Comparing Drug Testing and Self-Report of Drug Use among Youths and Young Adults in the General Population

Lana D. Harrison
Steven S. Martin
Tihomir Enev
Deborah Harrington

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Substance Abuse and Mental Health Services Administration
Office of Applied Studies

Acknowledgments

This report was prepared by the Division of Population Surveys, Office of Applied Studies, SAMHSA; the Center for Drug and Alcohol Studies, University of Delaware, Newark, Delaware; and RTI International, a trade name of Research Triangle Institute, Research Triangle Park, North Carolina. Work by the University of Delaware was performed under Grant No. R01 DA10930 from the National Institute on Drug Abuse and SAMHSA Contract No. 283-98-9008. Work by RTI was performed under SAMHSA Contract Nos. 283-97-9007 and 283-98-9008. Contributors at the University of Delaware listed alphabetically include Tihomir Enev, Deborah Harrington, Lana D. Harrison, and Steven S. Martin. Contributors at SAMHSA listed alphabetically include Joseph Gfroerer and Arthur Hughes. Contributors at RTI listed alphabetically include Brian Burke, Rachel Caspar, Lucinda Casson, Patrick Chen, James Chromy, Lee Ellen Coffey, Teresa Davis, Elizabeth (Wiebe) Dean, Steven Emrich, G.G. Frick, Rebecca Granger, David Heller, Donna Hewitt, Marjorie Hinsdale-Shouse, Larry A. Kroutil, Mary Ellen Marsden, Martin Meyer, Susan Meyers, Dawn Odom (now at Inveresk), Lisa E. Packer, Lanny Piper, and Thomas G. Virag (Project Director). Staff in the Division of Workplace Programs at SAMHSA and RTI's Center for Forensic Sciences reviewed the document. Reviewers at the National Institute on Drug Abuse listed alphabetically include James Colliver (now at SAMHSA), Marilyn Huestis, Lucinda Miner, and Yonette Thomas. At RTI, Richard S. Straw edited the report. Also at RTI, Joyce Clay-Brooks and Roxanne Snaauw word processed the report, and Pamela Couch Prevatt, Teresa G. Bass, Shari B. Lambert, and Wendy Broome prepared its press and Web versions. Final report production was provided by Beatrice Rouse, Coleen Sanderson, and Jane Feldmann at SAMHSA.

Public Domain Notice

Obtaining Additional Copies of Publication

Copies may be obtained, free of charge, from the National Clearinghouse for Alcohol and Drug Information (NCADI), a service of SAMHSA. Write or call NCADI at:

National Clearinghouse for Alcohol and Drug Information
P.O. Box 2345, Rockville, MD 20847-2345
1-240-221-4017, 1-800-729-6686, TDD 1-800-487-4889

Electronic Access to Publication

This publication can be accessed electronically through the Internet connections listed below:
http://www.samhsa.gov and http://www.oas.samhsa.gov

Originating Office

SAMHSA, Office of Applied Studies
1 Choke Cherry Road, Room 7-1044
Rockville, MD 20857

May 2007

Table of Contents

Table of Contents (continued)

List of Tables

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

Highlights

This report presents the results of a Validity Study conducted in 2000 and 2001 in conjunction with the National Household Survey on Drug Abuse (NHSDA), an annual survey conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies (OAS), to track the prevalence of substance use in the United States.[1] The purpose of the Validity Study was to provide information on the validity of self-reported drug use in a general population survey by comparing the self-reports of respondents with the results of drug tests of urine and hair specimens obtained from those same respondents. The study also was designed to investigate methodological issues, such as technical aspects of collecting urine and hair, the willingness of respondents to provide specimens, and questionnaire strategies.

The Validity Study was sponsored jointly by SAMHSA and the National Institute on Drug Abuse (NIDA), and it was conducted by researchers at the University of Delaware in collaboration with the NHSDA data collection contractor, RTI International.[2] A brief description of the design and results of the study is given below.

