J Quant Criminol (2007) 23:41–58
DOI 10.1007/s10940-006-9018-7

ORIGINAL PAPER

Differences in the Validity of Self-Reported Drug Use Across Five Factors: Gender, Race, Age, Type of Drug, and Offense Seriousness

André B. Rosay · Stacy Skroban Najaka · Denise C. Herz

Published online: 13 October 2006
© Springer Science+Business Media, LLC 2006

Abstract This study expands our knowledge about the validity of self-reported drug use by examining how gender, race, age, type of drug, and offense seriousness interact to affect the validity of self-reported drug use. This study also provides a conceptual framework that can be used to examine the validity of self-reported drug use. Differences in the validity of self-reported drug use are explained by examining differences in underreporting and overreporting. Differences in underreporting and overreporting are then further examined while controlling for differences in base rates of drug use. As shown, whether one controls for base rates of use may drastically affect estimates of underreporting and overreporting. By using hierarchical loglinear, logit, and logistic regression models with the Drug Use Forecasting data, we show that Black offenders provide less accurate self-reports than White offenders. Black offenders do so because they are more likely to underreport crack/cocaine use than White offenders. This difference, however, disappears once differences in base rates are controlled. A Black offender who tests positive is not more likely to underreport crack/cocaine use than a White offender who tests positive. Black offenders are also more likely to overreport both marijuana and crack/cocaine use relative to White offenders. Contrary to the first, this difference is not attributable to a difference in base rates. Methodological and substantive implications of this distinction are discussed. No differences across gender, age, or offense seriousness were found.

Keywords Drug use · Self-reports · Drug testing · Validity · Underreporting · Overreporting

A. B. Rosay (✉)
Justice Center, University of Alaska Anchorage, 3211 Providence Drive, Anchorage, AK 99508, USA
e-mail: afabr@uaa.alaska.edu

S. S. Najaka
Maryland State Commission on Criminal Sentencing Policy, College Park, MD, USA

D. C. Herz
California State University, Los Angeles, CA, USA

Introduction

The majority of studies examining drug use have relied on self-reported measures of drug use (Magura and Kang 1995). The results from these studies have determined how to plan and allocate drug prevention and rehabilitation services (Fendrich and Xu 1994) and the effectiveness of such services (Falck et al. 1992). These results have also influenced policy decisions such as which drug programs should be funded and expanded. In addition, individual self-reports are used every day in our justice system to determine which drug services should be offered to whom (Magura et al. 1987; Andrews et al. 1990). As we progress through an era in which drug use prevention and rehabilitation are pivotal concerns, self-reports are continuously becoming a widely used technique to measure drug use.

A well-known problem with self-reports, however, is the uncertainty about their ability to accurately indicate what is being measured (Richter and Johnson 2001). Many investigations have shown that the validity of self-reported data is questionable, especially when the topic is as sensitive as drug use. Reporting drug use, particularly while in the justice system, can have serious consequences. Individuals in the justice system may fear that disclosing drug use will intensify their involvement in the justice system, and may therefore be unlikely to disclose such information (Bale et al. 1981; Harrell 1985; Falck et al. 1992; Nelson et al. 1998). The validity of self-reports may also be affected by the general tendency to deny socially undesirable behaviors (Harrison 1995; Nelson et al. 1998; Sloan et al. 2004). Finally, unintentional errors (e.g., errors due to recall inaccuracy, the interview process, drug misidentification, and psychopharmacological effects of drug use) may also weaken the validity of self-reports (Harrison 1995; Nelson et al. 1998; Falck et al. 1992; Katz et al. 1997). Nonetheless, ''self-report measurement techniques are often preferred over biological testing, such as urinalysis, because they are more practical, less intrusive and less expensive'' (Nelson et al. 1998, p. 484). In addition, self-reports (unlike drug tests) can also measure ''the duration, frequency, intensity and other patterns of drug use, as well as the routes of administration and social context of use'' (Magura and Kang 1995, p. 9; McElrath et al. 1995).

Many investigations have examined whether self-reported drug use is a valid indicator of actual drug use. In one of the most comprehensive reviews of the literature, Magura and Kang (1995) presented a meta-analysis of 24 studies published since 1985 examining the validity of drug use reported by high risk populations. These 24 studies compared self-reported drug use with urinalysis or hair analysis results. Magura and Kang (1995) noted that ''positive self-reports were given by 42% of those subjects who had a positive urinalysis or hair analysis.'' The validity of self-reported drug use, however, varied greatly across studies. Magura and Kang (1995) hypothesized that these differences across studies were due, in part, to sample differences such as type of high risk population and type of drug use. This paper assesses the extent to which differences in the validity of self-reported drug use are due in part to sample differences.

Two types of sample differences are examined. First, we examine the extent to which the validity of self-reported drug use varies across samples stratified by base rates of drug use. Second, we examine the extent to which the validity of self-reported drug use varies across samples stratified by gender, race, age, type of drug, and offense seriousness. Stated differently, we examine the validity of self-reported

drug use across five factors (gender, race, age, type of drug, and offense seriousness). We do so with and without controlling for differences in base rates to document the effect that base rates have on the validity of self-reported drug use.

Effect of Base Rates

An individual's self-report is defined as valid if it is corroborated by a drug test result. More precisely, an individual provides a valid self-report if s/he denies using drugs and the drug test is negative or admits using drugs and the drug test is positive. Conversely, an individual provides an invalid self-report if s/he admits using drugs and the drug test is negative or denies using drugs and the drug test is positive. If an individual provides an invalid self-report, s/he either underreported their drug use (i.e., denied using drugs and the drug test is positive) or overreported their drug use (i.e., admitted using drugs and the drug test is negative). Invalid self-reports may therefore be explained in terms of underreporting or overreporting.

Mathematically, both the probability of underreporting and the probability of overreporting are affected by base rates of use, the probability of a positive drug test. As base rates of use increase, the probability of underreporting will necessarily increase. Mathematically, the probability of underreporting is the probability of a negative self-report (NS) with a positive test (PT), defined as $P\ (NS \cap PT)$. It can then be shown that $P\ (NS \cap PT) = P\ (NS \mid PT)\ P\ (PT)$, the probability of a negative self-report given a positive test multiplied by the probability of a positive test. As the probability of a positive test increases (i.e., as base rates increase), the probability of underreporting will necessarily increase. A similar argument can be developed to show that the probability of overreporting will necessarily increase as base rates of use decrease. Mathematically, the probability of overreporting is the probability of a positive self-report (PS) with a negative test (NT), defined as $P\ (PS \cap NT)$. It can then be shown that $P\ (PS \cap NT) = P\ (PS \mid NT)\ P\ (NT)$, the probability of a positive self-report given a negative test multiplied by the probability of a negative test. As the probability of a negative test increases (i.e., as base rates decrease), the probability of overreporting will necessarily increase.

