Published in final edited form as:

Addict Behav. 2013 October ; 38(10): 2568–2574. doi:10.1016/j.addbeh.2013.05.015.

Concordance Between Self-Report and Urine Drug Screen Data in Adolescent Opioid Dependent Clinical Trial Participants

Claire E Wilcox, MD¹, Michael P Bogenschutz, MD^{1,2}, Masato Nakazawa, PhD², and George Woody, MD³

Claire E Wilcox: cewilcox@salud.unm.edu; Michael P Bogenschutz: mbogenschutz@salud.unm.edu; Masato Nakazawa: masatobach@gmail.com

¹University of New Mexico Department of Psychiatry, 1 University of New Mexico, MSC 09-5030, Albuquerque, NM, 87131 USA

²Center on Alcoholism, Substance Abuse, and Addictions. University of New Mexico CASAA. 2650 Yale SE, MSC 11-6280, Albuquerque, NM 87106

³Perelman School of Medicine at the University of Pennsylvania and Treatment Research Institute. 500 Public Ledger Bldg; 150 South Independence Mall (W); Philadelphia, PA 19106

Abstract

Objective measures of drug use are very important in treatment outcome studies of persons with substance use disorders, but obtaining and interpreting them can be challenging and not always practical. Thus, it is important to determine if, and when, drug-use self-reports are valid. To this end we explored the relationships between urine drug screen results and self-reported substance use among adolescents and young adults with opioid dependence participating in a clinical trial of buprenorphine-naloxone. In this study, 152 individuals seeking treatment for opioid dependence were randomized to a 2-week detoxification with buprenorphine-naloxone (DETOX) or 12 weeks buprenorphine-naloxone (BUP), each with weekly individual and group drug counseling. Urine drug screens and self-reported frequency of drug use were obtained weekly, and patients were paid \$5 for completing weekly assessments. At weeks 4, 8, and 12, more extensive assessments were done, and participants were reimbursed \$75. Self-report data were dichotomized (positive vs. negative), and for each major drug class we computed the kappa statistic and the sensitivity, specificity, positive predictive value, and negative predictive value of self-report using urine drug screens as the “gold standard”. Generalized linear mixed models were used to explore the effect of treatment group assignment, compensation amounts, and participant characteristics on self-report. In general, findings supported the validity of self-reported drug use. However, those in the BUP

© 2013 Elsevier Ltd. All rights reserved.

Corresponding Author: Claire Wilcox, MD, University of New Mexico Department of Psychiatry, 1 University of New Mexico, MSC 09-5030, Albuquerque, NM 87131, Phone: 505-272-2223, cewilcox@salud.unm.edu.

Contributors

Drs. Bogenschutz, Wilcox, and Nakazawa designed the secondary analyses for this paper. Dr. Woody designed and ran the parent study and participated in writing this paper. Dr. Wilcox conducted literature searches and provided summaries of previous research studies. Dr. Nakazawa conducted the statistical analyses. Dr. Wilcox wrote the first draft of the manuscript and all authors contributed to and have approved the final manuscript.

Conflicts of Interest

All authors declare they have no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

group were more likely to under-report cocaine and opioid use. Therefore, if used alone, self-report would have magnified the treatment effect of the BUP condition.

Keywords

Concordance; Treatment Research; Self-Report; Urine Drug Screen; Adolescent; Substance Use Disorder

1. Introduction

In treatment studies of patients with substance use disorders, obtaining valid drug use outcome data can be challenging. Urine test data are often used as primary outcome measures because self-reported drug use data can be invalid (Lavori, Bloch, Bridge, et al., 1999; Winhusen, Somoza, Singal, et al., 2003). However, objective measures come with their own complications including high cost, varying and sometimes narrow windows for detection, and inaccuracy (Lavori, et al., 1999; Winhusen, et al., 2003). Rather than simply dismissing self-report data, it may be more useful to identify factors that influence the accuracy of self-reports and characteristics of individuals that are more, or less, likely to give accurate reports, as, often, self-report is an adequate measure of substance use (Babor, Steinberg, Anton, & Del Boca, 2000; Brown, Kranzler, & Del Boca, 1992; Del Boca & Darkes, 2003; Del Boca & Noll, 2000). Certain study design factors may increase accuracy, such as more rigorous information-gathering methods or not having contingencies for drug use (Darke, 1998; Del Boca & Noll, 2000; Sherman & Bigelow, 1992). Ongoing examination of this issue is important because of continuing changes in the nature, distribution, and demography of drug use. More research is needed to identify factors that may influence concordance between self-reports and urine toxicology test results, and how they vary in different populations. This study extends concordance examination to a new population: youthful opioid abusers being treated with a relatively new pharmacotherapy (buprenorphine) for opioid dependence (Woody, Poole, Subramaniam, et al., 2008).