Study Design

- NHSDA (now NSDUH) is an annual survey of the civilian, noninstitutionalized population of the United States aged 12 or older. Since 1999, the survey has collected data on the use of alcohol, tobacco, and illicit substances from nearly 70,000 respondents at their places of residence. Data are collected using computer-assisted interviewing (CAI), with sensitive questions administered via audio computer-assisted self-interviewing (ACASI) methods. The Validity Study was conducted as a supplement to the 2000 and 2001 NHSDAs.

- A separate national (excluding Alaska and Hawaii) sample of almost 6,000 persons aged 12 to 25 was selected for the Validity Study, and more than 4,400 completed an interview. These respondents were interviewed using the NHSDA methodology, with a slightly altered questionnaire to eliminate questions not needed for the study and to obtain key information needed for the study.

- Some of the questionnaire changes included adding questions about drug use corresponding to shorter time periods in order to be comparable with the window of detection of most drugs in urine and hair. In addition, a persuasion experiment or appeal was embedded in the study in which half of the respondents were given a statement emphasizing the importance of accurate reporting before other questions specific to the Validity Study were asked.

[1] As of 2002, the survey is called the National Survey on Drug Use and Health (NSDUH).
[2] RTI International is a trade name of Research Triangle Institute.

- At the end of each interview, respondents were asked to provide a hair and a urine specimen, with an incentive payment of $25 for each. Exactly 4,000 respondents provided at least one specimen. Specimens were mailed to a testing laboratory, which conducted the drug tests and sent the results to the study team at the University of Delaware.

- Urine and hair specimens were screened using an immunoassay test for the following drug classes: marijuana/hashish (cannabinoids), cocaine, amphetamines, and opiates. Specimens that screened positive were tested using gas chromatography/mass spectrometry (GC/MS) as the confirmatory drug test. Urine specimens also were analyzed for the presence of cotinine, the principal metabolite of nicotine, using only an immunoassay test. No confirmatory testing was performed for cotinine.

- For urine specimens, self-reported drug use (i.e., 30-day, 7-day, 3-day) was compared with confirmatory test results (i.e., positive or negative using the Validity Study cutoffs).

- Self-reported tobacco use (i.e., 30-day, 7-day, 3-day) was compared with the results of the urine test for cotinine (i.e., positive or negative using a 100 nanograms per milliliter [ng/mL] cutoff).

Methodological Findings and Limitations

- Comparisons of self-reports and hair test results were not included in this report due to technical and statistical problems in the study, as well as some unresolved issues concerning the interpretation of test results for drugs of abuse in hair. Hair test results are presented in Chapter 1 and Appendix E.

- Approximately 81 percent of those interviewed provided both a urine and a hair specimen, and about 89 percent provided at least one specimen.

- For validity studies such as this, when comparing self-reports and urine drug test results, the drug test is assumed to be 100 percent accurate. However, one must keep in mind that detection time windows for drugs in biological specimens are inexact, estimated time ranges based on research and case studies. Many biological factors, as well as analytical factors, affect the window of detection for a drug and its metabolites in urine.

- The Validity Study's survey and urinalysis protocols did not allow differentiation of respondents reporting the use of codeine, morphine, or heroin from those reporting the use of prescription pain relievers other than codeine or morphine.

- The Validity Study's survey and urinalysis protocols did not allow differentiation of respondents reporting the use of amphetamine or methamphetamine from those reporting the use of stimulants other than amphetamine and methamphetamine. In addition, the protocols did not allow respondents to report all legitimate or prescription amphetamine or methamphetamine use.

- It appears that the added persuasion (appeal) improved the accuracy of self-reports.

Results of Comparison of Self-Reports and Drug Tests

- For tobacco, there was 84.6 percent agreement between self-report in the past 30 days and urine drug test results. About 5.8 percent reported no use and tested positive, but nearly 1 in 10 (9.6 percent) reported use and did not test positive.

