182363

Differences in the Validity of Self-Reported Drug Use Across Five Factors:

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

André B. Rosay
University of Delaware

Stacy Skroban Najaka
University of Maryland at College Park

Denise C. Herz
University of Nebraska - Omaha

Final Report

Grant No. 97-IJ-CX-0051

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 for the use of the CESAR library.

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

TABLE OF CONTENTS

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

EXECUTIVE SUMMARY

Practitioners, researchers, and policy makers all rely extensively on measures of self-reported drug use. Self-reported measures of drug use are utilized to determine which drug prevention and rehabilitation services should be offered to whom, which services are successful, and finally, which services should be expanded and continually funded. A well known-problem with self-reports is the uncertainty about their ability to accurately indicate what is being measured. In particular, the extent to which self-reported drug use is an equally valid indicator of actual drug use across groups has been repeatedly questioned. However, little work has examined whether the validity of self-reported drug use varies across social groupings. Patterns across studies suggest that validity differences in self-reported drug use do exist. However, these differences cannot be statistically evaluated because each study utilized different types of high-risk populations and measurement procedures.

This study expands our knowledge about the accuracy of self-reported drug use in three directions. First, this study examines differences in the accuracy of self-reported drug use across gender, race, age, type of drug, and offense seriousness. All differences in the accuracy of self-reported drug use across these five factors and across interactions between these five factors are evaluated. Second, this study explains differences in the accuracy of self-reported drug use in terms of differences in underreporting and overreporting. Inaccurate self-reports can emerge due to underreporting and overreporting. The specific sources of inaccurate self-reports are determined. Third, this study explains differences in underreporting and overreporting in terms

of true differences or differences in opportunity. Individuals can underreport drug use only if they test positive for drug use. Similarly, individuals can overreport drug use only if they test negative for drug use. In order to uncover true differences in underreporting and overreporting, we must control for differences in the opportunity to underreport and overreport.

This study uses data collected in 1994 as part of the Drug Use Forecasting (DUF) Program. The DUF program interviews arrestees on lifestyles and drug use, and collects urine specimens. This allows one to check the accuracy of self-reported drug use with a biological criterion, namely, urine tests. The sample used consists of the 1994 data for White and Black adults from Indianapolis, Ft. Lauderdale, Phoenix, and Dallas. The exogenous measures included in this study consist of type of drug (marijuana and crack/cocaine), age group (18 through 30 and 31 or over), offense seriousness (misdemeanor and felony), race (Black and White), and gender (male and female). The endogenous measures included in this study consist of accuracy (whether the self-report and the drug test were both positive or negative), underreporting (whether the self-report was negative when the drug test was positive), and overreporting (whether the self-report was positive when the drug test was negative).

The first analyses examine differences in the accuracy 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, the sources of differences in the underreporting and overreporting of drug use are examined with 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. In the sub-sample with positive drug tests, all arrestees have an equal opportunity to underreport. In the sub-sample with negative drug tests, all arrestees have an equal opportunity to overreport. When using these sub-samples, differential opportunities are controlled for. As a result, any remaining difference across groups is attributable to a true difference. This allows us to separate differences in underreporting and overreporting into true differences and differences in opportunity.

Results showed that accuracy was a function of race. Black offenders provided less accurate self-reports than White offenders. This difference was explained by differences in underreporting and overreporting. The logistic regression results showed that Black offenders were more likely to underreport crack/cocaine use than White offenders, but that Black offenders were not more likely to underreport marijuana use than White offenders. The race difference in underreporting existed only for crack/cocaine use. In addition, this difference disappeared once opportunity was controlled for. This race difference was solely due to differences in opportunity. Blacks were more likely to underreport crack/cocaine use than Whites simply because a higher proportion of Black offenders tested positive for crack/cocaine use than White offenders. Black offenders were also more likely to overreport both marijuana and crack/cocaine use relative to White offenders. This difference was attributable to a true difference. When controlling for opportunity, Black offenders were still more likely to overreport both marijuana and crack/cocaine use relative to White offenders.

The analyses presented here clearly showed that some true differences in the accuracy, underreporting, and overreporting of drug use exist. Theoretical frameworks should be developed to explain these true differences. Nevertheless, the analyses presented here also clearly showed that differences in the accuracy, underreporting, and overreporting of drug use are

relatively rare. Some of these rare differences can simply be attributed to differences in opportunity. 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.

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

1. INTRODUCTION

The majority of studies examining drug use have relied on self-reported measures of drug use (Tims and Ludford, 1984; Young, 1994; Magura and Kang, 1995). The results from these studies have been used to determine how to plan and allocate drug prevention and rehabilitation services (Fendrich and Xu, 1994) and to determine the effectiveness of these drug prevention and rehabilitation services (Falck, Siegal, Forney, Wang and Carlson, 1992). These results have influenced policy decisions such as which drug prevention and rehabilitation programs should be funded and expanded and which ones should not. In addition, individual self-reports are used every day in our justice system to determine which drug prevention and rehabilitation services should be offered to whom (Magura, Goldsmith, Casriel, Goldstein and Lipton, 1987; Andrews, Zinger, Hoge, Bonta, Gendreau and Cullen, 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 is the uncertainty about their ability to accurately indicate what is being measured. Many investigations have shown that the validity of self-reported data is questionable, especially when the topic is as sensitive as drug use (Harrell, 1985). 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 are therefore unlikely to disclose such information (Maddux and Desmond, 1975; Bale, Van Stone, Engelsing and Zarcone, 1981; Wish, Johnson,

Strug, Chedekel and Lipton, 1983; Falck et al., 1992). Because drug use reported by individuals in the justice system is used for policy development and evaluation, it is important to determine the extent to which drug use reported by these individuals is a valid indicator of actual drug use.

Many investigations have examined this issue. 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 a biological criterion such as urinalysis or hair analysis. Magura and Kang (1995) noted that the validity of self-reported drug use varied greatly across studies[1]. They 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.

Reviewing the literature since 1985 where drug use measures anonymously reported by high risk populations were compared to urinalysis or hair analysis results, there indeed appears to be differences in the validity of self-reported drug use across types of high risk populations and across types of drug use. First, high risk females are less likely than high risk males to underreport drug use. Although no study has compared drug use reported by high risk populations to urinalysis or hair analysis results across gender groups, studies using samples consisting of high risk females (Marques, Tippetts and Branch, 1993; Funkhouser, Butz, Feng, McCaul and Rosenstein, 1994; Gray, 1996) reported that self-reported drug use was a more valid indicator of actual drug use than studies using samples consisting of high risk males (Anglin, Hser and Chou, 1993; Magura, Kang, and Shapiro, 1995). Second, high risk minorities are less likely than high risk non-minorities to underreport drug use (Fendrich and Xu, 1994; Gray,

---

[1] Magura and Kang (1995) reported that, on average, "positive self-reports were given by 42% of those subjects who had a positive urinalysis or hair analysis."