The unique benefits of self-report procedures are flexibility, adaptability, relatively low cost, efficiency, portability, and the possibility of collecting data through a variety of technologies such as telephone, computer and even video (Del Boca & Noll, 2000). Some have found that self-reports are as sensitive as, and may sometimes be more sensitive than, objective measures when data are collected with clear instructions to respondents combined with methods to improve their motivation and facilitate cognitive processing (Babor, et al., 2000; Del Boca & Noll, 2000). For example, a study of psychiatric patients in an emergency department showed that for marijuana, self-report was more sensitive than urinalysis (Perrone, De Roos, Jayaraman, & Hollander, 2001). Urine assay procedures can be inaccurate, increasing the relative validity of self-report (Akinci, Tarter, & Kirisci, 2001; Brown, et al., 1992; Jain, 2004; Magura, Goldsmith, Casriel, Goldstein, & Lipton, 1987; Perrone, et al., 2001; Sherman & Bigelow, 1992; Solbergdottir, Bjornsson, Gudmundsson, Tyrfinsson, & Kristinsson, 2004; Zanis, McLellan, Cnaan, & Randall, 1994).

Under reporting of drug use may vary according to drug class, though there is little consensus on which classes are more affected (Brown, et al., 1992; Darke, 1998; Falck, Siegal, & Carlson, 1992; Magura, et al., 1987; Perrone, et al., 2001; Sherman & Bigelow, 1992; Solbergdottir, et al., 2004; Zanis, et al., 1994). Over-reporting use (reporting positive when urine screen is negative) may also occur but is less frequent than under-reporting, and findings of over reporting may be due to inaccuracy of the assay procedure (Akinci, et al., 2001; Brown, et al., 1992; Jain, 2004; Magura, et al., 1987; Perrone, et al., 2001; Sherman & Bigelow, 1992; Solbergdottir, et al., 2004; Zanis, et al., 1994). Contingencies also affect the

validity of self-reports. For example, patients applying for methadone treatment may over-report opioid use because they are afraid that they will not qualify for treatment or that the physician will not prescribe a dose that prevents withdrawal (Digiusto, Seres, Bibby, & Batey, 1996; Sherman & Bigelow, 1992), while those on methadone treatment may under-report to avoid disapproval, termination of treatment, or loss of take-home privileges. Other contextual factors may also affect self-report accuracy, such as whether interviewers are para-professionals or professionals, the way questions are asked, whether strategies to enhance recall are used, conditions under which the data are obtained (treatment, research, occupational), perceived confidentiality, and whether the patient directly enters self-reports into a computer or provides them during an interview with a clinician or research technician (Del Boca & Noll, 2000; Digiusto, et al., 1996; Schumacher, Milby, Raczynski, et al., 1995; Sherman & Bigelow, 1992).

Finally, patient factors can also influence the validity of self-report. For example, pregnancy is associated with more under-reporting, likely related to fear of losing custody or criminal retribution (Marques, Tippetts, & Branch, 1993). Employment, African American race, diagnosis of histrionic personality disorder and cognitive deficits have been associated with under-reporting, whereas diagnoses of dependent personality, passive-aggressive personality or axis I affective disorders have been associated with less under-reporting (Babor, et al., 2000; Del Boca & Noll, 2000; Fendrich, Mackesy-Amity, Johnson, Hubbell, & Wislar, 2005). Some (Solbergdottir, et al., 2004) but not all studies (Kilpatrick, Howlett, Sedgwick, & Ghodse, 2000) have found younger age to be negatively correlated with under-reporting; adolescents may be especially influenced by social pressure of peers, characteristics of the adult examiner, and perceived threat to confidentiality (Schwarz, 1999). Factors that have not reliably been predictive include gender, past criminality, and antisocial personality disorder (Digiusto, et al., 1996; Magura, et al., 1987).

In view of these inconsistent findings on the validity of self-reports, we conducted a secondary analysis of self-report and urine test data from a randomized trial of buprenorphine-naloxone treatment for opioid addicted youth done by the NIDA Clinical Trials Network (CTN) (Woody, et al., 2008). Although the primary outcome was opioid use as measured by urine test results at weeks 4, 8 and 12, weekly self-report and urine test data were collected on use of cocaine, opiates, amphetamines, benzodiazepines, and cannabis. These data allowed us to explore predictors of concordance between urine drug tests and self-reports. Consistent with existing evidence, we hypothesized that concordance would be reasonably high for most drugs, and that self-report would be more specific than sensitive since patients tend to under-report more than over-report. We also hypothesized that self-reported positives would be lower in the BUP than DETOX group regardless of drug screen results due to greater engagement in treatment and desire to please the providers. Finally, in an exploratory analysis, we evaluated other subject factors that were previously shown to be associated with the validity of self-reported drug use.