- The 7-day results for tobacco show that 8.6 percent reported no use of cigarettes or cigars but tested positive and 5.4 percent reported use and did not test positive. The 3-day results show that 7.7 percent reported no use of cigarettes or cigars and tested positive and 3.6 percent reported use and did not test positive. Some of those who reported no use but tested positive may have used tobacco products other than cigarettes and cigars.

- For marijuana, there was 89.8 percent agreement between self-reported use in the past 30 days and urine drug test results, although this was dominated by 82.9 percent who reported no use and tested negative. About 4.4 percent reported no use and tested positive, and 5.8 percent reported use in the past 30 days and did not test positive.

- Using a 7-day period for comparison of marijuana self-report and urine drug test results, 4.8 percent reported no use and tested positive, and 2.7 percent reported use and did not test positive. Agreement between self-reported use and urine confirmatory drug test results was 92.5 percent for 7-day self-reported use and 93.0 percent for 3-day self-reported use. Again, the majority of these were individuals who reported no use and tested negative (i.e., 86.0 percent for 7-day and 86.9 percent for 3-day). In the 3-day period, 5.2 percent reported no use and tested positive, and 1.7 percent reported use and did not test positive.

- Only 1.4 percent of the sample tested positive for cocaine in their urine, and 0.9 percent reported cocaine use in the past 30 days. Also, 0.6 percent reported use in the past 30 days and tested negative, and 1.1 percent reported no use and tested positive.

- Comparison of the 7-day self-reports for cocaine with the urine drug test results showed 98.5 percent agreement, with 98.2 percent reporting no use and testing negative (0.3 percent reported use and tested positive). There was 98.6 percent agreement between the 3-day self-reports and urine drug test results: 98.4 percent reported no use in the past 3 days and tested negative, while another 0.2 percent reported use in the past 3 days and tested positive. Also, 0.1 percent reported use in the past 3 days and tested negative.

- Small sample sizes for opiates in terms of both self-reported use and positive urine drug test results precluded drawing conclusions about the validity of self-reported opiate use.

- Small sample sizes for stimulants in terms of both self-reported use and positive urine drug test results precluded drawing conclusions about the validity of self-reported stimulant use.

Conclusions

- The Validity Study demonstrated that it is possible to collect urine and hair specimens with a high response rate from persons aged 12 to 25 in a household survey environment.

- The results of tests conducted on hair collected in this study could not be used to compare with self-reports because there were technical and statistical problems related to the hair tests and unresolved issues concerning the interpretation of the analytical results.

- Most youths aged 12 to 17 and young adults aged 18 to 25 reported their recent drug use accurately. However, there were some reporting differences in either direction—with some not reporting use and testing positive, and some reporting use and testing negative.

- Biological drug test results can be used as objective markers of drug use to verify self-reports. Researchers employing drug tests in epidemiological studies must be knowledgeable of the performance characteristics of analytical procedures used for the drug tests (e.g., capabilities of the test methods and validation of procedures used by the testing laboratory), as well as the pharmacology of the drugs tested to enable acceptable study design and correct interpretation of the drug test results in the different biological specimen matrices.

1. Introduction

1.1. Background

This report examines the methodology and results from a study on the validity of self-reported drug use. Referred to as the Validity Study in this report, it was conducted in 2000 and 2001 as a methodological substudy of the National Household Survey on Drug Abuse (NHSDA—since 2002 called the National Survey on Drug Use and Health or NSDUH). The NHSDA (and now the NSDUH) is the primary source of statistical information on the use of illegal drugs by the U.S. population. The survey is sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services, and is planned and managed by SAMHSA's Office of Applied Studies (OAS). Data collection is conducted under contract with RTI International, Research Triangle Park, North Carolina.[3] The Validity Study was modeled on the NHSDA in collecting self-report data on drug use.

A series of methodological experiments has improved the NHSDA and NSDUH methodology to enhance the reporting of sensitive behaviors. Starting in 1999, the questionnaire has been delivered via computer-assisted interviewing (CAI), with the sensitive drug questions administered via audio computer-assisted self-interviewing (ACASI). Survey introductions and written material have been designed and improved over more than a quarter century to assuage respondents' concerns about how the confidentiality of the information they provide is protected. Therefore, the Validity Study was coupled with a survey that has already addressed issues of survey integrity and reliability.