Unfortunately, prior investigations on the validity of self-reported drug use have often overlooked the important difference between $P\ (NS \cap PT)$ and $P\ (NS \mid PT)$, calling both underreporting, and the important difference between $P\ (PS \cap NT)$ and $P\ (PS \mid NT)$, calling both overreporting. In our own review of the literature in the next section, it is often unclear whether prior researchers are operationalizing underreporting as $P\ (NS \cap PT)$ or $P\ (NS \mid PT)$ and operationalizing overreporting as $P\ (PS \cap NT)$ or $P\ (PS \mid NT)$. To differentiate between all of these probabilities, we will now refer to $P\ (NS \cap PT)$ as the unconditional probability of underreporting, $P\ (NS \mid PT)$ as the conditional probability of underreporting, $P\ (PS \cap NT)$ as the unconditional probability of overreporting, and $P\ (PS \mid NT)$ as the conditional probability of overreporting. We argue that underreporting and overreporting should be operationalized as conditional probabilities (i.e., $P\ (NS \mid PT)$ and $P\ (PS \mid NT)$, respectively) because these remain unaffected by base rates.

This is particularly important when examining differences in underreporting and overreporting across groups, as groups likely differ in base rates. As an example, suppose that one group has a $P\ (NS \mid PT) = \theta$, and a $P\ (PT) = \lambda$, and that a second group also has a $P\ (NS \mid PT) = \theta$, but has a $P\ (PT) = \lambda + \delta$. The unconditional probability of underreporting will then be lower in group #1 than in group #2 by $\theta\delta$,

even though the conditional probabilities are equal. This potentially large difference is solely attributable to a difference in base rates (i.e., $P(PT)$), and is clearly not due to a difference in the conditional probability of underreporting (which is θ in both groups). More substantively, differences across groups in the unconditional probabilities of underreporting and overreporting may be due to differences across groups in base rates and conditional probabilities. To uncover the differences in the conditional probabilities, we must simply control for differences in the base rates.

To summarize, we study differences in the validity of self-reported drug use by examining differences in the unconditional probabilities of underreporting and overreporting. The unconditional probabilities of underreporting and overreporting are further studied by examining differences in conditional probabilities and base rates (see Fig. 1a). Unbiased estimates of underreporting and overreporting can only be obtained via conditional probabilities.

It is important to emphasize the difference between conditional and unconditional probabilities because they have different implications. Differences in unconditional probabilities may simply imply differences in base rates (e.g., differences in the probability of testing positive). On the other hand, differences in conditional

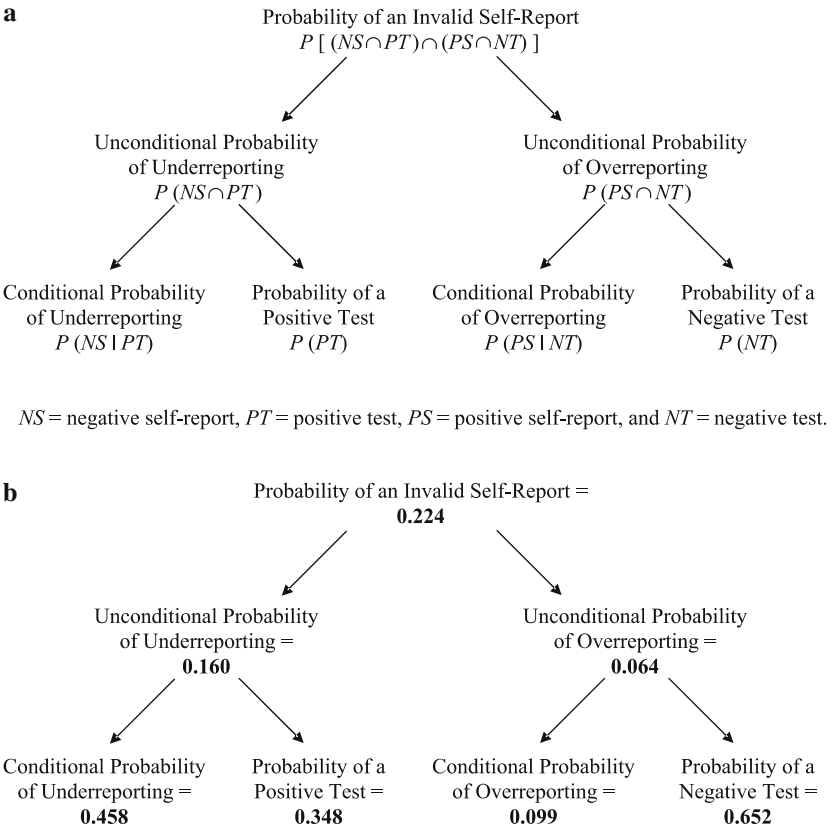

Fig. 1 (**a**) Mathematical decomposition of the probability of an invalid self-report. (**b**) Empirical decomposition of the probability of an invalid self-report

probabilities do not imply differences in base rates. To make inferences about conditional probabilities from unconditional probabilities would be misleading. Valid inferences about conditional probabilities across groups cannot be achieved without controlling for differences in base rates. Stated differently, unconditional probabilities of underreporting and overreporting confound conditional probabilities and base rates.

Effect of Gender, Race, Age, Type of Drug, and Offense Seriousness

This study focuses on five factors—gender, race, age, type of drug, and offense seriousness. Although significant research exists on the effects of these five factors, it is often unclear whether the dependent variables are conditional or unconditional probabilities. We briefly review the literature on each of these factors, focusing on the prior literature that clearly examines conditional probabilities of underreporting and overreporting (while citing other literature for further information).

No study has determined the statistical significance of differences between marijuana and crack/cocaine use (because independent samples are not created). Nonetheless, several have descriptively examined differences between marijuana and crack/cocaine (or provided enough information to do so). In particular, three studies provided enough descriptive statistics so that we could calculate differences in the validity of self-reported drug use between marijuana and crack/cocaine and determine if these differences were due to differences in conditional probabilities (Mieczkowski 1990; Stephens and Feucht 1993; Harrison 1995).