1996)[2]. Third, high risk populations involved in serious criminality are less likely to accurately report drug use than high risk populations involved in less serious criminality (Magura et al., 1987)[3]. Finally, some have hypothesized that high risk populations are less likely to underreport marijuana use because it is more socially acceptable than other drug use. But while Dembo, Williams, Wish and Schmeidler (1990), Fendrich and Xu (1994), and Katz, Webb, Gartin and Marshall (1997) reported that high-risk populations were less likely to inaccurately report marijuana use than other drug use, Brown, Kranzler and Del Boca (1992) and Harrison (1995) reported that the validity of self-reported drug use was not a function of the type of drug use.

While the comparisons across these few studies are indicative of validity differences across types of high risk populations and across types of drug use, patterns cannot be statistically evaluated because of 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, Hoffman and Nemes, 1997). For example, different studies have operationalized validity in different ways. Some operationalized validity as the accuracy of self-reported drug use while others operationalized validity as the underreporting of drug use. As noted by Magura and Kang (1995), these differences do not allow us to make valid comparisons across studies.

This study expands our knowledge about the accuracy of self-reported drug use in three directions. First, this study examines differences in the accuracy of self-reported drug use across gender, race, age, type of drug, and offense seriousness. All differences in the accuracy of self-

---

[2] See also Page, Davies, Ladner, Alfassa and Tennis (1977).

[3] See also Eckerman, Bates, Rachal and Poole (1971); Page et al. (1977). On the other hand, McGlothlin, Anglin and Wilson (1977) reported that the validity of drug use reported by high risk populations is independent of the legal status of these high risk populations.

reported drug use across these five factors and across interactions between these five factors are evaluated. Second, this study explains differences in the accuracy of self-reported drug use in terms of differences in underreporting and overreporting. Inaccurate self-reports can emerge due to underreporting and overreporting. The specific sources of inaccurate self-reports are determined. Third, this study explains differences in underreporting and overreporting in terms of true differences or differences in opportunity. Individuals can underreport drug use only if they test positive for drug use. Similarly, individuals can overreport drug use only if they test negative for drug use. In order to uncover true differences in underreporting and overreporting, we must control for differences in the opportunity to underreport and overreport.

We must do so because the likelihood of underreporting is, in part, a function of the likelihood of testing positive. Similarly, the likelihood of overreporting is, in part, a function of the likelihood of testing negative. Because only individuals who test positive can underreport drug use, a group with a high rate of positive tests will underreport drug use to a greater extent than a group with a low rate of positive tests. Similarly, because only individuals who test negative can overreport drug use, a group with a high rate of negative tests will overreport drug use to a greater extent than a group with a low rate of negative tests. This will occur even if the true propensities to underreport and overreport drug use are equal across groups (see Appendix A for both a mathematical and empirical proof of this phenomenon). For example, males may underreport drug use to a greater extent than females because (1) males truly underreport drug use to a greater extent, or (2) males have more positive tests, and, as a result, more opportunities to underreport. True differences in underreporting and overreporting can only be discovered when opportunities to underreport and overreport are fixed across groups. Once opportunities to underreport and overreport are fixed across groups, certain groups of individuals may take

advantage of these opportunities to a greater extent than others. If so, true differences in underreporting and overreporting would exist.

Overall, this study examines differences in the accuracy of self-reported drug use and explains these differences in terms of differences in underreporting and overreporting. This study further examines differences in underreporting and overreporting to determine whether these are attributable to true differences or to differences in opportunity. 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 models, logit models, and logistic regression models with the 1994 Drug Use Forecasting data.

2. METHODS

2.1 Drug Use Forecasting (DUF) Data

This study uses data collected in 1994 as part of the Drug Use Forecasting Program. Self-report surveys on lifestyles and drug use and urine specimens were collected from adult arrestees across 23 sites in the United States[4]. The target population for all sites included male and female arrestees being held in a particular jurisdiction's detention facility. All arrestees were interviewed and asked for a urine specimen within 48 hours of their arrest. Although two sites collected data from less than 100 females each quarter, DUF sites typically collected data from

---

[4] These sites include: New York, NY; Washington, DC; Portland, OR; San Diego, CA; Indianapolis, IN; Houston, TX; Ft. Lauderdale, FL; Detroit, MI; New Orleans, LA; Phoenix, AZ; Chicago, IL; Los Angeles, CA; Dallas, TX; Birmingham, NY; Omaha, NE; Philadelphia, PA; Miami, FL; Cleveland, OH; San Antonio, TX; St. Louis, IL; San Jose, CA; Denver, CO; and Atlanta, GA.

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.

Once the urine specimens were collected, they were sent to a lab and analyzed for ten drugs: cocaine, opiates, marijuana, PCP, methadone, benzodiazepines, methaqualone, propoxphene, barbiturates, and amphetamines. Positive results for amphetamines were "confirmed by gas chromatography to eliminate those caused only by over-the-counter medications. For most drugs, urinanalysis can detect use within the previous 2 to 3 days; use of marijuana and PCP can sometimes be detected several weeks after use" (U.S. Department of Justice, 1996).

2.2 Advantages and Disadvantages of DUF Data

The DUF data provide a unique opportunity to examine the validity of self-reported drug use for several reasons. First, arrestees report their drug use for the past three days at the same time that urine is collected. Second, all urine specimens are collected within 48 hours of their arrest. Finally, testing the validity of self-reported drug use with arrestee data contributes significantly to the debate on the validity of self-report data regarding sensitive information such as drug use and criminality. Using arrestees to examine the validity of self-reported drug use is particularly beneficial from a policy standpoint. Arrestee drug use is arguably different than that

measured by other national drug use indicators because DUF has the ability to reach hidden populations often missed by these indicators (e.g., homeless and transient populations). Further, it seems plausible that the drug users in the arrestee population often represent chronic users who pose the greatest threat to themselves and society, and who could benefit from treatment the most. Assessing treatment needs often relies on the accuracy of self-reported drug use as well as the consideration of other arrestee characteristics. This study considers these interactions and their relationship to the validity of self-report drug use.

The primary disadvantage to using the DUF data is that interviewers and interview procedures are not completely standardized across sites. These differences across sites (e.g., being interviewed in front of a detention guard vs. 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 four sites. Due to these low sample sizes, we are unfortunately unable to determine whether significant differences across sites exist. The statistical power of our analyses is too low to examine site differences. More simplistic analyses are required in order to examine site differences. However, even if site differences were examined, interpretational confounding would likely occur. Very little documentation (if any) on site-specific protocols is available. While site differences may be uncovered, it would be very difficult to link such differences to specific factors.