The Validity Study sought to establish how valid self-reported drug use is in the NHSDA and, more generally, to provide information on the validity of survey research methods that utilize self-report. Surveys using self-report measures were developed initially as an alternative to indicator data such as that obtained from clinical, police, or court records. Those data are important to help assess the prevalence of illicit drug use. However, if drug-using behaviors are examined only among those seeking substance abuse treatment or those in the criminal justice system, we miss many who use drugs, but whose drug use does not lead them to treatment or cause them to come to the attention of the criminal justice system. To define more fully the nature and breadth of drug use among the population requires using unobtrusive measures typically employed in survey research methods. There is no other way to gather meaningful information on drug use among the general population than by surveying people about their behaviors.

However, surveying even a representative sample may still result in an inaccurate assessment due to misreporting. Respondents may misreport due to memory lapse, poor comprehension, or desire to deceive—either to conceal undesirable attributes or to exaggerate desirable ones (Schuman & Presser, 1981). The possibility that respondents "edit" their survey answers to what they perceive is the socially desirable response has long been a concern in survey research (Crowne & Marlowe, 1960; Schuman & Presser, 1981), and the concern is

[3] RTI International is a trade name of Research Triangle Institute.

particularly salient for drug use (Babor, Brown, & DelBoca, 1990; Rouse, Kozel, & Richards, 1985).

Using current technology, it is possible to determine drug use with reasonable accuracy by testing specimens of the respondents' urine and, more controversially, their hair. In this study, urine and hair specimens were analyzed for the presence of drugs. Urine has obvious advantages, such as its wide acceptability and the availability of laboratories for analysis. Its accuracy is generally accepted. Urine testing generally detects drug use in the past 2 to 7 days, although marijuana metabolites have been detected for up to a month in heavy users. The study also collected and analyzed hair, a less widely implemented and accepted specimen matrix for drug testing, but one that has the potential to determine drug use over a longer period of time.

1.2. Summary of the Validity Study Methodology

The sampling and fieldwork for the Validity Study reproduced the NHSDA in 2000-2001 with great fidelity but with four systematic differences: (1) the Validity Study was representative of the coterminous U.S. population, which does not include Alaska and Hawaii; (2) the Validity Study interviews were not conducted in Spanish, creating a nonresponse category for those who only spoke Spanish; (3) the Validity Study was limited to persons aged 12 to 25 years; and (4) a maximum of one person was selected per sampled dwelling unit. With these exceptions, the Validity Study replicated the NHSDA sample. Also consistent with the NHSDA, the Validity Study utilized CAI methods, with the drug questions administered by ACASI methods.

The study team developed an abbreviated version of the NHSDA questionnaire (the core plus auxiliary questions) to obtain information on the frequency and recency of alcohol and drug use. Questions also were asked about the correlates and consequences of drug use. Standard demographic information (e.g., race, gender, income) was collected. After completing the **core** portion of the study following NHSDA protocols, respondents were introduced seamlessly to the validity portion of the study. Nothing differed between the NHSDA and Validity Study until after the core sections on drug use had been completed.

After the core questions, **follow-up** questions were asked corresponding to shorter time periods that are more comparable with the window of detection for the drugs of interest in urine. Those reporting past month use of cigarettes or cigars, marijuana, cocaine, heroin, prescription pain relievers (including opiates), or stimulants (including amphetamines) were asked whether that use occurred in the past 3 days. Respondents also were asked about use during the past 6 months for those who reported their last drug use was during the past 12 months.

For half the respondents, a persuasion experiment was embedded in the questionnaire after the follow-up questions. This **appeal** was designed to increase the respondents' willingness to provide valid responses. They were apprised of the scientific need to measure accurately the prevalence of recent drug use in the United States. Respondents were assured again of the confidentiality of the survey and the data they provided, as well as the anonymity of their survey responses. Half of the respondents did not receive the appeal, but rather were told they would be asked a few questions about what they thought about the study and were requested to "please answer these questions as honestly as you can."