These three studies utilized the Drug Use Forecasting data to examine the validity of offenders' self-reported marijuana and crack/cocaine use. Descriptive statistics reveal that marijuana self-reports are consistently more accurate than cocaine self-reports (see also Fendrich and Xu 1994; Katz et al. 1997; Kim et al. 2000; Golub et al. 2002). Furthermore, all three studies reveal that offenders are more likely to overreport marijuana use than cocaine use and to underreport cocaine use than marijuana use (see also Fendrich and Xu 1994; Gray and Wish 1999; Kim et al. 2000; Wish et al. 2000). Stephens and Feucht's (1993) and Harrison's (1995) data reveal that the difference in underreporting is solely attributable to a difference in base rates. Stated differently, it is not a difference in conditional probabilities. Offenders are more likely to underreport cocaine use than marijuana use only because they have higher rates of cocaine use than marijuana use. On the other hand, Mieczkowski's (1990) data reveal that differences in underreporting are attributable to differences in both conditional probabilities and base rates. When controlling for differences in base rates, differences in unconditional probabilities remain. Data from all three studies indicate that the difference in overreporting is attributable to differences in both base rates and conditional probabilities.

Though not examined by Stephens and Feucht (1993), their data again provide us the opportunity to examine differences in the validity of self-reported drug use across gender groups and to examine whether these differences are attributable to differences in conditional probabilities. Their data reveal no gender differences in the conditional probabilities of underreporting or overreporting (see also Magura et al. 1987; Falck et al. 1992; Nelson et al. 1998; Hser et al. 1999; Messina et al. 2000; Kim et al. 2000; Golub et al. 2002). There are two exceptions to this general conclusion. Controlling for differences in base rates, Lu et al. (2001) found that males

were more likely to underreport crack-cocaine use than females and McElrath et al. (1995) found that males underreported more than females in Manhattan, but underreported less than females in Phoenix. No gender differences were observed in Ft. Lauderdale, Los Angeles, or St. Louis.

While several studies have examined the validity of self-reported drug use across racial groups, no consistent pattern can be noted. McElrath et al. (1995) reported no race differences in the validity of self-reported drug use in Manhattan, Ft. Lauderdale, Los Angeles, and Phoenix (see also Nelson et al. 1998). In St. Louis, however, Blacks provided less valid self-reports than Whites (see also Falck et al. 1992; Fendrich and Xu 1994; Katz et al. 1997; Kim et al. 2000). Few studies have focused on conditional probabilities. Lu et al. (2001) found that conditional probabilities of underreporting crack use were significantly higher for Whites (see also Page et al. 1977; McNagny and Parker 1992). Race had no effect on conditional probabilities of underreporting marijuana use. Similarly, Magura et al. (1987), Gray and Wish (1999), Hser et al. (1999), and Golub et al. (2002) reported no race differences in the conditional probabilities of underreporting drug use. No study has examined race differences in overreporting.

Age differences in the validity of self-reported drug use were found by Nelson et al. (1998) who showed that younger respondents provided more accurate self-reports than older respondents. On the other hand, Falck et al. (1992) and McElrath et al. (1995) reported no age differences in the validity of self-reported drug use. Furthermore, a variety of studies have shown that age does not affect the underreporting of drug use (Page et al. 1977; McNagny and Parker 1992; Fendrich and Xu 1994; Hser et al. 1999; Gray and Wish 1999; Messina et al. 2000; Kim et al. 2000). While some studies have found that age affects the underreporting of drug use (Magura et al. 1987; Falck et al. 1992; Katz et al. 1997; Sloan et al. 2004), only Lu et al. (2001) clearly show that age affects the conditional probability of underreporting. Lu et al. (2001) found that being younger significantly increased the conditional probability of underreporting crack use but significantly decreased the conditional probability of underreporting marijuana use. No study has examined age differences in overreporting.

Differences in the validity of self-reported drug use between felony and misdemeanor offenders were found by McElrath et al. (1995) in Los Angeles, but not in Manhattan, Ft. Lauderdale, Phoenix, or St. Louis (see also Katz et al. 1997). Controlling for differences in base rates, Gray and Wish (1999) reported that drug offenders were more likely to underreport than non-drug offenders. On the other hand, Kim et al. (2000) showed that when controlling for differences in base rates, drug offenders were less likely to underreport past 3-day marijuana use than non-drug offenders. No differences in conditional probabilities were found in the underreporting of past 30-day or lifetime marijuana use, or of past 3-day, past 30-day, or lifetime crack/cocaine use. Furthermore, Fendrich and Xu (1994) found no difference in conditional probabilities of underreporting across property, person, and drug offenders. No study has examined the effect of offense category on overreporting drug use.

Given the diversity in the operational definitions of underreporting and overreporting, it is difficult to synthesize the previous literature. This problem is exacerbated by the diversity in the types of high risk populations studied, the types of drug use measured, and the measurement procedures and conditions of each study (Magura and Kang 1995; Wish et al. 1997; Gray and Wish 1999). Nonetheless, three

general conclusions can be reached from our review of the previous literature. First, explaining validity differences in terms of underreporting and overreporting and explaining differences in underreporting and overreporting in terms of differences in conditional probabilities and base rates can help us organize research on the validity of self-reported drug use. Second, as shown by Stephens and Feucht (1993), Fendrich and Xu (1994), Katz et al. (1997), Hser et al. (1999), and Kim et al. (2000) significant interactions are likely to exist (e.g., gender and type of drug, race and type of drug, age and type of drug). As noted by Page et al. (1977), complex interactions in the covariates of prevarication rates should be examined. Finally, we can generally conclude that (1) marijuana self-reports are more accurate than cocaine self-reports, (2) respondents are more likely to underreport cocaine use than marijuana use, (3) respondents are less likely to overreport cocaine use than marijuana use, (4) gender, race, age, and offense category have mixed effects on the validity of self-reported drug use and may significantly interact to affect the validity of self-reported drug use. What is certain though, both mathematically and from the prior literature, is that how one operationalizes underreporting and overreporting matters a great deal.

Purpose of This Study

This study examines differences in the validity of self-reported drug use. This study further explains differences in the validity of self-reported drug use in terms of differences in underreporting and overreporting. Finally, differences in underreporting and overreporting are examined to determine whether they are attributable to differences in conditional probabilities or base rates (again, see Fig. 1a). Differences are examined across five factors—gender, race, age, type of drug, and offense seriousness—and across all possible interactions between these five factors. This is accomplished using hierarchical loglinear, logit, and logistic regression models with the 1994 Drug Use Forecasting data.