Other disadvantages to using the DUF data are important to note but do not pose significant threats to this particular study. The DUF data are not representative of everyone arrested or everyone who uses drugs because samples are based on convenience and purposive sampling procedures rather than on a random sampling procedure. Consequently, several biases

are inherent in these data. First, the sample is not representative of the general drug-using population and does not capture arrestees who are arrested, booked, and released. While it certainly would be interesting to examine the extent to which measures of drug use reported by non-criminal samples are valid indicators of actual drug use, this is neither possible nor our intent. Such data (i.e., containing both self-report data and urinalysis results) for non-criminal samples are scarce. Second, certain types of offenses will be over represented while others are under represented. The fact that certain offenses will be over represented relative to others is not problematic. The degree to which self-reported drug use is corroborated by the urinalysis results will not be affected by such disproportionate sampling.

2.3 Sample

The sample consists of the 1994 data for White and Black adults from Indianapolis, Ft. Lauderdale, Phoenix, and Dallas. These four sites were chosen 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. More specifically, these four sites were the only sites which contained at least 20 respondents with a negative test for marijuana use and a negative self-report, at least 20 respondents with a positive test for marijuana use and a negative self-report, at least 20 respondents with a negative test for marijuana use and a positive self-report, and at least 20 respondents with a positive test for marijuana use and a positive self-report. Furthermore, these four sites were the only sites which contained at least 20 respondents with a negative test for crack/cocaine use and a negative self-report, at least 20 respondents with a positive test for crack/cocaine use and a negative self-report, at least 20 respondents with a negative test for crack/cocaine use and a positive self-report, and at least 20 respondents with a positive test for crack/cocaine use and a positive self-

report. 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. More specifically, cases were randomly assigned to contribute information either on marijuana use or on crack/cocaine use (i.e., no case was allowed to contribute information on both marijuana use and on crack/cocaine use). To ensure that the proportions of positive and negative self-reports and drug tests of marijuana and crack/cocaine use were not altered, a stratified randomization procedure was used. This stratified randomization procedure is illustrated in Appendix B.

The marijuana group includes approximately half of those who tested positive for marijuana. The marijuana comparison group includes approximately half of those who tested negative for marijuana. The crack/cocaine group includes approximately half of those who tested positive for crack/cocaine. Finally, the crack/cocaine comparison group includes approximately half of those who tested negative for crack/cocaine. As an example, consider two individuals who tested positive for marijuana but negative for crack/cocaine. One was randomly chosen to provide information on marijuana use and was included in the marijuana group. The other was randomly chosen to provide information on crack/cocaine use and was included in the crack/cocaine comparison group. In the end, 680 cases were gathered for the marijuana sample, 1,689 for the marijuana comparison sample, 976 for the crack/cocaine sample, and 1,407 for the crack/cocaine comparison sample. The final sample consists of all cases from these four groups merged into a single data file.

The adequacy of this stratified sampling technique was checked in several ways. First, the random assignment procedure was forced to provide two groups of roughly equal size. If the two groups created varied in size by more than 10 cases or by more than 10% of the combined sample size, the randomization sequence was rejected and a new one was created. Second, the randomization sequence was rejected and a new one created if the distributions of gender, race, age, and offense seriousness within drug test categories were significantly altered from the original data.

2.4 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 accuracy (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 marijuana, crack, and cocaine within the previous three days. The drug tests can generally detect the use of these drugs for two to three 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. In addition, it would not be entirely surprising if individuals were more likely to underreport marijuana use than crack/cocaine use. It would, however, be surprising if individuals were more likely to overreport marijuana use than crack/cocaine use.

2.5 Sample Characteristics

Descriptive statistics for all endogenous and exogenous measures are shown in Table 1. The data contain 4,752 cases. Subjects were fairly equally distributed across the four chosen site locations. As indicated, roughly half of the subjects contributed information on marijuana use while the other half contributed information on crack/cocaine use. Slightly more than half of the cases were between 18 and 30 years of age. The majority of individuals were male. Females comprised 32% of data. With regards to race, approximately half of the subjects were Black. The remaining 50% of cases were White. Thirty-eight percent of offenders were arrested for a misdemeanor while 62% were arrested for a felony. The majority (78%) of self-reports were accurate. Respectively, only 16% and 6% of arrestees underreported and overreported drug use.

2.6 Procedures

The first analyses examine differences in the accuracy 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, the sources of differences in the underreporting and overreporting of drug use are examined with 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. In the sub-sample with positive drug tests, all arrestees have an equal opportunity to underreport. In the sub-sample with negative drug tests, all arrestees have an equal opportunity to overreport. When using these sub-samples, differential opportunities are controlled for. As a result, any remaining difference across groups

is attributable to a true difference. The following sections describe in more detail the use of hierarchical loglinear, logit, and logistic regression models.

2.6.1 Hierarchical Loglinear Models

The data represent a 2^6 (i.e., 2X2X2X2X2X2) 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 necessary (i.e., 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 (Bishop et al., 1975; Fienberg, 1980; Dillon and Goldstein, 1984; Agresti, 1990; Ishii-Kuntz, 1994). More specifically, the following seven models are evaluated for our analyses:

(1) Model #1: no main effects and no interaction terms

(2) Model #2: all six main effects but no interaction terms,

(3) Model #3: all six main effects and all 15 two-way interactions,

(4) Model #4: all six main effects, all 15 two-way interactions, and all 20 three-way interactions,

(5) Model #5: all six main effects, all 15 two-way interactions, all 20 three-way interactions, and all 15 four-way interactions,

(6) Model #6: all six main effects, all 15 two-way interactions, all 20 three-way interactions, all 15 four-way interactions, and all six five-way interactions, and

(7) Model #7: all six main effects, all 15 two-way interactions, all 20 three-way interactions, all 15 four-way interactions, all six five-way interactions, and the six-way interaction.

Model #7 is called the saturated model and provides a perfect fit to the data (i.e., it is able to exactly reproduce the observed cell counts).

For each of the seven models, 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). Conversely, if this Chi-Square statistic is significant, one can conclude that the model provides a poor fit to the data (i.e., the predicted cell counts are significantly different than the observed ones). More interestingly, the seven 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. A difference between two Chi-Square statistics is itself a Chi-Square statistic with degrees of freedom equal to the difference in degrees of freedom. As an example, consider model #7 and the more parsimonious model #5. The terms omitted from model #5 (i.e., all five- and six-way interactions) are significant if the increase in the Chi-Square statistic between models #7 and #5 is significant. More precisely, if the difference in the Chi-Square statistic between model #7 and model #5 is significant, then at least one of the terms omitted from model #5 is significant (i.e., model #5 provides a significantly worse fit to the data than model #7). Conversely, if the difference in the

Chi-Square statistic between model #7 and model #5 is not significant, then none of the terms omitted from model #5 are significant (i.e., model #5 does not provide a significantly worse fit to the data than model #7).