2. Materials and Methods

2.1 Participants and Outcomes

In the parent study, 152 subjects aged 15–21 seeking treatment for opioid dependence were randomized to a 2-week detoxification with buprenorphine-naloxone (DETOX; N=78), or 12 weeks of buprenorphine-naloxone (BUP; N=74), with a dose taper beginning in week 9 and ending in week 12, each with weekly individual and group drug counseling (Woody, et al., 2008). Subjects were paid \$5 for weekly assessments which included urine drug screen and self-report of drug use, and \$75 for more extensive assessments at weeks 4, 8 and 12. Weekly assessments took approximately 30 minutes, and monthly assessments (weeks 4, 8, and 12) took approximately 90 minutes. Participants were asked “In the past week how

many days did you use: [heroin, methadone, other opiates, benzodiazepines, cocaine, amphetamines, methamphetamines, cannabis?]" A dichotomous self-report response was created as follows: for cocaine, cannabis, and benzodiazepines, if participants indicated non-zero days of use, the response was coded as "1" for each drug; otherwise, as "0". For amphetamines, participants' responses to methamphetamine and amphetamine were first combined, and non-zero responses in either group were coded as "1". Similarly, for opioids, participants' responses to heroin, methadone, and other opiates were first combined, and non-zero responses were coded as "1". The same questions were used for the more extensive monthly assessments.

The urinalyses for drugs of abuse were performed on site utilizing the SureStep drug screen card (which tests for all drugs noted above except Oxycodone but does include a test for tricyclic antidepressants) and the Rapid One OXY on-site urine drug screen for Oxycodone. Cutoffs in ng/ml were as follows: amphetamines (1000ng/ml), barbiturates (300 ng/ml), benzodiazepines (300 ng/ml), cocaine (benzoylecgonine) (300ng/ml), methadone (300 ng/ml), methamphetamine (1000 ng/ml), morphine (hydrocodone, hydromorphone, heroin) (2000 ng/ml), phencyclidine (PCP) (25 ng/ml), tetrahydrocannabinol (THC) (50 ng/ml) and oxycodone (100 ng/ml). Urine samples were assigned a positive or negative value for each of five groups: opioids (morphine/opiates, methadone, and oxycodone), cocaine, cannabis, benzodiazepines, and amphetamines (methamphetamine and amphetamine).

Based on these values, five measures of concordance of self-report with urine samples were computed: Cohen's kappa (κ), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). In this case we used the urine toxicology result as the "gold standard"; thus "true positive" was defined as having a positive urine toxicology screen result. κ is a statistical measure of inter-rater agreement or inter-annotator agreement for categorical items, sensitivity reflects the proportion of true positives (based on urine drug screen results) which are correctly identified as such by self-report, specificity reflects the proportion of true negatives which are correctly identified as such, positive predictive value reflects the proportion of positive self-reports that are true positives, and negative predictive value reflects the proportion of negative self-reports that are true negatives.

2.2 Statistical Methods

To analyze the concordance between two longitudinal dichotomous variables, we used Generalized Linear Mixed Modeling (GLMM), an extension of the generalized linear model that can accommodate random effects, as well as the correlated nature of repeated measures data (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). Based on maximum-likelihood estimation, GLMM provides results in terms of logit (x) = $\log(x)/(1-x)$. Logit (x) can be converted into an odds ratio (OR) by exponentiating it. GLMM was performed based on the R lme4 package's *lmer* function.

Aim 1—Examine overall relationships between self-report and urine test results for the five drugs tested. This primarily descriptive aim involved computing the five measures of concordance (κ , sensitivity, specificity, PPV, NPV) for the drugs at the 12 assessment time points and then calculating the mean and standard deviation of those twelve values, weighted based on the number of observations at each of the time-points.

Aim 2—Determine if treatment assignment was associated with differences in validity of self-report, using urine toxicology results as the "gold standard". Specifically, we tested the hypothesis that participants assigned to BUP would be more likely to have negative self-reports of drug use regardless of whether urine drug screens were positive or negative. To test this hypothesis, we created a GLMM with self-report as the outcome variable for each

of the five drugs. In this model, an interaction between treatment and urine test results would imply that treatment assignment moderated the relationship between urine and self-report results. In the absence of any such effect, a treatment effect would represent the effect of treatment condition on self-report independent of the urine results, the urine result representing the actual “true” outcome, and the treatment effect representing the probability of over-reporting or under-reporting depending on the sign of the logit. For each drug, we initially included the main effects of treatment (DETOX vs. BUP: 0, 1) and urine (NEGATIVE vs. POSITIVE: 0, 1) factors as well as their interaction (Full Model; FM). When the interaction was not significant, we removed it and refit an Alternative Model (AM) excluding the interaction. Then we compared the two models using a log likelihood-ratio test [$-2\ln(L_{AM}/L_{FM}) = \chi^2(1)$], and Akaike’s and Bayesian Information Criteria (AIC and BIC). If AM fit the data as well as, or better than, FM [i.e., $\chi^2(1) < 3.84$; AIC_{FM-AM} and $BIC_{FM-AM} > 0$], results from the AM were used preferentially.¹ This model-comparison approach was necessary because removal of a non-significant effect can still result in a significant detriment of model fit.

Aim 3—Explore if participant characteristics or compensation amounts were associated with differences in the likelihood of positive self-report and/or the relationship between urine samples and self-report (moderator analysis). The following predictors were examined: month, compensation, age, sex, the number of crimes committed, employment past 30 days, employment past 3 years, Heroin Severity of Dependence Scale (SDSS) score (Miele, Carpenter, Smith Cockerham, et al., 2000), and education completed.