All respondents then were asked to complete a series of questions about the accuracy of the survey responses they just provided, as well as how they believed "most people" would respond to the questions. These questions are referred to as **debriefing** questions. This series of questions preceded the **repeat** questions, which again asked respondents about the recency of their use of cigarettes and cigars, marijuana, cocaine, stimulants, and opiates (e.g., past 7 days, past 3 days). Also included in the repeat section were questions about passive exposure to drugs (i.e., how often they had been around people smoking cigarettes or any other tobacco product, marijuana/hashish, cocaine or crack, heroin, or methamphetamine). At the end of the formal interview, consent procedures for collecting biological specimens were introduced. Respondents were offered an incentive of $50 ($25 each for a urine and hair specimen), with the explanation that the specimens would be tested for the presence of drugs. Before the urine and hair specimens were collected, a few interviewer-administered questions were asked to obtain data on possible modifications to the hair, such as dyeing or permanent waves, because some studies have shown that hair treatments can decrease drug concentrations in hair.

For urine collections, the interviewer provided a plastic container to the respondent for the specimen. Collections were not observed. The interviewer noted the temperature of the specimen at the time of collection.

For hair collections, the interviewer cut a portion of hair close to the scalp on the posterior vertex region of the head in a way that was least cosmetically noticeable. For the first quarter of the study period, only one specimen was collected from the respondent. Thereafter, in an attempt to obtain a sufficient amount for testing, interviewers were instructed to collect two hair specimens from respondents. The two specimens were combined, with the proximal ends aligned, and packaged together for shipment to the laboratory. The laboratory segmented the hair for two separate analyses—the first 1.3 centimeters (cm) measured from the proximal end were analyzed for past month drug use, and the remaining specimen (up to 6.5 cm in length) was analyzed for drug use occurring in the 5 months prior to the past month.

Urine and hair specimens were analyzed by U.S. Drug Testing Laboratory for the presence of the following drugs or metabolites: marijuana metabolite (delta-9-tetrahydrocannabinol carboxylic acid, carboxy-THC), cocaine metabolite (benzoylecgonine or BZE), amphetamines (amphetamine and methamphetamine), and opiates (codeine and morphine). The testing laboratory used two different immunoassay methods for the urine drug screening test during the study: fluorescence polarization immunoassay (FPIA) was used from January 2000 to May 2001, and enzyme-multiplied immunoassay technique (EMIT) was used from June 2001 to December 2001. Hair specimens were screened for the same drug classes using a different immunoassay test (i.e., enzyme-linked immunosorbent assay, ELISA). Specimens screened positive (i.e., with results at or above the Validity Study cutoff concentrations) were tested using gas chromatography/mass spectrometry (GC/MS) for targeted drug analytes as the confirmatory test. Urine specimens also were screened for cotinine, the principal metabolite of nicotine, using ELISA. No confirmatory testing was performed for cotinine.

1.3. Urine and Hair Drug Testing

Over the past several decades, sophisticated technology-based methods have been developed to analyze drug metabolites in bodily fluids or tissues. The development of urine testing technology for drugs of abuse began in the late 1960s and has continued over the subsequent decades. Today, a variety of available test methods have been proven to produce scientifically valid and forensically acceptable evidence of recent drug use. Hair and oral fluid specimens also are used in many drug testing programs. Drugs also can be detected in other biological specimens (e.g., blood, perspiration, nails, semen, and meconium). Each biological specimen is unique and offers a somewhat different pattern of information regarding drug use over time. Also, each specimen has particular strengths and weaknesses regarding the type of information that may be obtained from drug testing.

Testing methods may be designated as either screening or confirmatory tests. Screening tests are used to provide a preliminary result to indicate the drug(s) and/or metabolites that may be present in a specimen. Confirmatory tests are more specific and are used to identify drug analytes definitively.