Methods

Drug Use Forecasting (DUF) Data

This study uses data collected in 1994 as part of the DUF program. Self-report surveys of drug use and urine specimens were collected from adult arrestees across 23 sites in the United States. The target population for all sites included male and female arrestees held in detention facilities. All arrestees were interviewed and asked for a urine specimen within 48 h of their arrest. Although two sites collected data from less than 100 females each quarter, DUF sites typically collected data from approximately 225 male and 100 female arrestees. Compliance rates for arrestees (both male and female) were typically high across sites, with more than 90% agreeing to the interview and over 80% agreeing to provide a urine specimen. Each site determined who would be interviewed from their detention population. As a result, some sites prioritized certain offenses over others. DUF protocol, however, encouraged site personnel to interview non-drug felony and misdemeanor offenders before those charged with a drug offense. With the exception of Omaha, traffic offenses were excluded from the target population.

Urine specimens were analyzed for ten drugs: cocaine, opiates, marijuana, PCP, methadone, benzodiazepines, methaqualone, propoxphene, barbiturates, and amphetamines. Marijuana and cocaine tests were performed using EMIT^{TM} (Enzyme Multiplied Immunoassay). ''For most drugs, urinalysis can detect use within the previous 2–3 days; use of marijuana and PCP can sometimes be detected several weeks after use'' (U.S. Department of Justice, 1996). Using EMIT^{TM}, rates of false positives are quite low (2.1% and 2.5% for marijuana and cocaine, respectively) but rates of false negatives are higher (29.0% and 22.8% for marijuana and cocaine, respectively; see Harrison 1995).[1] There is no reason to believe, however, that rates of false positives and false negatives vary across social groups.

Disadvantages of DUF Data

The primary disadvantage to using the DUF data is that interview procedures are not completely standardized across sites. These differences across sites (e.g., being interviewed in front of a detention guard versus being interviewed in a closed area away from all criminal justice personnel) may bias response rates and the willingness of arrestees to answer honestly. Because sample sizes per site are rather low, we are forced to use data from multiple sites. Due to these low sample sizes, we are unfortunately unable to fully determine whether significant differences across sites exist. The statistical power of our analyses is too low to fully examine site differences. More simplistic analyses are required in order to examine site differences.

Previous research, however, has generally not reported differences across sites or procedures. Wish et al. (2000), for example, utilized an experimental design to determine whether the type of informed consent (standard versus enhanced) and the sequence of drug testing (interview versus urine specimen first) affected the validity of self-reported drug use. Results clearly indicated that the validity of self-reported drug use was not affected by these procedural differences. In addition, Rosenfeld and Decker (1993) examined the consistency of underreporting across time and space. Their results indicated that the magnitude of the difference between self-reports and urine tests is consistent across both time and space. On the other hand, Yacoubian (2001) concluded that urinalysis and self-report agreements are less stable across jurisdictions than across time. McElrath et al. (1995) further noted that the correlates of inaccurate self-reports and of underreporting do vary across sites. When such differences are uncovered, interpretational confounding is likely to occur. It is very difficult to explain such differences because very little documentation on site-specific protocols is available.

Sample

The sample consists of the 1994 data for White and Black adults from Indianapolis, Ft. Lauderdale, Phoenix, and Dallas. These four sites were purposefully chosen

[1] Cutoff levels for marijuana testing were changed in 1996 from 100 ng/ml to 50 ng/ml, thus increasing the percentage of positive tests (and decreasing the percentage of negative tests) by 5–7% points (U.S. Department of Justice, 1996). In the end, we must recognize that urine testing is not a ''gold'' criterion for self-reported drug use. Recent advances in Bayesian statistics allow us examine the validity of self-reports in the absence of a true ''gold'' criterion (see Joseph et al. 1995). Future research should capitalize on these advances to examine the validity of self-reports when the truth is never known with certainty.

because each contained over 500 respondents and contained at least 20 respondents per cell in two-by-two tables of marijuana self-report versus marijuana test and of crack/cocaine self-report versus crack/cocaine test. The minimum requirement of 20 respondents per cell is important to ensure adequate power for our analytic methods. Of the 4,899 White and Black adults from these four sites, 147 (3%) were eliminated due to missing data on the variables used in this analysis.

Because differences in the validity of self-reported drug use across drug categories (i.e., marijuana and crack/cocaine) were of interest, a sampling technique was used to create independent observations on the validity of self-reported marijuana use and of crack/cocaine use. By creating statistically independent observations, we gain the ability to calculate the statistical significance of differences in the validity of self-reported drug use across drug categories. In order to create independent observations, cases were randomly assigned to contribute information either on marijuana use or on crack/cocaine use. To not alter the proportions of positive and negative self-reports and drug tests of marijuana and crack/cocaine use, a stratified randomization procedure was used. The adequacy of this procedure was checked to ensure that the distributions of gender, race, age, and offense seriousness within drug test categories were not significantly altered from the original data. Data were archived with ICPSR (Study No. 2706).

Measures

The exogenous measures included in this study consist of type of drug (coded 0 for marijuana and 1 for crack/cocaine), age (coded 0 for 18 through 30, and 1 for 31 or over), offense seriousness (coded 0 for misdemeanor and 1 for felony), race (coded 0 for Black and 1 for White), and gender (coded 0 for male and 1 for female). The endogenous measures included in this study consist of validity (coded 1 if the self-report and the drug test were both positive or negative and 0 otherwise), underreporting (coded 1 if the self-report was negative when the drug test was positive and 0 otherwise), and overreporting (coded 1 if the self-report was positive when the drug test was negative and 0 otherwise). Self-reports were obtained by asking respondents to indicate their use of marijuana, crack, and cocaine within the previous 3 days. The drug tests can generally detect the use of these drugs for 2–3 days. Marijuana use can generally be detected longer than crack/cocaine use. It would therefore not be entirely surprising if individuals were less likely to have accurate self-reports of marijuana use than of crack/cocaine use, were more likely to underreport marijuana use than crack/cocaine use, and were less likely to overreport marijuana use than crack/cocaine use.