Because some of the weeks contained few observations, especially for infrequently used drugs, four weeks were grouped into each month; that is, month 1 = weeks 1–4; month 2 = weeks 5–8; month 3 = weeks 9–12. This “Monthly” factor was centered so that the intercepts represented average values at month 1 instead of month 0. The compensation factor consisted of two levels: HIGH weeks (4, 8, and 12; \$75) and LOW weeks (all other weeks; \$5). Both age and education completed were measured in years and centered so that the intercepts would indicate average values at the mean age of 19.7 and at the mean years of 11.2, respectively. Employment past 30 days and employment past 3 years consisted of three categories: 1 = NON-EARNER (service, retired/disability, in controlled environment); 2 = PART-TIMER part-time, students, home-makers, retirees); and 3 = FULL-TIMER. Both variables were centered so that the intercept indicated the likelihood of positive self-report among part-time employees. The number of crimes committed factor represented how many crimes the participants had committed in their lifetime. Close to 77% (156/202) reported no previous crimes. The Heroin SDSS score was computed by averaging the self-reported days of meeting one of the diagnostic criteria for heroin abuse or dependence (e.g., withdrawal symptoms, unsuccessful attempts to cut down use etc.) during the past 30 days (median = 24, interquartile range = 0–29).

Many of the predictors were highly correlated, causing multi-collinearity, leading to skewed estimated parameter values. Also, including several predictors and their interactions with the urine factor often resulted in non-convergence. To deal with these issues, we fit a separate model for each predictor with three effects: two main effects for the predictor and for urine and their interaction effect.

¹Because smaller AIC and BIC indicate better model fit given the number of parameters estimated, AIC_{FM-AM} and $BIC_{FM-AM} > 0$ indicates that AM fits the data better than FM.

3. Results

Aim 1: Examine the overall relationship between urine results and self-reports

Here we describe relationships between self-report and urine results for the five drugs tested (Table 1; Figure 1). Cohen's κ averaged across 12 weeks was relatively high, equal to or exceeding 0.7 in all drugs except benzodiazepines (0.55). For some drugs κ fluctuated widely across weeks: for benzodiazepines the range was 0.04~1.00 and the weighted standard deviation (wSD) was 0.32; for amphetamines the range was 0.00~1.00 and the wSD was 0.28. For other drugs, there was less fluctuation (SD for opioids, cocaine, and cannabis = 0.06, 0.12, 0.04 respectively). Regarding sensitivity, self-reported benzodiazepine use had the lowest sensitivity with a weighted mean of 0.56 while self-report for the other drugs had a weighted mean sensitivity equal to or greater than 0.67. Like Cohen's κ , sensitivity fluctuated widely for benzodiazepines (wSD = 0.34, range = 0.00~1.00) and for amphetamines (wSD = 0.30, range = 0.33~1.00, excluding one missing result). Self-report for all drugs had very high specificity averaged across 12 weeks with the highest of 0.99 for amphetamines and benzodiazepines, and the lowest of 0.89 for cannabis. Specificity was more stable than the other measures: the greatest wSD was 0.04 in cannabis and the lowest of 0.01 in amphetamines, cocaine, and benzodiazepines.

Self-report for all drugs had a relatively high probability that a positive report reflected a positive urine screen (PPV), with benzodiazepines having the lowest weighted mean of 0.64 and amphetamines the second lowest PPV of 0.76. Like Cohen's κ and sensitivity, PPV for self-reported use fluctuated widely for benzodiazepines (wSD = 0.38, range = 0.00~1.00, excluding one missing result) and for amphetamines (wSD = 0.27, range = 0.00~1.00). In terms of the negative predictive value of self-report (probability that a negative report reflected a negative urine drug screen), the drugs had high means with opioids having the lowest weighted mean of 0.90. Like specificity, NPV was very stable. The greatest wSD was 0.04 in opioids and the lowest of 0.01 in amphetamines, cocaine, and benzodiazepines.

Aim 2: Determine if specific treatment assignment was associated with differences in the relationship between urine test results and self-report

Here we tested whether treatment assignment altered the relationship between self-report and the urine tests (Table 2). An interaction between treatment and urine test results would imply that treatment assignment moderated this relationship. In the absence of any such effect (as was seen here), the treatment effect would be the effect of treatment condition on self-report independent of urine results. A positive logit value for treatment effect would therefore reflect greater under-reporting in those assigned to BUP in comparison to those assigned to DETOX.

To account for multiple comparisons for the five drug classes tested we used a Bonferroni-corrected p-value ($0.05/5=0.01$) as a threshold of significance for Aim 2. The full model fit to the benzodiazepines data did not converge, (the statistical model was too complicated to provide a valid result given the small amount of data). Therefore, this drug was excluded from further analysis. The urine by treatment interaction was not significant for any of the drugs examined ($p>0.20$, Table 2, FM). Furthermore, in all drugs examined, removing the interaction did not result in a significant detriment of the model fit, indicated by non-significant χ^2 values as well as positive AIC_{FM-AM} and BIC_{FM-AM} . Therefore, the FM (including the interaction term in model) was rejected in favor of the AM (excluding the interaction term from the model) for all classes of drug. In the AM, treatment assignment significantly altered participants' self-report for opioids and cocaine. Specifically, participants who had been assigned to BUP were significantly less likely to self-report use of opioids and cocaine, regardless of Urine ($p < 0.01$, Table 2, AM). This implies that

treatment assignment may modify self-report, regardless of the urine findings, indicating that participants assigned to BUP were more likely to under-report.