1.3.1 Urinalysis

When designing a validity study and interpreting urine drug test results for comparison with self-reports, an important consideration is the window of detection for drugs and their metabolites in urine. The window of detection is affected by both physiological and analytical factors. Physiological factors affecting drug pharmacokinetics and the excretion of the drug and metabolites into the urine include the drug dose, ingestion method, level of use, drug chemical structure, fluid intake, and urine pH, as well as individual metabolism, diet, and disease states. Analytical factors include the performance characteristics of the test methods used for screening and confirmation (e.g., accuracy, sensitivity, specificity), the testing laboratory's procedures for sample preparation and analysis, and the cutoff concentrations used to distinguish positive and negative specimens. The procedures for sample handling also affect drug recovery in urine. For example, the recovery of some drugs in urine is affected by specimen changes in pH over time and the degree of mixing of thawed specimens after freezing. In addition, recovery of marijuana analytes may be affected by the type of specimen container used.

For the majority of illicit drugs, a single use occasion should be detected in the urine for 2 to 7 days. Frequent, multiple dosing over an extended period may extend the detection period. Normally, specimens collected within 6 hours of drug use contain the highest concentrations of the parent drugs and their metabolites. Drug excretion in urine occurs at an exponential rate, with the majority of the drug dose eliminated within 2 days (Cone, 1997). Typically, cocaine metabolites (e.g., BZE) and opiates are detectable for 1 to 3 days in the urine. Amphetamines are generally detectable for 2 to 4 days. Marijuana metabolites, however, may be detected for up to 30 days in heavy users (Cone, 1997).

1.3.2 Hair Analysis

Hair is a relatively new specimen matrix for drugs of abuse testing. Hair testing is currently an option for epidemiological studies on drug use as well as in other settings (e.g.,

private-sector workplace programs, criminal justice applications, court systems). The drug testing industry continues to develop procedures and technologies for hair testing, and researchers continue to study hair as a drug testing specimen matrix. However, important issues may limit the applicability of hair testing for some purposes.

Hair as a drug testing specimen has characteristics of significant benefit for some uses, including epidemiology studies and workplace drug testing programs. Advantages of hair as a specimen matrix include its ease of collection, transport, and storage. In addition, hair is less likely to transmit bio-organisms than urine, and it is less susceptible to tampering than urine (by donors attempting to alter drug test results). After hair has undergone sample preparation steps, the same analytical methods commonly used for screening and confirmatory urine tests (i.e., immunoassay and GC/MS, respectively) can be used for most drugs of abuse testing in hair. Perhaps the most important advantage in the context of validating self-report in epidemiology studies is that hair has a longer drug detection window than urine. While many drugs of abuse are detected in urine only up to 2 to 7 days after use, it may be possible to detect drugs in hair for up to several months or even years after use. Some research supports that hair maintains a chronological record of drug use, so the detection window depends mainly on the length of the hair tested. Due to the differing windows of detection, urine analysis and hair analysis are generally considered complementary tests for establishing drug use.

Although these characteristics make hair an attractive option for drug testing, researchers must be cognizant of the limitations imposed by the testing methodology and the biological specimen (Cone, 1997; Harrison, 1997; Kidwell & Blank, 1995; Musshoff & Madea, 2006). Although the technology of hair testing has progressed since this study was conducted, several issues concerning testing of drugs in hair continue to be investigated.