Descriptive statistics are shown in Table 1. Overall, 22.4% of self-reports were invalid. The decomposition of the probability of an invalid self-report is shown in Fig. 1b. Unconditional probabilities show that 22.4% of self-reports were invalid because 16.0% underreported drug use and 6.4% overreported drug use (16.0% + 6.4% = 22.4%). The unconditional probability of underreporting is decomposed into a conditional probability of 45.8% and a probability of a positive test of 34.8% (45.8% * 34.8% = 16.0%). The conditional probability of underreporting is therefore 186% higher than the unconditional probability. The unconditional probability of overreporting is decomposed into a conditional probability of 9.9% and a probability of a negative test of 65.2% (9.9% * 65.2% = 6.4%). The conditional probability of overreporting is therefore 55% higher than the

Table 1 Descriptive statistics for endogenous and exogenous measures

Measure	Number (percent)
Drug	
Marijuana	2,369 (49.9)
Crack/cocaine	2,383 (50.1)
Age	
18–30	2,648 (55.7)
31 or over	2,104 (44.3)
Gender	
Male	3,238 (68.1)
Female	1,514 (31.9)
Race	
Black	2,428 (51.1)
White	2,324 (48.9)
Offense	
Misdemeanor	1,787 (37.6)
Felony	2,965 (62.4)
Validity of self-report	
Valid	3,687 (77.6)
Invalid	1,065 (22.4)

unconditional probability. Again, how we operationally define underreporting and overreporting matters a great deal. This will be particularly true when examining differences in the probabilities of underreporting and overreporting across groups, as groups likely differ in base rates.

Procedures

The first analyses examine differences in the validity of self-reported drug use across gender, race, age, type of drug, and offense seriousness. These differences are examined with hierarchical loglinear, logit, and logistic regression models. These differences are then explained by examining differences in the underreporting and overreporting of drug use across gender, race, age, type of drug, and offense seriousness. These differences are again examined with hierarchical loglinear, logit, and logistic regression models. Finally, we re-examine differences in the underreporting and overreporting of drug use while controlling for differences in base rates using logistic regression models. Final logistic regression models are estimated on the full sample, on the sub-sample with positive drug tests, and on the sub-sample with negative drug tests. Using the full sample does not control for base rates. These models estimate unconditional probabilities of underreporting and overreporting. Using the sub-sample with positive drug tests allows us to examine the conditional probability of underreporting while controlling for differences in base rates. Conversely, using the sub-sample with negative drug tests allows us to examine the conditional probabilities of overreporting while controlling for differences in base rates. The following sections describe in more detail the use of hierarchical loglinear, logit, and logistic regression models.

Hierarchical Loglinear Models

The data represent a 2^6 contingency table (i.e., endogenous measure by five exogenous measures). Hierarchical loglinear models and logit models are used to reduce,

or collapse, this contingency table to include only significant main effects and interactions. In the hierarchical loglinear models, the dependent variable is the count in each cell of the 2^6 contingency table. As a result, all possible interactions are considered, including those without the endogenous measure (e.g., type of drug by age by race). Interactions without the endogenous measure are eliminated in the logit analyses described in the next section.

Hierarchical loglinear models are primarily useful to determine the significance of higher-order interactions. Unsaturated models (i.e., ones which do not contain all main effects or interactions) are systematically compared to a saturated model to determine whether variables interact as well as the level of their interactions (Fienberg 1980). For each model, a Chi-Square statistic can be computed to indicate the degree to which the predicted cell counts approach the observed ones. If this Chi-Square statistic is not significant, one can conclude that the model provides a good fit to the data (i.e., the predicted cell counts are not significantly different than the observed ones). More interestingly, models can be compared to determine if the six-, five-, four-, three-, and two-way interactions, and the main effects are significant.

Models are compared using differences in Chi-Square statistics to determine whether the six-way interaction is significant, all five-way, all four-way, all three-way, all two-way, and all main effects are significant. Furthermore, models are compared to determine whether all six- and five-way interactions are jointly significant, whether all six-, five-, and four-way interactions are jointly significant, whether all six-, five-, four-, and three-way interactions are jointly significant, whether all interactions are jointly significant, and whether all interactions and main effects are jointly significant.

Logit Models

In logit models, the dependent variable is the endogenous measure. Therefore, logit models inherently consider only main effects and interactions, which are related to the endogenous measure. All main effects and interactions, which do not involve the endogenous measure are instantly dropped from the model. Whether these main effects and interactions are significant is of no interest. A backward elimination procedure was used to eliminate the remaining non-significant interaction terms and main effects. The backward elimination procedure starts with the model suggested by the hierarchical loglinear analysis and systematically eliminates the least significant interaction terms and main effects until all interaction terms or main effects included in the model are significant. At each step of the backward elimination procedure, the least significant main effect and all interaction terms involving this main effect were eliminated. Main effects and interaction terms were eliminated only if the resulting increase in the Chi-Square statistic was non-significant (i.e., if the difference between observed and expected cell counts did not significantly increase). The accuracy of all backward elimination procedures was checked with forward selection procedures. Identical results were always obtained.

Logistic Regression Models

For ease of interpretation and presentation, the final logit models are converted to logistic regression models. In these models, the slopes represent the expected effect of the independent variables on the log-odds of the dependent variable. Predicted probabilities can also be computed.

Results

Models of Validity

The results from the hierarchical loglinear model for validity are presented in Table 2. This table shows the 11 comparisons mentioned in section ''Hierarchial loglinear models''. More precisely, the first row presents the significance of the six-way interaction. The second row presents the significance of all five-way interactions followed by the joint significance of all five- and six-way interactions. The third row presents the significance of all four-way interactions followed by the joint significance of all four-, five-, and six-way interactions. The fourth row presents the significance of all three-way interactions followed by the joint significance of all three-, four-, five-, and six-way interactions. The fifth row presents the significance of all two-way interactions followed by the joint significance of all interactions. Finally, the last row presents the significance of all main effects followed by the joint significance of all main effects and interactions.

Results show that all six-, five-, four-, and three-way interactions are not statistically significant. Removing all six-, five-, four-, and three-way interactions would not significantly reduce the fit provided to the data ($p = 0.527$). However, at least one of the two-way interactions is significant ($p < 0.001$). Eliminating all two-way interactions would significantly reduce the fit provided to the data. In addition, eliminating all interactions would significantly reduce the fit provided to the data as well ($p < 0.001$). The final model therefore contains all main effects and two-way interactions. This model predicts that validity is a function of type of drug, race, offense seriousness, age, and gender.