Aim 3. Explore if participant characteristics or compensation amounts were associated with differences in the likelihood of positive self-report and/or the relationship between urine samples and self-report

By analogy to Aim 2, an interaction effect between participant characteristic and urine test results would imply that participant characteristics may be modifying the relationship between self-report and urine result. In the absence of any such effect, the participant characteristic is the effect of said characteristic on the probability of either over-or under-reporting. We report all significant main effects of participant characteristic and significant interaction effects ($p < 0.05$) in Table 3. Non-significant ($p > 0.05$) effects are not reported. We did not correct for multiple comparisons for Aim 3, as this aim was exploratory. For benzodiazepines, all the models failed to converge and were excluded. For amphetamines, employment past 3 years, employment past 30 days, number of crimes committed, and Heroin SDSS score models converged falsely and were excluded.

For amphetamines, compensation amount was a significant moderator in that higher compensation amounts predicted a greater likelihood of self-report positive independent of urine results. For Cannabis, gender emerged as a significant moderator, in that males were more likely to report positive given positive urine than females, indicated by the significant positive sex by urine interaction ($p = 0.03$). Furthermore, for cannabis, employment past 3 years and employment past 30 days interacted with urine results such that employment predicted a lower odds of reporting positive given positive urine (significant interactions, $p < 0.05$), and of reporting positive given negative urine (significant main effects, $p = 0.059$ for employment past 30 days; $p < 0.05$ for employment past 3 years). Finally, also for cannabis, Heroin SDSS score also interacted with the urine results: a lower severity of opiate dependence was associated with a higher probability of self-reporting positive given positive urine results. For cocaine and opioids, there were no significant main effects of participant characteristics or characteristic X urine interactions in any of the models.

4. Discussion

In this sample of adolescents and young adults participating in an opioid dependence treatment study, concordance between self-report and urine drug screen results for most drugs (except benzodiazepines) was reasonably high (κ 's > 0.7). In addition, the specificity of self-report was quite high [i.e., the percentage of non-drug using people who "correctly" (based on urine drug screen result) report negative use] but sensitivity [i.e., the percentage of drug-using people who "correctly" (based on urine drug screen result) report positive] was somewhat lower, although usually still > 0.7 . PPV and NPV (using urine drug screen result as the "gold standard") were high as well (> 0.7) indicating that self-report of drug use for cannabis, cocaine, opioids and, for the most part amphetamines, was valid, and usually consistent with urine toxicology.

These values are higher than those reported in many previous studies (Brown, et al., 1992; Darke, 1998; Digiusto, et al., 1996; Magura, et al., 1987; Perrone, et al., 2001; Sherman & Bigelow, 1992; Solbergstodt, et al., 2004; Zanis, et al., 1994). Some studies have seen higher rates of self-report accuracy in adolescent populations compared to adults (Akinci, et al., 2001; Magura, et al., 1987; Solbergstodt, et al., 2004), which could explain the surprisingly high validity of self-report in this study. Also, consistent with previous studies demonstrating that individuals under-report drug use more than they over-report, sensitivity tended to be slightly lower than specificity in our study (Akinci, et al., 2001; Brown, et al., 1992; Jain, 2004; Magura, et al., 1987; Perrone, et al., 2001). The reasonably high NPV in

this study alleviates concerns that negative self-report is invalid as a high NPV indicates that a self-report of non-use is most likely correct (Magura, et al., 1987; Solbergdottir, et al., 2004).

Self-report of opioid or cocaine use was less likely to be positive in participants assigned to BUP, regardless of urine drug screen status, most likely due to under-reporting in the BUP group. These results argue against reliance on self-report as a primary outcome in open label clinical trials where participants are expected to prefer one medication over the placebo medication, as use of self-report could result in a spurious magnification of treatment effects.

There were some interesting but inconclusive findings (not corrected for multiple tests) in the exploratory analyses. These findings should be interpreted with caution, and would need to be replicated in other studies to be considered generalizable. That said, we saw in this study that employment was associated with a decreased probability of positive self-reports given positive urine results for cannabis, consistent with previous literature (Myrick, Henderson, Dansky, Pelic, & Brady, 2002). Also for cannabis, male gender was associated with an increased probability of self-report positive when urines were positive. Although gender-related differences in substance use appear to be shrinking over time, males may be more likely to start using illicit substances at a younger age than females, and report greater use of illicit substances, except for prescription medications, which females use in greater quantity than males (Shannon, Havens, Oser, Crosby, & Leukefeld, 2011). Given the strong relationship between social factors and substance use (Galea, Nandi, & Vlahov, 2004), it would not be surprising if males were more inclined than females to report true drug use to an interviewer due to less shame about drug use, however previous literature has not identified gender as a predictor of under-reporting (Solbergdottir, et al., 2004) and it is not clear why this effect would be specific to cannabis. In addition, higher severity of heroin dependence was associated with a lower probability of reporting cannabis use (under-reporting of cannabis use), consistent with prior studies showing that heavier heroin users were more likely to be non-disclosers of other drug (cocaine) use (Tassiopoulos, Bernstein, Heeren, et al., 2004), and with studies showing that higher levels of cocaine dependence predict a greater a likelihood of under-reporting (Myrick, et al., 2002).