The ability of hair testing to distinguish drug use from external contamination remains controversial. Currently, it is still unclear how drugs enter the hair. This creates concern about detected drugs in a hair test arising from external contamination from drug-dust particles, smoke, vapor, or drug solutions. Several studies have found cocaine in the hair of children, suggesting that contamination is an important consideration (Randall, 1992; Rosenberg, Marino, Meert, & Kauffman, 1995; Smith, Kidwell, & Cook, 1994). There is conflicting information from studies on the efficacy of decontamination procedures in removing externally deposited drug (Romano, Barbera, Spadaro, & Valenti, 2003; Welch, Sniegoski, Allgood, & Habram, 1993). Although some contend that procedures, including extensive washing, can distinguish drug present from external contamination (Cairns, Hill, Schaffer, & Thistle, 2004; Schaffer, Hill, & Cairns, 2005; Schaffer, Wang, & Irving, 2002), others have demonstrated that externally deposited drug remains even after extensive washing (Romano, Barbera, & Lombardo, 2001; Stout, Ropero-Miller, Baylor, & Mitchell, 2006; Wang & Cone, 1995).

Further investigation is needed to determine whether criteria (e.g., analyte cutoff, metabolite-to-parent drug ratio) could be established to effectively distinguish a drug user from an individual who has been exposed to drug in their environment. Differentiating an externally deposited drug from a consumed drug is not definitive in hair tests for drugs in which the analyte is the parent drug (e.g., cocaine, amphetamine, methamphetamine, morphine, codeine) or a compound for which there are nonmetabolic pathways for their formation (e.g., cocaine metabolites: benzoylecgonine, cocaethylene, and norcocaine). External contamination is not an

issue in marijuana testing because the analyte detected in the confirmatory test, carboxy-THC, is a true metabolite and will only be present in the hair following marijuana ingestion. However, carboxy-THC has a low rate of incorporation into hair (Nakahara, Takahashi, & Kikura, 1995). Therefore, it is necessary to use a test method more sensitive than GC/MS (e.g., GC/MS/MS, GC/GC/MS, liquid chromatography [LC]/MS/MS) to detect this analyte at the concentrations found in hair.

Drug dose and time relationships for drugs in hair are not clear. Some research has suggested that the amount of drugs in the hair is proportional to the amount of use (DuPont & Baumgartner, 1995). Studies with labeled cocaine have found only a limited dose and time relationship (Cone, 1994a; Henderson, Harkey, & Jones, 1993; Henderson, Harkey, Zhou, Jones, & Jacob, 1996; Kidwell & Blank, 1995), although heavier drug users tend to have higher concentrations in their hair (Henderson et al., 1996; Kidwell & Blank, 1995; Mieczkowski, Landress, Newel, & Coletti, 1991b; Welch & Sniegoski, 1995). One study used deuterium-labeled cocaine that could be detected separately from street cocaine use (Henderson et al., 1996). Different doses and dosing regimens of cocaine were given to 25 moderate cocaine users for up to 10 months. The study found, after a single dose of cocaine, that the drug was not always confined to a discrete area adjacent to the root. In some subjects, the drug was distributed over multiple segments extending far from the root. There was also considerable variability in the amount of drug incorporated into hair and the time until the drug first appeared in hair.

Some studies support that segmental hair analysis can provide a chronological record of drug use (Beumer, Bosman, & Maes, 2001; Kronstrand, Nystrom, Josefsson, & Hodgins, 2002; Pichini et al., 2006). However, others found high variability in segmental analysis results (Charles et al., 2003; Clauwaert, Van Bocxlaer, Lambert, & De Leenheer, 2000; Wilkins et al., 1999). Researchers should use caution when interpreting drug test results from hair segments.

There is evidence that hair color and type may bias some results in hair testing. Studies have shown an association between different hair color or type and drug incorporation into the hair, with coarse dark hair retaining more drug than other hair types (Cone, 1994b; Henderson et al., 1993; Holl, Grabert, Heinze, & Debatin, 1998; Kidwell & Blank, 1994, 1995). Researchers have shown that basic drugs may be present in higher concentrations in dark hair compared with lighter hair (Borges, Wilkins, & Rollins, 2001; Gygi, Joseph, Cone, Wilkins, & Rollins, 1996; Kronstrand, Forstberg-Peterson, Kagedal, Ahlner, & Larson, 1999; Potsch, Skopp, & Moeller, 1997; Rollins et al., 2003; Scheidweiler, Cone, Moolchan, & Huestis, 2005; Wilkins, Haughey, Krueger, & Rollins, 1995). However, the limited number of population studies in the literature do not indicate a significant association between hair color and drug analyte (Hoffman, 1999; Kelly, Mieczkowski, Sweeney, & Bourland, 2000; Mieczkowski & Newel, 2000).