Logit analyses (also shown in Table 2) were performed to eliminate specific non-significant effects. For each model in Table 2, the likelihood ratio Chi-Square statistic is reported along with its degrees of freedom and significance. Of more importance in the backward elimination procedure, the differences in Chi-Square statistics between the first model and subsequent models are also reported. These differences in Chi-Square statistics are used to show that the fit provided to the data

Table 2 Significance of parameters in loglinear and logit models for accuracy

Loglinear parameters	Likelihood ratio Chi-Square	df	*p*-value	Sum in Chi-Square	df	*p*-value
Six-way interaction	0.039	1	0.842			
Five-way interactions	8.043	6	0.235	8.082	7	0.325
Four-way interactions	6.640	15	0.967	14.722	22	0.874
Three-way interactions	25.197	20	0.194	39.919	42	0.527
Two-way interactions	237.045	15	<0.001	276.964	57	<0.001
Main Effects	2530.667	6	<0.001	2807.631	63	<0.001
Logit model[a]	Likelihood ratio Chi-Square	df	*p*-value	Difference in Chi-Square	df	*p*-value
[D] [R] [O] [A] [G]	31.53	26	0.209			
[D] [R] [O] [A]	31.71	27	0.243	0.18	1	0.671
[D] [R] [O]	31.89	28	0.279	0.36	2	0.835
[D] [R]	32.53	29	0.297	1.00	3	0.801
[R]	33.51	30	0.301	1.98	4	0.739

[a] D, Drug; R, race; O, offense; A, age; G, gender

is never significantly worse than the fit provided to the data by the first model (suggested from the hierarchical loglinear models).

The main effect of gender was removed first because doing so produced the smallest increase in the Chi-Square statistic. In addition, the increase in the Chi-Square statistic was not significant ($p = 0.671$). Following this logic, the main effects of age, offense seriousness, and type of drug were subsequently removed. Removing these terms did not reduce the fit provided to the data ($p = 0.835$, 0.801, and 0.739 for age, offense seriousness, and type of drug, respectively). No further terms could be removed. Removing the main effect of race would have significantly reduced the fit provided to the data (comparison not shown, $p < 0.001$). The final model therefore shows that validity is solely a function of race.

The results from the logistic regression models (not shown) indicate that the log-odds of a self-report being valid are significantly higher for Whites than for Blacks. More specifically, the predicted probability of a valid self-report is 0.74 for Whites and 0.66 for Blacks. This small, but significant, difference may emerge due to differences in underreporting and overreporting. The following sections examine such differences.

Models of Underreporting

Results shown in Table 3 reveal that all six-, five-, and four-way interactions are not significant ($p = 0.843$). While results show that eliminating all three-, four-, five-, and six-way interactions would not significantly reduce the fit provided to the data ($p = 0.162$), results also show that at least one of the three-way interactions is significant ($p = 0.018$). Given the conflicting results about the significance of the three-way interactions, we chose to be conservative and hypothesized that at least one of the three-way interactions was significant. The final model therefore contains all main effects and all two- and three-way interactions. This model predicts that underreporting is a function of type of drug, race, offense seriousness, age, and gender, and of all two-way interactions between these five factors.

Table 3 Significance of parameters in loglinear and logit models for underreporting

Log linear parameters	Likelihood ratio Chi-Square	df	*p*-value	Sum in Chi-Square	df	*p*-value
Six-way interaction	0.09	1	0.764			
Five-way interactions	9.31	6	0.157	9.40	7	0.225
Four-way interactions	6.04	15	0.979	15.44	22	0.843
Three-way interactions	35.49	20	0.018	50.93	42	0.162
Two-way interactions	283.89	15	<0.001	334.82	57	<0.001
Main effects	3413.20	6	<0.001	3748.02	63	<0.001
Logit model[a]	Likelihood ratio Chi-Square	df	*p*-value	Difference in Chi-Square	df	*p*-value
[DR] [DO] [DA] [DG] [RO] [RA] [RG] [OA] [OG] [AG]	12.10	16	0.737			
[DR] [DA] [DG] [RA] [RG] [AG]	13.57	21	0.887	1.47	5	0.916
[DR] [DG] [RG]	19.74	25	0.760	7.64	9	0.571
[DR]	25.26	28	0.614	13.16	12	0.357

[a] D, Drug; R, race; O, offense; A, age; G, gender. All models contain lower interaction terms and main effects (i.e., [DR] contains drug by race interaction and main effects of drug and race)

The results from the logit models are also presented in Table 3. The main effect of offense seriousness and all interactions involving offense seriousness were removed first. All terms involving offense seriousness were removed because doing so produced the smallest increase in the Chi-Square statistic. In addition, the increase in the Chi-Square statistic was not significant ($p = 0.916$). For the same reasons, the main effect of age and all interactions involving age were then removed. Finally, the main effect of gender and all interactions involving gender were removed. Once again, removing these terms did not significantly reduce the fit provided to the data ($p = 0.571$ and 0.357 for age and gender, respectively). No further terms could be removed. Removing the interaction between type of drug and race would have significantly reduced the fit provided to the data (comparison not shown, $p = 0.004$). The final model shows that underreporting is a function of type of drug, race, and of the type of drug by race interaction.

The results from the logistic regression models are shown in Table 4. Two logistic regression models are shown in Table 4. The first is estimated on the full sample that includes both positive and negative tests. In this model, differences in base rates are not controlled for. As a result, this model examines the unconditional probability of underreporting. The second is estimated on the sub-sample that tested positive. In this model, there are no differences in base rates (all tested positive). As a result, this model examines the conditional probability of underreporting. Full sample results indicate that the log-odds of underreporting are significantly higher for reports of crack/cocaine use than of marijuana use. The effect of race is non-significant, but the log-odds of underreporting are significantly higher for reports of crack/cocaine use from Blacks than from Whites. The log-odds of underreporting are also significantly higher for reports of crack/cocaine use from Blacks than for reports of marijuana use from both Blacks and Whites. The predicted probabilities of underreporting marijuana use from Whites and Blacks, and of underreporting crack/cocaine use from Whites and Blacks are 0.12, 0.12, 0.15, and 0.25, respectively.

When controlling for differences in base rates (in the sub-sample that tested positive), results reveal that the interaction between race and type of drug becomes non-significant. More specifically, when differences in base rates between Black and White offenders are controlled, Black offenders do not underreport to a greater extent. Black offenders underreport crack/cocaine use to a greater extent than White offenders because, and solely because, their base rate of crack/cocaine use is higher. The main effect of type of drug is still statistically significant. Offenders are more likely to underreport crack/cocaine use than marijuana use. This difference is not attributable to differences in base rates.