Finally, higher compensation increased the probability that patients with positive urines for amphetamines self-reported positive. In this study, there were much higher no-show rates (>50%) for low compensation visits compared to high compensation visits (25–50%). Compensations amounts may affect the characteristics of individuals that show up for assessments, and in particular higher compensation may be more likely to bring in active drug users for assessments (Wilcox, Bogenschutz, Nakazawa, & Woody, 2011).

There was wide variability in the ability of self-report to predict urine results for benzodiazepines and amphetamines. For benzodiazepines and amphetamines, concordance, sensitivity, specificity, PPV, and NPV of self-report were generally lower and less stable. These results may be due to smaller numbers of positive cases. For instance, there were only 32 (+Self-Report | +Urine) cases summed across 12 weeks and 152 participants in the benzodiazepines group and 24 cases in the cocaine group, whereas in other drugs the numbers ranged from 103 to 372. In the case of benzodiazepines, variability could also be due to the long half-life of some drug metabolites (diazepam, for example), the unreliability of urine screens to pick up certain benzodiazepines, and the potential for other medications to trigger false positive urine benzodiazepine results (Tenore, 2010).

A few limitations of this study warrant mention. Although the application of mixed-effects modeling (i.e., GLMM) enhances the generalizability of our findings, this was a secondary

analysis rather than a prospective, hypothesis-driven study. In addition, the participants knew they were going to have urine toxicology tests when they were asked for self-reports. Therefore, these findings may not generalize to situations where urines are not obtained. Also, as mentioned previously, the findings for Aim 3 were not corrected for multiple comparisons and must be considered exploratory. Finally, the finding that compensation moderated the probability of self-reporting positive when urines were positive for amphetamines should be interpreted with caution. It could be related to the compensation amount, or to other factors associated with the higher compensation visits (longer visits, higher attention).

In summary, our data imply that self-report of drug use may be a valid outcome measure in treatment studies of adolescents and young adults with opioid use disorders. Future studies could further define the predictors of greater or lower probability of over- and under-reporting, so that interpretations of treatment studies using self-report as primary outcome measures can be more accurate.

Acknowledgments

The work was supported by the NIDA Clinical Trials Network grants U10DA15833 (Bogenschutz) and U10DA13043 and KO5 DA-17009 (Woody).

Role of Funding Source

This study was supported by the NIDA Clinical Trials Network grants U10DA15833 (Bogenschutz) and U10DA13043 and KO5 DA-17009 (Woody). NIDA had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

References

- Akinci IH, Tarter RE, Kirisci L. Concordance between verbal report and urine screen of recent marijuana use in adolescents. *Addict Behav.* 2001; 26(4):613–619. [PubMed: 11456081]
- Babor TF, Steinberg K, Anton R, Del Boca F. Talk is cheap: measuring drinking outcomes in clinical trials. *J Stud Alcohol.* 2000; 61(1):55–63. [PubMed: 10627097]
- Brown J, Kranzler HR, Del Boca FK. Self-reports by alcohol and drug abuse inpatients: factors affecting reliability and validity. *Br J Addict.* 1992; 87(7):1013–1024. [PubMed: 1322747]
- Darke S. Self-report among injecting drug users: a review. *Drug Alcohol Depend.* 1998; 51(3):253–263. discussion 267–258. [PubMed: 9787998]
- Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction.* 2003; 98(Suppl 2):1–12. [PubMed: 14984237]
- Del Boca FK, Noll JA. Truth or consequences: the validity of self-report data in health services research on addictions. *Addiction.* 2000; 95(Suppl 3):S347–360. [PubMed: 11132362]
- Digiusto E, Seres V, Bibby A, Batey R. Concordance between urinalysis results and self-reported drug use by applicants for methadone maintenance in Australia. *Addict Behav.* 1996; 21(3):319–329. [PubMed: 8883483]
- Falck RS, Siegal HA, Carlson RG. Case management to enhance AIDS risk reduction for injection drug users and crack cocaine users: practical and philosophical considerations. *NIDA Res Monogr.* 1992; 127:167–180. [PubMed: 1435994]
- Fendrich M, Mackesy-Amiti ME, Johnson TP, Hubbell A, Wislar JS. Tobacco-reporting validity in an epidemiological drug-use survey. *Addict Behav.* 2005; 30(1):175–181. [PubMed: 15561458]
- Galea S, Nandi A, Vlahov D. The social epidemiology of substance use. *Epidemiol Rev.* 2004; 26:36–52. [PubMed: 15234946]
- Jain R. Self-reported drug use and urinalysis results. *Indian J Physiol Pharmacol.* 2004; 48(1):101–105. [PubMed: 15270376]
- Kilpatrick B, Howlett M, Sedgwick P, Ghodse AH. Drug use, self report and urinalysis. *Drug Alcohol Depend.* 2000; 58(1–2):111–116. [PubMed: 10669061]