Using this strategy, models #6 and #7 are compared to determine whether the six-way interaction is significant, #5 and #6 to determine whether all five-way interactions are significant, #4 and #5 to determine whether all four-way interactions are significant, #3 and #4 to determine whether all three-way interactions are significant, #2 and #3 to determine whether all two-way interactions are significant, and #1 and #2 to determine whether all main effects are significant. Furthermore, models #5 and #7 are compared to determine whether all six- and five-way interactions are significant, #4 and #7 to determine whether all six-, five-, and four-way interactions are significant, #3 and #7 to determine whether all six-, five-, four-, and three-way interactions are significant, #2 and #7 to determine whether all interactions are significant, and #1 and #7 to determine whether all interactions and main effects are significant.

2.6.2 Logit Models

The hierarchical loglinear models are useful to eliminate all interactions of a specified order (i.e., all five-way interactions). Hierarchical loglinear models are not useful, however, to eliminate specific interactions which are not of interest (i.e., interactions not involving the endogenous measure). Because the dependent variable in hierarchical loglinear models is the cell count from the contingency tables, all possible interactions are considered, including those not involving the endogenous measure. 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.

The hierarchical loglinear models are also not useful to eliminate specific interactions which are not significant. A backward elimination procedure was used to eliminate nonsignificant interaction terms and main effects in the logit models. 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. More specifically, the backward elimination procedure considers the final model from the hierarchical loglinear analysis (henceforth the HLL model) as the best model. The following five logit models and Chi-Square statistics are then estimated:

(1) best model without main effect of gender or interactions involving gender,

(2) best model without main effect of race or interactions involving race,

(3) best model without main effect of offense seriousness or interactions involving offense seriousness,

(4) best model without main effect of age or interactions involving age, and

(5) best model without main effect of type of drug or interactions involving type of drug.

One of these five logit models then becomes the new best model if it provides the smallest increase in the Chi-Square statistic from the HLL model and if this increase is nonsignificant (i.e., if the fit provided to the data does not become significantly worse). This procedure is repeated until no model provides an increase in the Chi-Square statistic from the HLL model that is not significant. When the backward elimination procedure is complete, the best model becomes the final model. The accuracy of all backward elimination procedures was checked with a forward selection procedure. Identical results were always obtained.

2.6.3 Logistic Regression Models

One problem with logit models is that they are difficult to interpret. For ease of interpretation and presentation, the final logit models are converted to logistic regression models. In these model, the slopes represent the expected effect of the independent variables on the log-odds of the dependent variable. Predicted probabilities are then computed as:

$$p = \frac{e^{b_0 + \sum b_i x_i}}{1 + e^{b_0 + \sum b_i x_i}} \quad , \tag{1}$$

where b_0 is the intercept, b_i are slopes, and x_i are main effects or interactions.

3. RESULTS

3.1 Models of Accuracy

3.1.1 Hierarchical Loglinear Models

The results from the hierarchical loglinear model for accuracy are presented in Table 2. This table shows the 11 comparisons mentioned in section 2.6.1. More precisely, the first row presents the significance of the six-way interaction. The second row presents the significance of all five-way interactions and the joint significance of all five- and six-way interactions. The third row presents the significance of all four-way interactions and the joint significance of all four-, five-, and six-way interactions. The fourth row presents the significance of all three-way interactions and the joint significance of all three-, four-, five-, and six-way interactions. The fifth row presents the significance of all two-way interactions and the joint significance of all interactions. Finally, the last row presents the significance of all main effects and 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 ($p = 0.53$). 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 HLL model (i.e., the final model from the hierarchical loglinear analysis) therefore contains all main effects and two-way interactions.

3.1.2 Logit Models

The results from the logit models are presented in Table 3. In this table, the first model is the HLL model (i.e., all main effects and two-way interactions). All interactions not involving accuracy are now dropped from the model. Accuracy becomes the dependent variable. As a result, the two-way interactions involving accuracy (e.g., accuracy by gender) become main effects. The main effect of accuracy becomes the constant. The backward selection procedure therefore starts with the following model[5]: [D] [R] [O] [A] [S], where D= Drug, R= Race, O= Offense, A= Age, and S= Sex. This model hypothesizes that accuracy is a function of type of drug, race, offense seriousness, age, and gender. The backward elimination procedure then attempts to eliminate nonsignificant main effects.

For each model in Table 3, 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 HLL model and subsequent

---

[5] This notation is an abbreviation used to uniquely identify loglinear and logit models (see, for example, Fienberg, 1980). [DR] is an abbreviation for the main effect of type of drug, the main effect of race, and the interaction between drug and race. [R] would simply be an abbreviation for the main effect of race. All models include a constant term.

models are also reported. These differences in Chi-Square statistics are used to show that the fit provided to the data is never significantly worse than the fit provided to the data by the HLL model (i.e., all p-values are nonsignificant).

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 ($p = .671$). Second, the main effect of age was removed. Removing this main effect did not significantly reduce the fit provided to the data ($p = .835$). Third, the main effect of offense seriousness was removed. Again, removing this term did not significantly reduce the fit provided to the data ($p = 0.801$). Finally, the main effect of type of drug was removed. Removing this term did not reduce the fit provided to the data ($p = 0.739$). 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 shows that accuracy is solely a function of race.

3.1.3 Logistic Regression Models

The results from the logistic regression models are presented in Table 4. The results indicate that the log-odds of a self-report being accurate are significantly higher for Whites than for Blacks. More specifically, the predicted probability of accuracy 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 differences in underreporting and overreporting.

3.2 Models of Underreporting

3.2.1 Hierarchical Loglinear Models

Results shown in Table 5 reveal that all six-, five-, and four-way interactions are not

significant. Removing all six-, five-, and four-way interactions would not significantly reduce the fit provided to the data ($p = 0.843$). However, at least one of the three-way interactions is significant ($p = 0.018$). Eliminating all three-way interactions would significantly reduce the fit provided to the data. On the other hand, removing all six-, five-, four-, and three-way interactions would not significantly reduce the fit provided to the data ($p = 0.162$). Overall, eliminating all three-, four-, five-, and six-way interactions would not significantly reduce the fit provided to the data, but at least one of the three-way interactions is significant. 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 HLL model therefore contains all main effects and all two- and three-way interactions.