Table 4 Logistic regression model for underreporting

Parameter	Full sample β (s.e.)	Sample with positive test β (s.e.)
Constant	–2.020 (0.089)[c]	–0.424 (0.107)[c]
Drug[a]	0.920 (0.111)[c]	0.372 (0.134)[c]
Race[b]	0.047 (0.127)	0.182 (0.156)
Race by drug	–0.686 (0.165)[c]	–0.174 (0.205)
Model χ^2 (df)	95.527 (3)[c]	9.647 (3)[d]
–2 Log likelihood	4174.24	2274.543

[a] Drug: 0 = Marijuana, 1 = Cocaine

[b] Race: 0 = Black, 1 = White

[c] $p < 0.01$

[d] $p = 0.02$

Models of Overreporting

Results shown in Table 5 reveal that all six-, five-, four-, and three-way interactions are not statistically significant. Removing all these interactions would not significantly reduce the fit provided to the data ($p = 0.794$). However, at least one of the two-way interactions is significant ($p < 0.001$). The final model therefore contains all main effects and two-way interactions. This model predicts that underreporting is a function of type of drug, race, offense seriousness, age, and gender.

The results from the logit models are also presented in Table 5. The main effect of gender was removed first because doing so produced the smallest increase in the Chi-Square statistic. In addition, the increase in the Chi-Square statistic was not significant ($p = .417$). The main effects of age and offense seriousness were subsequently removed. Removing these main effects did not significantly reduce the fit provided to the data ($p = 0.415$ and 0.399 for age and offense seriousness, respectively). Removing either the main effect of type of drug or of race would have significantly reduced the fit provided to the data (comparisons not shown, $p < 0.001$). The final model shows that overreporting is a function of type of drug and race.

The results from the logistic regression models (shown in Table 6) indicate that the log-odds of overreporting are significantly higher for reports of marijuana use than of crack/cocaine use. In addition, the log-odds of overreporting are significantly higher for Blacks than for Whites. The predicted probabilities of overreporting marijuana use for Whites and Blacks, and of overreporting crack/cocaine use for Whites and Blacks are 0.08, 0.11, 0.02, and 0.03, respectively. Overall, offenders are more likely to overreport marijuana use than crack/cocaine use, and Black offenders are more likely to overreport the use of marijuana and crack/cocaine than White offenders.

These differences may again be due to differences in conditional probabilities or to differences in base rates. The logistic regression model of overreporting was also evaluated in the sub-sample of offenders with negative drug tests. Results (also shown in Table 6) reveal that all effects remain statistically significant even when differences in base rates are controlled for. Offenders are more likely to overreport marijuana use than crack/cocaine use and Black offenders are more likely to

Table 5 Significance of parameters in loglinear and logit models for overreporting

Loglinear parameters	Likelihood ratio Chi-Square	df	*p*-value	Sum in Chi-Square	df	*p*-value
Six-way interaction	1.007	1	0.316			
Five-way interactions	5.176	6	0.521	6.183	7	0.518
Four-way interactions	10.260	15	0.803	16.443	22	0.793
Three-way interactions	17.890	20	0.595	34.333	42	0.794
Two-way interactions	314.782	15	<0.001	349.115	57	<0.001
Main effects	5317.034	6	<0.001	5666.149	63	<0.001
Logit model[a]	Likelihood ratio Chi-Square	df	*p*-value	Difference in Chi-Square	df	*p*-value
[D] [R] [O] [A] [G]	25.92	26	0.467			
[D] [R] [O] [A]	26.58	27	0.487	0.66	1	0.417
[D] [R] [O]	27.68	28	0.482	1.76	2	0.415
[D] [R]	28.87	29	0.472	2.95	3	0.399

D, Drug; R, race; O, offense; A, age; G, gender

Table 6 Logistic regression model for overreporting

Parameter	Full sample β (s.e.)	Sample with negative test β (s.e.)
Constant	–2.068 (0.085)[c]	–1.627 (0.087)[c]
Drug[a]	–1.273 (0.139)[c]	–1.074 (0.141)[c]
Race[b]	–0.307 (0.121)[c]	–0.436 (0.124)[c]
Model χ^2 (df)	106.272 (2)[c]	84.576 (2)[c]
–2 Log likelihood	2164.141	1912.471

[a] Drug: 0 = Marijuana, 1 = Cocaine

[b] Race: 0 = Black, 1 = White

[c] $p < 0.01$

overreport drug use than White offenders, even when controlling for differences in base rates.

Summary and Conclusions

The logistic regression model for validity revealed that validity was a function of race. Black offenders provided less valid self-reports than White offenders. This difference was explained by differences in underreporting and overreporting. We found that Black offenders were more likely to underreport crack/cocaine use than White offenders. This race difference, however, disappeared once differences in base rates were controlled for. Black offenders were more likely to underreport crack/cocaine use simply because a higher proportion of Black offenders (51.2%) tested positive for crack/cocaine use than White offenders (30.5%). Black offenders were also more likely to overreport both marijuana and crack/cocaine use relative to White offenders. These differences were not attributable to differences in base rates. When controlling for differences in base rates, Black offenders were still more likely to overreport both marijuana and crack/cocaine use relative to White offenders.

We should also note that while accuracy was not a function of type of drug, both underreporting and overreporting were. More specifically, offenders were more likely to underreport crack/cocaine use and were more likely to overreport marijuana use. This is striking given that the window of detection is longer for marijuana use than for crack/cocaine use. The underreporting and overreporting effects canceled each other out in the validity analyses. Because offenders were more likely to underreport and overreport different types of drugs, the validity of self-reported drug use was not affected by type of drug. Underreporting and overreporting differences across types of drug could be explained by differences in base rates.

The results also indicated that gender, offense seriousness, age, and type of drug do not affect the validity of self-reported drug use. These results strongly support the further use of self-report data to examine patterns of drug use and for research and policy development purposes. Nevertheless, there are four important limitations. First, while type of drug does not have an effect on the validity of self-reported drug use, offenders are more likely to underreport crack/cocaine use than marijuana use and are more likely to overreport marijuana use than crack/cocaine use. Second, Black offenders provide significantly less valid reports of drug use than White offenders. Third, Black offenders have higher rates of crack/cocaine use than White offenders, and thus underreport crack/cocaine use to a greater extent. Finally, Black offenders are more likely to overreport both marijuana and crack/cocaine use than White offenders.