- Lavori PW, Bloch DA, Bridge PT, Leiderman DB, LoCastro JS, Somoza E. Plans, designs, and analyses for clinical trials of anti-cocaine medications: where we are today. NIDA/VA/SU Working Group on Design and Analysis. *J Clin Psychopharmacol*. 1999; 19(3):246–256. [PubMed: 10350031]
- Magura S, Goldsmith D, Casriel C, Goldstein PJ, Lipton DS. The validity of methadone clients' self-reported drug use. *Int J Addict*. 1987; 22(8):727–749. [PubMed: 3679632]
- Marques PR, Tippetts AS, Branch DG. Cocaine in the hair of mother-infant pairs: quantitative analysis and correlations with urine measures and self-report. *Am J Drug Alcohol Abuse*. 1993; 19(2):159–175. [PubMed: 8484355]
- Miele GM, Carpenter KM, Smith Cockerham M, Trautman KD, Blaine J, Hasin DS. Substance Dependence Severity Scale (SDSS): reliability and validity of a clinician-administered interview for DSM-IV substance use disorders. *Drug Alcohol Depend*. 2000; 59(1):63–75. [PubMed: 10706976]
- Myrick H, Henderson S, Dansky B, Pelic C, Brady KT. Clinical characteristics of under-reporters on urine drug screens in a cocaine treatment study. *Am J Addict*. 2002; 11(4):255–261. [PubMed: 12584868]
- Perrone J, De Roos F, Jayaraman S, Hollander JE. Drug screening versus history in detection of substance use in ED psychiatric patients. *Am J Emerg Med*. 2001; 19(1):49–51. [PubMed: 11146019]
- Schumacher JE, Milby JB, Raczynski JM, Caldwell E, Engle M, Carr J, et al. Validity of self-reported crack cocaine use among homeless persons in treatment. *J Subst Abuse Treat*. 1995; 12(5):335–339. [PubMed: 8583516]
- Schwarz N. Self-Report: How the Questions Shape the Answers. *American Psychologist*. 1999; 54(2):93–105.
- Shannon LM, Havens JR, Oser C, Crosby R, Leukefeld C. Examining gender differences in substance use and age of first use among rural Appalachian drug users in Kentucky. *Am J Drug Alcohol Abuse*. 2011; 37(2):98–104. [PubMed: 21142705]
- Sherman MF, Bigelow GE. Validity of patients' self-reported drug use as a function of treatment status. *Drug Alcohol Depend*. 1992; 30(1):1–11. [PubMed: 1317286]
- Solbergstodtir E, Bjornsson G, Gudmundsson LS, Tyrfinngsson T, Kristinsson J. Validity of self-reports and drug use among young people seeking treatment for substance abuse or dependence. *J Addict Dis*. 2004; 23(1):29–38. [PubMed: 15077838]
- Tassiopoulos K, Bernstein J, Heeren T, Levenson S, Hingson R, Bernstein E. Hair testing and self-report of cocaine use by heroin users. *Addiction*. 2004; 99(5):590–597. [PubMed: 15078233]
- Tenore PL. Advanced urine toxicology testing. *J Addict Dis*. 2010; 29(4):436–448. [PubMed: 20924879]
- Wilcox CE, Bogenschutz M, Nakazawa M, Woody GE. Compensation effects on clinical trial data collection in opioid dependent young adults. *American Journal of Drug and Alcohol Abuse*. 2011 (In press).
- Winhusen TM, Somoza EC, Singal B, Kim S, Horn PS, Rotrosen J. Measuring outcome in cocaine clinical trials: a comparison of sweat patches with urine toxicology and participant self-report. *Addiction*. 2003; 98(3):317–324. [PubMed: 12603231]
- Woody GE, Poole SA, Subramaniam G, Dugosh K, Bogenschutz M, Abbott P, et al. Extended vs short-term buprenorphine-naloxone for treatment of opioid-addicted youth: a randomized trial. *JAMA*. 2008; 300(17):2003–2011. [PubMed: 18984887]
- Zanis DA, McLellan AT, Cnaan RA, Randall M. Reliability and validity of the Addiction Severity Index with a homeless sample. *J Subst Abuse Treat*. 1994; 11(6):541–548. [PubMed: 7884837]
- Zuur, AF.; Leno, EN.; Walker, NJ.; Saveliev, AA.; Smith, GM. *Mixed Effects Models and Extensions in Ecology in R*. New York: Springer-Verlag; 2009.

Highlights

We examine concordance between urine drug screen results and drug use self-report.

We examine factors that influence self-report validity.

In general, self-report is a valid measure of drug use, but adding urine tests improves detection of drug use.