3.2.2 Logit Models

The results from the logit models are presented in Table 6. In this table, the first model is the HLL model (i.e., all main effects and all two- and three-way interactions). The backward selection procedure therefore starts with the following model: [DR] [DO] [DA] [DS] [RO] [RA] [RS] [OA] [OS] [AS], where D= Drug, R= Race, O= Offense, A= Age, and S= Sex. This model hypothesizes 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. The backward elimination procedure is then utilized to eliminate nonsignificant main effects and the nonsignificant interactions involving these nonsignificant main effects.

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 ($p = .916$). Second, the main effect of age and all interactions

involving age were removed. Removing this main effect and interactions did not significantly reduce the fit provided to the data (p= .571). 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= 0.357). 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.

3.2.3 Logistic Regression Models

The results from the logistic regression models are presented in Table 7. The 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 nonsignificant, 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. Predicted probabilities of underreporting across race groups and types of drug are plotted in Figure 1. 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.

As previously noted, these differences may be due to true differences or to differences in opportunity. The logistic regression model of underreporting was also evaluated in the sub-sample of offenders with positive drug tests. In this sub-sample, all offenders have the opportunity to underreport drug use. Results (also shown in Table 7) reveal that the interaction between race and type of drug becomes non-significant when controlling for differences in

offenders to underreport crack/cocaine use, Black offenders do not take advantage of this opportunity to a greater extent. Black offenders underreport crack/cocaine use to a greater extent than White offenders because, and solely because, they have more opportunities to do so. The race difference in underreporting is not a true difference. The main effect of type of drug is still statistically significant. Offenders are more likely to underreport crack/cocaine use than marijuana use. This is a true difference.

3.3 Models of Overreporting

3.3.1 Hierarchical Loglinear Models

Results show that all six-, five-, four-, and three-way interactions are not statistically significant. Removing all six-, five-, four-, and three-way interactions does not appear to significantly reduce the fit provided to the data ($p = 0.7938$). 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. The HLL model therefore contains all main effects and two-way interactions.

3.3.2 Logit Models

The results from the logit models are presented in Table 9. In this table, the first model is the HLL model (i.e., all main effects and two-way interactions). The backward selection procedure therefore starts with the following model: [D] [R] [O] [A] [S], where D= Drug, R= Race, O= Offense, A= Age, and S= Sex. This model hypothesizes that underreporting is a function of type of drug, race, offense seriousness, age, and gender. The backward elimination procedure then attempts to eliminate nonsignificant main effects.

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

increase in the Chi-Square statistic. In addition, the increase in the Chi-Square statistic was not significant ($p = .417$). Second, the main effect of age was removed. Removing this main effect did not significantly reduce the fit provided to the data ($p = .415$). Finally, the main effect of offense seriousness was removed. Once again, removing this term did not significantly reduce the fit provided to the data ($p = 0.399$). No further terms could be removed. 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.

3.3.3 Logistic Regression Models

The results from the logistic regression models are presented in Table 10. The results 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. Predicted probabilities of overreporting across race groups and drug categories are plotted in Figure 2. 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.

As previously noted, these differences may again be due to true differences or to differences in opportunity. The logistic regression model of overreporting was also evaluated in the sub-sample of offenders with negative drug tests. In this sub-sample, all offenders have the opportunity to overreport drug use. Results (also shown in Table 10) reveal that all effects remain statistically significant even when differences in opportunity are controlled for.

are more likely to overreport drug use than White offenders. These are true differences.

3.4 Summary of Results

The logistic regression model for accuracy revealed that accuracy was a function of race. Black offenders provided less accurate self-reports than White offenders. This difference was explained by differences in underreporting and overreporting. The logistic regression results showed that Black offenders were more likely to underreport crack/cocaine use than White offenders, but that Black offenders were not more likely to underreport marijuana use than White offenders. The race difference in underreporting existed only for crack/cocaine use. In addition, this difference disappeared once opportunity was controlled for. Black offenders were more likely to underreport crack/cocaine use simply because a higher proportion of Black offenders tested positive for crack/cocaine use than White offenders. The race difference in underreporting is attributable to an opportunity difference rather than to a true difference. Black offenders were also more likely to overreport both marijuana and crack/cocaine use relative to White offenders. This difference was attributable to a true difference. When controlling for opportunity, Black offenders were still more likely to overreport both marijuana and crack/cocaine use relative to White offenders.

To briefly summarize, Black offenders have less accurate self-reports of marijuana and crack/cocaine use than White offenders. More specifically, Black offenders are more likely to underreport crack/cocaine use and more likely to overreport both marijuana and crack/cocaine use. However, Black offenders are more likely to underreport crack/cocaine use simply because a higher proportion of Black offenders test positive for crack/cocaine than White offenders. The race difference in overreporting, on the other hand, appears to be a true difference.

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 urinalysis test is more sensitive to marijuana use than to crack/cocaine use. Because marijuana use can generally be detected for longer periods than crack/cocaine use, we would expect that offenders would be more likely to underreport marijuana use rather than crack/cocaine use. This result should be investigated further. The underreporting and overreporting effects canceled each other out in the accuracy analyses. Because offenders were more likely to underreport and overreport different types of drugs, the accuracy of self-reported drug use was not affected by type of drug. None of the underreporting and overreporting differences across types of drug could be explained by differences in opportunity.

4. CONCLUSIONS

This investigation was designed to examine the validity of self-reported drug use across five factors - gender, race, age, offense seriousness, and type of drug. More specifically, we examined how these five factors and the interactions between these five factors would predict the accuracy of self-reported drug use. Only the main effect of race emerged as a significant predictor of the accuracy of self-reported drug use. Black offenders provided less accurate reports of drug use than White offenders. This difference was then explained by showing that Black offenders were more likely to underreport crack/cocaine use than White offenders. This difference, however, was due solely to an opportunity difference. The difference in the accuracy of self-reported drug use across racial groups was also explained by showing that Black offenders were more likely to overreport both marijuana and crack/cocaine use than White offenders. These differences were attributable to true differences.

The results indicated that gender, offense seriousness, age, and type of drug do not affect the accuracy of self-reported drug use. These results strongly support the further use of the DUF data to examine patterns of drug use. In addition, they strongly support the use of self-report data on drug use for research and policy development purposes. Nevertheless, there are four important limitations. First, while type of drug does not have an effect on the accuracy 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 accurate reports of drug use than White offenders. Third, Black offenders are more likely to underreport crack/cocaine use than White offenders. 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 opportunity does not mean that self-reports of crack/cocaine use are equally valid across racial groups. The fact that the race effect disappears when opportunity is controlled for does not mean that valid inferences can be reached when comparing self-reports of crack/cocaine use across racial groups. 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 have the opportunity to do so. Among offenders who test positive for crack/cocaine use, race does not affect the likelihood of underreporting. As shown in Appendix A, the effect of race on underreporting will increase as the differences in opportunity 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. However, while race will not affect the likelihood of underreporting in samples where different racial groups have equal opportunities to underreport,

race will still affect the likelihood of overreporting, even in samples where different racial groups have equal opportunities to overreport. Black offenders are more likely to overreport both marijuana and crack/cocaine use than White offenders. This difference is not attributable to an opportunity difference.

