. [PubMed: 10847215]
- Hser YI, Anglin MD, Chou CP. Reliability of retrospective self-report by narcotics addicts. *Psychological Assessment*. 1992; 4:207–213.
- Johnson TP, Mott JA. The reliability of self-reported age of onset of tobacco, alcohol, and illicit drug use. *Addiction*. 2001; 96:1187–1198. [PubMed: 11487424]
- Kenkel D, Lillard DR, Mathios A. Smoke or fog? The usefulness of retrospectively reported information about smoking. *Addiction*. 2003; 98:1307–1313. [PubMed: 12930218]
- Koenig LB, Jacob T, Haber JR. Validity of the lifetime drinking history: A comparison of retrospective and prospective quantity-frequency measures. *Journal of Studies on Alcohol and Drugs*. 2009; 70:296–303. [PubMed: 19261242]
- Loftus EF, Marburger W. Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition*. 1983; 11:114–120. [PubMed: 6865744]
- McGlothlin, WH.; Anglin, MD.; Wilson, BD. An evaluation of the California civil addict program (NIDA Services Research Monograph Series). Rockville, MD: National Institute on Drug Abuse; 1977.
- Mensch BS, Kandel DB. Underreporting of substance use in a national longitudinal youth cohort. *Public Opinion Quarterly*. 1988; 52:100–124.
- Morris NA, Slocum LA. The validity of self-reported prevalence, frequency, and timing of arrest: An evaluation of data collection using a life event calendar. *Journal of Research in Crime & Delinquency*. 2010; 47:210–240.

- O'Malley PM, Bachman JG, Johnston LD. Reliability and consistency in self-reports of drug use. *International Journal of the Addictions*. 1983; 18:805–824. [PubMed: 6605313]
- Prendergast M, Hall E, Wexler H, Melnick G, Cao Y. Amity prison-based therapeutic community: Five-year outcomes. *The Prison Journal*. 2004; 84:36–60.
- Shillington AM, Cottler LB, Mager DE, Compton WM. Self-report stability for substance use over 10 years: Data from the St. Louis Epidemiological Catchment study. *Drug and Alcohol Dependence*. 1995; 40:103–109. [PubMed: 8745131]
- Sobell, LC.; Sobell, MB. Timeline followback: User's manual. Toronto, Ontario, Canada: Addiction Research Foundation; 1996.
- Sobell LC, Sobell MB, Riley DM, Schuller R, Pavan DS, Cancilla A, Leo GI. The reliability of alcohol abusers' self-reports of drinking and life events that occurred in the distant past. *Journal Studies on Alcohol*. 1988; 49:225–232.
- Wexler HK, De Leon G, Thomas G, Kressel D, Peters J. The Amity prison evaluation: Reincarceration outcomes. *Criminal Justice and Behavior*. 1999; 26:147–167.
- Wexler HK, Melnick G, Lowe L, Peters J. Three-year reincarceration outcomes for Amity in-prison and aftercare therapeutic community and aftercare in California. *The Prison Journal*. 1999; 79:321–336.

Biographies

Michael Prendergast is the director of the Criminal Justice Research Group, University of California, Los Angeles (UCLA), Integrated Substance Abuse Program (ISAP). A principal investigator of state and federally funded projects on criminal justice system drug treatment strategies, his research interests include treatment interventions for drug-abusing offenders, systematic reviews of treatment effectiveness, and strategies for implementing evidence-based practices.

Libo Li has worked at the UCLA ISAP as a social and behavioral research data analysis expert. His research interests include methodological development and evaluation of advanced models for longitudinal research and new methods' application to real data analysis. He has been actively involved in many ISAP longitudinal drug abuse studies.

Elizabeth Evans is a project director at UCLA ISAP. Her area of interest is a public health application approach to the understanding of multilevel determinants, correlates, and consequences of substance use disorders over the life course. She has published on health service utilization and outcomes, health disparities, and drug offender program treatment effectiveness.

Elizabeth Hall is a researcher in UCLA's ISAP's Criminal Justice Research Group. Over more than 15 years of substance abuse research, she has been involved in a variety of substance abuse treatment outcomes research and evaluation projects. Her areas of expertise include substance abuse treatment setting contingency management and criminal justice system substance abuse treatment for women and child welfare system families.

Concordance of Retrospective Recall: Primary Drug use Frequency Reported at 1- and 5-Year Follow-Up Interviews for the Same 12-Month Period ($N = 380$).

Table 1

		Total concordance ($n = 207$; 54.5%)		Partial concordance ($n = 120$; 31.6%)		No concordance ($n = 53$; 14.0%)	
Frequency of use reported at							
1-year follow-up	No use	Daily use		Daily use		Daily use	
5-year follow-up	No use	Daily use		Daily use		Daily use	
n	67	40		100		10	
%	17.6	10.5		26.3		2.6	
		40		100		37	
		10.5		26.3		9.7	
		10.5		26.3		12.4	
		40		100		47	
		10.5		26.3		26	
		10.5		26.3		6.8	
		40		100		22	
		10.5		26.3		5.8	
		10.5		26.3		8.2	
		40		100		31	
		10.5		26.3		No use	
		10.5		26.3		No use	

Table 2

Characteristics of Respondents at Baseline by Concordance Level ($N = 380$).

	Total concordance ($n = 207$)		Partial concordance ($n = 120$)		No concordance ($n = 53$)		Full sample (380)	
	%	or M (SD)	%	or M (SD)	%	or M (SD)	%	or M (SD)
Age	30.4	(6.5)	29.9	(7.2)	29.0	(6.7)	30.0	(6.8)
Race/ethnicity								
African American	33.3		40.0		34.0		35.5	
White	43.0		27.5		34.0		36.8	
Hispanic	19.3		27.5		30.2		23.4	
Other	4.4		5.0		1.9		4.2	
No. of years in prison in lifetime	4.3	(3.4)	4.4	(3.4)	4.1	(3.1)	4.3	(3.4)
Cognitive ability								
General proficiency score	7.6	(3.3)	7.2	(3.3)	7.0	(3.0)	7.4	(3.2)
IQ score (scaled)	86.3	(11.1)	85.9	(12.5)	86.6	(10.1)	86.2	(11.4)
Mental health disorders								
Antisocial personality	56.9		53.4		56.9		55.8	
Anxiety	21.6		24.1		25.5		22.9	
Major depression	8.8		9.5		15.7		10	
Primary drug type*								
Alcohol	8.7		15.8		3.8		10.3	
Marijuana	11.6		17.5		15.1		14.0	
Cocaine	20.3		25.0		26.4		22.6	
Opiates	27.1		17.5		32.1		24.7	
Methamphetamine	32.4		24.2		22.6		28.4	
Age at first use of primary drug	16.4	(4.6)	16.3	(5.7)	16.6	(4.6)	16.4	(4.9)

* $p < .05$.