Figure 1. Proportions of samples positive by self-report, by urinalysis, and by the two combined (i.e., either self report positive or urine toxicology positive counts as positive) for each drug class. N = 923 across all 12 weeks.

Table 1Cohen's κ , Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value for Five Drugs

Drug	Variable	Weighted Mean	Weighted SD
Opioids	Kappa	0.72	0.06
	Sensitivity	0.79	0.08
	Specificity	0.92	0.03
	PPV	0.84	0.07
	NPV	0.90	0.04
Cocaine	Kappa	0.75	0.12
	Sensitivity	0.73	0.14
	Specificity	0.98	0.01
	PPV	0.88	0.11
	NPV	0.95	0.02
Amphetamines	Kappa	0.69	0.28
	Sensitivity	0.67	0.30
	Specificity	0.99	0.01
	PPV	0.76	0.27
	NPV	0.99	0.01
Cannabis	Kappa	0.79	0.04
	Sensitivity	0.91	0.03
	Specificity	0.89	0.04
	PPV	0.87	0.03
	NPV	0.92	0.03
Benzodiazepines	Kappa	0.55	0.32
	Sensitivity	0.56	0.34
	Specificity	0.99	0.01
	PPV	0.64	0.38
	NPV	0.97	0.02

Note: PPV = Positive Predictive Value; NPV = Negative Predictive Value; NA = Not Available; Weighted Mean = The mean of weekly values calculated at each of the 12 weeks, weighted by the number of valid observations within each week; Weighted SD = The standard deviation of the weekly values at each of the 12 weeks, weighted by the number of valid observations within each week.

Table 2

Test of Treatment, Urine, and Interaction Effects for Five Drugs

Factors	FM			AM			Fit Difference (FM - AM)		
	Logit(p)	S.E.	p	Logit(p)	S.E.	p	$\chi^2(1)$	AIC	BIC
Opioids	Intercept	-1.98	0.34	0.000	-2.13	0.30	0.000	0.87	1.13
	Urine	4.07	0.42	0.000	4.36	0.29	0.000		
	TX	-1.46	0.47	0.002	-1.18	0.36	0.001		
	Urine by TX	0.53	0.57	0.359					
Cocaine	Intercept	-4.31	0.60	0.000	-4.02	0.48	0.000	1.70	0.3
	Urine	6.60	0.78	0.000	5.97	0.52	0.000		
	TX	-1.06	0.85	0.213	-1.64	0.57	0.004		
	Urine by TX	-1.20	1.05	0.250					
Amphetamines	Intercept	-11.66	4.85	0.016	-9.91	2.93	0.001	0.77	1.23
	Urine	6.50	4.50	0.149	5.23	1.67	0.002		
	TX	2.25	6.26	0.719	1.02	3.97	0.798		
	Urine by TX	-1.83	4.82	0.705					
Cannabis	Intercept	-2.65	0.58	0.000	-2.31	0.49	0.000	1.64	0.37
	Urine	5.27	0.75	0.000	4.64	0.53	0.000		
	TX	0.24	0.81	0.770	-0.41	0.62	0.510		
	Urine by TX	-1.34	1.06	0.210					
Benzodiazepines	Intercept								
	Urine								
	TX								
	Urine:TX								

Note: TX = Treatment; $\chi^2(df = 1)$ critical value = 3.84; FM = Full Model; AM = Alternative Model; Positive AIC and BIC indicates that AM is the preferred model given the number of parameters it estimates.

Table 3

Test of Participant Characteristic, Urine Sample, and Interaction Effects for Five Drugs

Amphetamines			
Factors	Logit	S.E.	P
(Intercept)	11.97	2.55	0.000
Urine	6.89	2.01	0.001
Compensation	2.49	1.24	0.045
Compensation x Urine	-1.90	1.98	0.338

Cannabis			
Factors	Logit	S.E.	P
(Intercept)	-1.80	0.53	0.001
Urine	3.41	0.71	0.000
Sex	-1.33	0.8	0.095
Sex x Urine	2.27	1.05	0.030
(Intercept)	-4.21	0.77	0.000
Urine	7.82	1.09	0.000
Employment Past 30 Days	0.73	0.39	0.059
Employment Past 30 Days x Urine	-1.28	0.51	0.012
(Intercept)	-5.65	1.40	0.000
Urine	9.24	1.86	0.000
Employment Past 3 Years	1.11	0.56	0.049
Employment Past 3 Years x Urine	-1.48	0.75	0.048
(Intercept)	-3.73	0.76	0.000
Urine	8.28	1.29	0.000
Heroin SDSS Score	0.03	0.03	0.359
Heroin SDSS Score x Urine	-0.11	0.05	0.024

Note: Only significant ($P < 0.05$) effects of Participant Characteristic and the Urine X Participant Characteristic reported (non-significant effects not reported). Participant characteristic: month, compensation, age, sex, number of crimes committed, employment past 30 days, employment past 3 years, Heroin Severity of Dependence Scale (SDSS) score, education completed. Drugs: Opiates, Amphetamines, Cannabis, Cocaine, Benzodiazepines.