In addition, the effects of type of drug on underreporting and overreporting could not simply be explained by differences in opportunity 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. These differences could not be explained by opportunity differences. While it is beyond the scope of this investigation, it is important to further examine the true differences uncovered in this research. Overreporting and underreporting may emerge due to a variety of factors. Future investigations should go beyond identifying individual characteristics that are related to underreporting and overreporting. Future investigations should explain why certain individuals are compelled to underreport and overreport drug use. Theoretical frameworks should be developed to explain the true differences in underreporting and overreporting. Future analyses should examine the mechanisms which connect individual demographic and social characteristics to underreporting and overreporting.

Future analyses should also describe the contexts in these mechanisms operate. Other factors may affect the accuracy of self-reported drug use and such factors should be explored. Specifically, it is important for future investigations to examine the extent to which data collection procedures affect the validity of self-reported drug use. As previously mentioned, it is unfortunate that site-specific differences in data collection procedures were not more carefully documented. Without careful documentation of these procedures, it is of little value to examine differences across sites. While differences across sites could be found, there would be no way to

interpret such differences. The ADAM project should more carefully document all data collection procedures. Only then will we be able to determine if these procedures have an effect on the validity of self-reported drug use.

The analyses presented here clearly showed that some true differences in the accuracy, 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 accuracy, underreporting, and overreporting of drug use are relatively rare. Some of these rare differences can simply be attributed to differences in opportunity. 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.

REFERENCES

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-Analysis. Criminology, 28, 369-404.

Anglin, M. D., Hser, Y. I., & Chou, C. P. (1993). Reliability and Validity of Retrospective Behavioral Self-Report by Narcotics Addicts. Evaluation Review, 17, 91-108.

Agresti, A. (1990). Categorical Data Analysis. New York: John Wiley & Sons.

Bale, R. N., Van Stone, W. W., Engelsing, T. M., & Zarcone, V. P. (1981). The Validity of Self-Reported Heroin Use. International Journal of the Addictions, 16, 1387-1398.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). Discrete Multivariate Analysis. Cambridge, MA: MIT.

Brown, J., Kranzler, H. R., & Del Boca, F. K. (1992). Self-Reports by Alcohol and Drug Abuse Inpatients: Factors Affecting Reliability and Validity. British Journal of Addictions, 87, 1013-1024.

Dembo, R., Williams, L., Wish, E. D., & Schmeidler, J. (1990). Urine Testing of Detained Juveniles to Identify High-Risk Youth. Washington: National Institute of Justice, U.S. Department of Justice.

Dillon, W. R., & Goldstein, M. (1984). Multivariate Analysis: Methods and Applications. New York: Wiley.

Eckerman, W., Bates, J. D., Rachal, J. V., & Poole, W. K. (1971). Drug Usage and Arrest Charges. Washington: Bureau of Narcotics and Dangerous Drugs.

Falck, R., Siegal, H. A., Forney, M. A., Wang, J., & Carlson, R. G. (1992). The Validity of Injection Drug Users Self-Reported Use of Opiates and Cocaine. Journal of 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. Journal of Drug Issues, 24, 99-116.

Fienberg, S. E. (1980). The Analysis of Cross-Classified Categorical Data. Cambridge, MA: MIT Press.

Funkhouser, A. W., Butz, A. M., Feng, T. I., McCaul, M. E., & Rosenstein, B. J. (1993). Prenatal Care and Drug Use in Pregnant Women. Drug and Alcohol Dependence, 33, 1-9.

Gray, T. A. (1996). The Validity of Self-Reported Drug Use: An Assessment of Female Arrestee Drug Users. Unpublished master's thesis, University of Maryland at College Park, College Park, MD.

Harrell, A. V. (1985). Validation of Self-Report: The Research Record. In B. Rouse, N. Kozel, & L. Richards (Eds.), Self-Report Methods of Estimating Drug Use [NIDA Research Monograph, 57]. Rockville, MD: NIDA.

Harrison, L. D. (1995). The Validity of Self-Reported Data on Drug Use. Journal of Drug Issues, 25, 91-111.

Ishii-Kuntz, M. (1994). Ordinal Log-Linear Models. Thousand Oaks, CA: Sage.

Katz, C. M., Webb, V. J., Gartin, P. R., & Marshall, C. E. (1997). The Validity of Self-Reported Marijuana and Cocaine Use. Journal of Criminal Justice, 25, 31-41.

Maddux, J. F., & Desmond, D. P. (1975). Reliability and Validity of Information from Chronic Heroin Users. Journal of Psychiatric Research, 12, 87-95.

Magura, S., Goldsmith, D., Casriel, C., Goldstein, P. J., & Lipton, D. S. (1987). The Validity of Methadone Clients' Self Reported Drug Use. International Journal of Addictions, 22, 727-749.

Magura, S., & Kang, S. Y. (1995). Validity of Self-Reported Drug Use in High Risk Populations: A Meta-Analytic Review. New York: National Development and Research Institutes.

Magura, S., Kang, S. Y., & Shapiro, J. L. (1995). Measuring Cocaine Use by Hair Analysis Among Criminally-Involved Youth. Journal of Drug Issues, 25, 683-701.

Marques, P. R., Tippetts, A. S., & Branch, D. G. (1993). Cocaine in the Hair of Mother-Infant Pairs: Quantitative Analysis and Correlations with Urine Measures and Self-Report. American Journal of Drug and Alcohol Abuse, 19, 159-175.

McGlothlin, W. H., Anglin, D. A., & Wilson, B. D. (1977). An Evaluation of the California Civil Addict Program. Rockville, MD: NIDA.

Page, W. F., Davies, J. E., Ladner, R. A., Alfassa, J., & Tennis, H. (1977). Urinalysis Screened Vs Verbally Reported Drug Use: The Identification of Discrepant Groups. International Journal of Addictions, 12, 439-450.

Tims, F. M., & Ludford, J. P. (1984). Drug Abuse Treatment Evaluation: Strategies, Progress and Prospects [NIDA Research Monograph, 51]. Rockville, MD: NIDA.

Wish, E. D., Hoffman, J. A., & Nemes, S. (1997). The Validity of Self-Reports of Drug Use at Treatment Admission and at Followup: Comparisons with Urinalysis and Hair Assays. In L. Harrison, & A. Hughes (Eds.), The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates. Rockville, MD: NIDA.