The disappearance of the race effect on underreporting when controlling for differences in base rates does not mean that self-reports of crack/cocaine use are equally valid across racial groups. The fact that the race effect disappears when differences in base rates are controlled for does not mean that valid inferences can be reached when comparing self-reports of crack/cocaine use across racial groups. It does not mean either that valid rates of crack/cocaine use can be calculated without adjustments. It simply explains why race has an effect on underreporting. Black offenders are more likely to underreport crack/cocaine use than White offenders because Black offenders are more likely to test positive for crack/cocaine use. Among offenders who test positive for crack/cocaine use, race does not affect the likelihood of underreporting. The effect of race on underreporting will increase as the differences in base rates increase. To make valid inferences from self-reports of crack/cocaine use across racial groups, we must choose racial groups with similar rates of positive drug tests. To compute valid rates of crack/cocaine use, we should take into account racial differences in positive drug tests. However, while race will not affect the likelihood of underreporting in samples with similar base rates, race will still affect the likelihood of overreporting, even in samples with similar base rates. Black offenders are more likely to overreport both marijuana and crack/cocaine use than White offenders. This difference is not attributable to a difference in base rates.

In addition, the effects of type of drug on underreporting and overreporting could not simply be explained by differences in base rates either. Offenders are more likely to underreport crack/cocaine use than marijuana use. In addition, offenders are more likely to overreport marijuana use than crack/cocaine use. The analyses presented here clearly showed that some true differences in the validity, underreporting and overreporting of drug use exist. Additional work is required to explain these differences. Nevertheless, the analyses presented here also clearly showed that differences in the validity, underreporting, and overreporting of drug use are relatively rare. Some of these rare differences can simply be attributed to differences in base rates. No differences across gender, age, or offense seriousness were found. Even though we actively searched for higher-order interactions, our final models were remarkably simple. This undoubtedly supports the further, though cautious, use of self-reports.

Acknowledgments This project is supported by Grant No. 97-IJ-CX-0051 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. We wish to thank Eric Wish and anonymous reviewers for their guidance.

References

Andrews DA, Zinger I, Hoge RD, Bonta J, Gendreau P, Cullen FT (1990) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. Criminology 28:369–404

Bale RN, Van Stone WW, Engelsing TM, Zarcone VP (1981) The validity of self-reported heroin use. Int J Addict 16:1387–1398

Falck R, Siegal HA, Forney MA, Wang J, Carlson RG (1992) The validity of injection drug users self-reported use of opiates and cocaine. J Drug Issues 22:823–832

Fendrich M, Xu Y (1994) The validity of drug use reports from juvenile arrestees: a comparison of self-report, urinalysis and hair assay. J Drug Issues 24:99–116
Fienberg SE (1980) The analysis of cross-classified categorical data. MIT, Cambridge
Golub A, Johnson BD, Taylor A, Liberty, HJ (2002) The validity of arrestees' self-reports: variations across questions and persons. Justice Q 19:477–502
Gray TA, Wish ED (1999) Correlates of underreporting recent drug use by female arrestees. J Drug Issues 29:91–106
Harrell AV (1985) Validation of self-report: the research record. In: Rouse B, Kozel N, Richards L (eds) Self-report methods of estimating drug use. NIDA, Rockville
Harrison LD (1995) The validity of self-reported data on drug use. J Drug Issues 25:91–111
Hser Y, Maglione M, Boyle K (1999) Validity of self-report of drug use among STD patients, ER patients, and arrestees. Am J Drug Alc Abuse 25:81–91
Joseph L, Gyorkos TW, Coupal L (1995) Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am J Epidemiol 141:263–272
Katz CM, Webb VJ, Gartin PR, Marshall CE (1997) The validity of self-reported marijuana and cocaine use. J Crim Jus 25:31–41
Kim JYS, Fendrich M, Wislar JS (2000) The validity of juvenile arrestees' drug use reporting: a gender comparison. J Res Crime Delinq 37:419–432
Lu NT, Taylor BG, Riley KJ (2001) The validity of adult arrestee self-reports of crack cocaine use. Am J Drug Alc Abuse 27:399–419
Magura S, Kang SY (1995) Validity of self-reported drug use in high risk populations: a meta-analytic review. National Development and Research Institutes, New York
Magura S, Goldsmith D, Casriel C, Goldstein PJ, Lipton DS (1987) The validity of methadone clients' self reported drug use. Int J Addict 22:727–749
McElrath K, Dunham R, Cromwell P (1995) Validity of self-reported cocaine and opiate use among arrestees in five cities. J Crim Jus 23:531–540
McNagny SE, Parker RM (1992) High prevalence of recent cocaine use and the unreliability of patient self-report in an inner-city walk-in clinic. JAMA 267:1106–1108
Messina NP, Wish ED, Nemes S, Wraight B (2000) Correlates of underreporting of post-discharge cocaine use among therapeutic community clients. J Drug Issues 30:119–132
Mieczkowski T (1990) The accuracy of self-reported drug use: an evaluation and analysis of new data. In: Weisheit R (ed) Drugs, crime and the criminal justice system. Anderson, Cincinnati
Nelson DB, Kotranski L, Semaan S, Collier K, Lauby J, Feighan K, Halbert J (1998) The validity of self-reported opiate and cocaine use by out-of-treatment drug users. J Drug Issues 28:483–494
Page WF, Davies JE, Ladner RA, Alfassa J, Tennis H (1977) Urinalysis screened vs. verbally reported drug use: the identification of discrepant groups. Int J Addic 12:439–450
Richter L, Johnson PB (2001) Current methods of assessing substance use: a review of strengths, problems, and developments. J Drug Issues 46:34–42
Rosenfeld R, Decker S (1993) Discrepant values, correlated measures: cross-city and longitudinal comparisons of self-reports and urine tests of cocaine use among arrestees. J Crim Jus 21:223–230
Sloan JJ, Bodapati MR, Tucker TA (2004) Respondent misreporting of drug use in self-reports: social desirability and other correlates. J Drug Issues 34:269–292
Stephens RC, Feucht TE (1993) Reliability of self-reported drug use and urinalysis in the dug use forecasting system. Prison J 73:279–289
Wish ED, Gray T, Sushinsky J, Yacoubian GS Jr, Fitzgerald N (2000) An experiment to enhance the reporting of drug use by arrestees. J Drug Issues 30:55–76
Wish ED, Hoffman JA, Nemes S (1997) The validity of self-reports of drug use at treatment admission and at followup: comparisons with urinalysis and hair assays. In: Harrison L, Hughes A (eds) The validity of self-reported drug use: improving the accuracy of survey estimates. NIDA, Rockville
Yacoubian GS (2001) Exploring the temporal validity of self-reported marijuana use among juvenile arrestees. J Alc Drug Educ 46:34–42
U.S. Department of Justice, National Institute of Justice (1996) Drug Use Forecasting: 1996 Annual Report. GPO, Washington