Wish, E., Johnson, B., Strug, D., Chedekel, M., & Lipton, D. (1983). Are Urine Tests Good Indicators of the Validity of Self-Reports of Drug Use? It Depends on the Test. New York: Narcotic and Drug Research.

Young, N. K. (1994). Invest in Treatment for Alcohol and Other Drug Problems: It Pays. Washington: National Association of State Alcohol and Drug Abuse Directors.

U.S. Department of Justice, National Institute of Justice (1996). Drug Use Forecasting: 1996 Annual Report. Washington: Government Printing Office.

Table 1

Descriptive Statistics for Endogenous and Exogenous Measures

	Number (Percent)
Site	
Indianapolis	1,214 (25.5%)
Ft. Lauderdale	1,215 (25.6%)
Phoenix	1,114 (23.4%)
Dallas	1,209 (25.4%)
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%)
Accuracy	
Not Accurate	1,065 (22.4%)
Accurate	3,687 (77.6%)
Underreporting	
Not Underreported	3,993 (84.0%)
Underreported	759 (16.0%)
Overreporting	
Not Overreported	4,446 (93.6%)
Overreported	306 (6.4%)

Table 2

Significance of Parameters in Loglinear Models for Accuracy

Parameters	Likelihood Ratio Chi-Square	df	p-value	Sum in Chi-Square	df	p-value
Six-way interaction	0.039	1	0.8425			
Five-way interactions	8.043	6	0.2350	8.082	7	0.3254
Four-way interactions	6.640	15	0.9669	14.723	22	0.8739
Three-way interactions	25.197	20	0.1940	39.919	42	0.5267
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

Table 3

Significance of Parameters in Logit Models for Accuracy

Model[1]	Likelihood Ratio Chi-Square	df	p-value	Difference in Chi-Square	df	p-value
[D] [R] [O] [A] [S]	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

Notes:
1 - D = Drug, R = Race, O = Offense, A = Age, S = Sex.

Table 4

Maximum Likelihood Results of Logistic Regression Model for Accuracy

Parameter	β (s.e.)
Constant	0.6859 (0.1069)*
Race[a]	0.3800 (0.0704)*
Model χ^2 (df)	29.478 (1)*
-2 Log Likelihood	5027.301

Notes:
a - Race: 0 = Black, 1 = White
* - $p < 0.01$

Table 5

Significance of Parameters in Loglinear Models for Underreporting

Parameters	Likelihood Ratio Chi-Square	df	p-value	Sum in Chi-Square	df	p-value
Six-way interaction	.09	1	.764			
Five-way interactions	9.31	6	.157	9.40	7	.225
Four-way interactions	6.04	15	.979	15.44	22	.843
Three-way interactions	35.49	20	.018	50.93	42	.162
Two-way interactions	283.89	15	<.001	334.82	57	<.001
Main Effects	3413.20	6	<.001	3748.02	63	<.001

Table 6

Significance of Parameters in Logit Models for Underreporting

Model[1]	Likelihood Ratio Chi-Square	df	p-value	Difference in Chi-Square	df	p-value
[DR] [DO] [DA] [DS] [RO] [RA] [RS] [OA] [OS] [AS]	12.10	16	.737			
[DR] [DA] [DS] [RA] [RS] [AS]	13.57	21	.887	1.47	5	.916
[DR] [DS] [RS]	19.74	25	.760	7.64	9	.571
[DR]	25.26	28	.614	13.16	12	.357

Notes:
1 - D = Drug, R = Race, O = Offense, A = Age, S = Sex. All models contain lower interaction terms and main effects (i.e., [DR] contains drug by race interaction and main effects of drug and race).

Table 7

Maximum Likelihood Results of Logistic Regression Model for Underreporting

	Full Sample	Sample with Opportunity
Parameter	β (s.e.)	β (s.e.)
Constant	-2.0204 (0.0887)*	-0.4238 (0.1072)*
Drug[a]	0.9196 (0.1110)*	0.3717 (0.1342)*
Race[b]	0.0472 (0.1267)	0.1821 (0.1560)
Race by Drug	-0.6859 (0.1648)*	-0.1742 (0.2047)
Model χ^2 (df)	95.527 (3)*	9.647 (3)**
-2 Log Likelihood	4174.24	2274.543

Notes:
a - Drug: 0 = Marijuana, 1 = Cocaine
b - Race: 0 = Black, 1 = White
* - $p < 0.01$
* - $p = 0.02$

Table 8

Significance of Parameters in Loglinear Models for Overreporting

Parameters	Likelihood Ratio Chi-Square	df	p-value	Sum in Chi-Square	df	p-value
Six-way interaction	1.007	1	0.3156			
Five-way interactions	5.176	6	0.5214	6.184	7	0.5185
Four-way interactions	10.260	15	0.8031	16.443	22	0.7933
Three-way interactions	17.890	20	0.5947	34.333	42	0.7938
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

Table 9

Significance of Parameters in Logit Models for Overreporting

Model[1]	Likelihood Ratio Chi-Square	df	p-value	Difference in Chi-Square	df	p-value
[D] [R] [O] [A] [S]	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

Notes:
1 - D = Drug, R = Race, O = Offense, A = Age, S = Sex.

Table 10

Maximum Likelihood Results of Logistic Regression Model for Overreporting

	Full Sample	Sample with Opportunity
Parameter	β (s.e.)	β (s.e.)
Constant	-2.0680 (0.0855)*	-1.6269 (0.0868)*
Drug[a]	-1.2735 (0.1388)*	-1.0736 (0.1412)*
Race[b]	-0.3069 (0.1210)*	-0.4365 (0.1236)*
Model χ^2 (df)	106.272 (2)*	84.576 (2)*
-2 Log Likelihood	2164.141	1912.471

Notes:
a - Drug: 0 = Marijuana, 1 = Cocaine
b - Race: 0 = Black, 1 = White
* - $p < 0.01$

Figure 1

Predicted Probabilities of Underreporting Across Groups

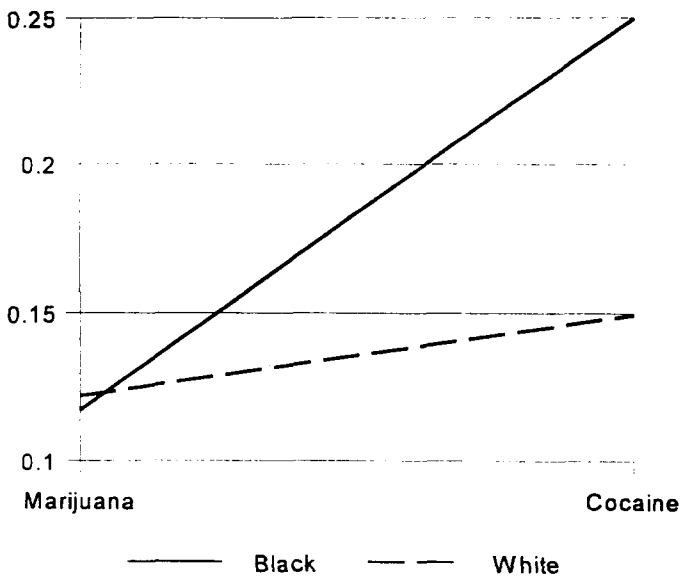
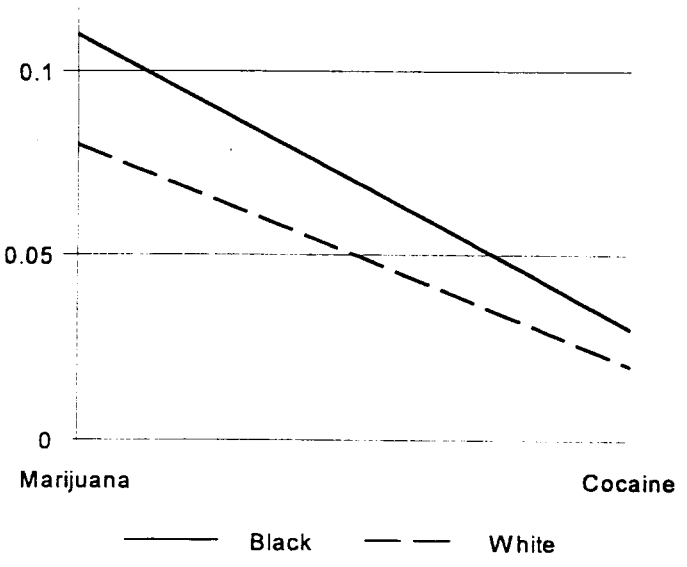


Figure 2

Predicted Probabilities of Overreporting Across Groups



Appendix A

Mathematical and Empirical Proof

Because only those who test positive for drug use can underreport drug use, a group with a high rate of positive drug tests will underreport drug use to a greater extent than a group with a low rate of positive drug tests. This will occur even if the true propensity to underreport drug use is equal across groups. A similar proof (not shown) can be derived for overreporting.

Mathematical Proof:

Define U as the event of underreporting drug use, PT as the event of a positive drug test, NT as the event of a negative drug test, PS as the event of a positive self-report, and NS as the event of a negative self-report. Then, operationalize U as 1 when NS and PT, and as 0 otherwise, so that $P(U) = P(NS \cap PT)$, or $P(U) = P(NS \mid PT)\, P(PT)$. The true propensity to underreport drug use is simply $P(NS \mid PT)$, the probability of a negative self-report given a positive drug test. But the probability of underreporting is also influenced by $P(PT)$, the probability of a positive drug test. As the probability of a positive drug test increases, the probability of underreporting also increases. As an example, suppose that group #1 has a $P(NS \mid PT) = \theta$, and a $P(PT) = \lambda$, and that group #2 also has a $P(NS \mid PT) = \theta$, but has a $P(PT) = \lambda + \delta$. Group #1 will then underreport drug use less than group #2, even though $P(NS \mid PT) = \theta$ in both groups. $P(U)$ will be lower in group #1 than in group #2 by $\theta\delta$. This difference is solely attributable to a difference in $P(PT)$, or a difference in opportunity.

Empirical Proof:

Suppose that for both group #1 ($G1$) and group #2 ($G2$), $P(U) = P(NS \mid PT) = .7$, so that members of $G1$ and $G2$ are equally likely to underreport drug use. Now suppose that $P(PT \mid G1)$

= .7, and that *P (PT | G2)* = .3, so that members of *G1* have more opportunities to underreport drug use than members of *G2*. Finally, suppose that the sample size for both *G1* and *G2* is 200. The following tables can then be constructed[6]:

Group #1			
Self-	Drug Test		
Report	Negative	Positive	Total
Negative	?	98	?
Positive	?	42	?
Total	60	140	200

Group #2			
Self-	Drug Test		
Report	Negative	Positive	Total
Negative	?	28	?
Positive	?	12	?
Total	160	40	200

We can then compute *P (U)* as follows:

$P(U \mid G1) = P(NS \cap PT) = 98 / 200 = .49$, and (A3)

$P(U \mid G2) = P(NS \cap PT) = 12 / 200 = .14$. (A4)

We now conclude that members of *G1* are substantially more likely than members of *G2* to underreport drug use[7]. But in this instance, members of *G1* are substantially more likely than members of *G2* to underreport drug use because, and only because, the proportion of positive drug tests is higher in *G1* than in *G2*. Members of *G1* have more opportunities to underreport drug use than members of *G2*. The difference in underreporting across groups is solely due to a difference in opportunity across groups. It is not a true difference.

---

[6] Not enough information is provided to complete these two tables. As a result, some cells remain undefined. However, as can be seen in the computation of *P (U)*, it makes no difference what these cells contain.

[7] In fact, this difference is statistically significant ($p < 0.01$).

Appendix B

Stratified Random Sampling Scheme to Obtain Independent Observations

Marijuana		Crack / Cocaine		Sample Size	Sampling
Test	Self-Report	Test	Self-Report		
Positive	Positive	Positive	Positive	180	86 Marijuana 94 Cocaine
Positive	Negative	Positive	Positive	64	29 Marijuana 35 Cocaine
Positive	Positive	Positive	Negative	184	93 Marijuana 91 Cocaine
Positive	Negative	Positive	Negative	163	81 Marijuana 82 Cocaine
Negative	Positive	Positive	Positive	176	88 Marijuana comparison 88 Cocaine
Negative	Negative	Positive	Positive	573	290 Marijuana comparison 283 Cocaine
Negative	Positive	Positive	Negative	87	44 Marijuana comparison 43 Cocaine
Negative	Negative	Positive	Negative	520	260 Marijuana comparison 260 Cocaine
Positive	Positive	Negative	Positive	22	12 Marijuana 10 Cocaine comparison
Positive	Negative	Negative	Positive	10	5 Marijuana 5 Cocaine comparison
Positive	Positive	Negative	Negative	418	206 Marijuana 212 Cocaine comparison
Positive	Negative	Negative	Negative	339	168 Marijuana 171 Cocaine comparison
Negative	Positive	Negative	Positive	31	16 Marijuana comparison 15 Cocaine comparison
Negative	Negative	Negative	Positive	81	40 Marijuana comparison 41 Cocaine comparison
Negative	Positive	Negative	Negative	170	87 Marijuana comparison 83 Cocaine comparison
Negative	Negative	Negative	Negative	1734	864 Marijuana comparison 870 Cocaine comparison