Hair treatments can alter drug concentrations in hair. Bleaching, dyeing, and permanent waves have been shown to decrease the concentration of drugs in hair (Harkey & Henderson, 1989; Henderson et al., 1993; Jurado, Kintz, Menendez, & Repetto, 1997; Skopp, Potsch, & Moeller, 1997), although the percentage decrease is similar regardless of hair type (Kidwell & Blank, 1995).

1.4. Hair Testing Results for the Validity Study

Problems were encountered with hair testing in the Validity Study.

The majority of hair specimens collected had insufficient quantity for testing. After an eighth of the interviews had been conducted (the first quarter of 2000), the testing laboratory reported that most of the specimens collected thus far had insufficient quantity for testing. At that point, interviewers were instructed to begin taking two hair specimens (rather than one), each estimated to be the area of a pencil eraser. Even with the revised collection procedures, it appears that interviewers routinely were collecting fewer than the 50 strands of hair necessary for the laboratory to test for multiple drug classes in the Validity Study. Over half of the hair specimens collected (53.6 percent) did not have sufficient quantity to conduct a screening test for drugs in the first 1.3 cm of hair. Another 2.2 percent of the specimens had insufficient quantity for confirmation testing, so that a total of 55.8 percent of the specimens could not be analyzed to identify drug use in the past 30 days. Other researchers also have noted a problem of insufficient hair specimens in a household study of the general population (Fendrich, Johnson, Sudman, Wislar, & Spiehler, 1999a; Fendrich, Johnson, Wislar, & Sudman, 1999b). The authors raised concern over sample bias, noting key demographic differences (e.g., race, gender) between participants and nonparticipants. Although this was not evaluated in this Validity Study, it is a subject for future investigation.

There were too few positive hair test results in the Validity Study to allow a meaningful analysis of the findings. As noted above, because of insufficient quantity, not all hair specimens collected could be analyzed. In addition, there were significant problems with opiates and marijuana testing. The hair testing cutoffs used in the Validity Study were suggested by the testing laboratory. The cutoffs used for opiates and marijuana exceeded industry standards at that time. The industry standards were incorporated into SAMHSA's draft *Mandatory Guidelines for Federal Workplace Drug Testing Programs* (draft 1 published in April 2000; SAMHSA, 2000-2001). In 2001, the testing laboratory lowered the opiates screening cutoff to be consistent with industry standards. There were only eight specimens confirmed positive for opiates in the 2 years of the Validity Study. The testing laboratory also lowered the marijuana screening and confirmatory cutoffs in 2001. However, the marijuana screening cutoff was still 5 times higher than the industry standard, and the confirmatory cutoff was still 20 times higher. As a result, there were no positive hair specimens for marijuana in year 1 and only 18 positive in year 2.

Because of these problems and the unresolved issues with hair testing for drugs of abuse (discussed in Section 1.3.2), the hair testing results compared with self-reports from the Validity Study have not been included in this report. Hair testing results are presented in Appendix E.

1.5. Organization of the Report

This report is divided into nine chapters and seven appendices. Chapter 2 compares the Validity Study with the parent NHSDA study, including similarities in the self-reported prevalence of drug use in both surveys. This chapter also includes response rates for the urine specimens. Chapter 3 examines the responses to the debriefing intervention, including memory

and comprehension, embarrassment and confidentiality concerns, and truthfulness. This chapter also considers the respondent's privacy in completing the survey. The following chapters are drug-specific: Chapter 4 examines tobacco, Chapter 5 marijuana, Chapter 6 cocaine, Chapter 7 opiates, and Chapter 8 amphetamines. Chapter 9 provides a summary of the results and their implications.

Appendix A provides an overview of prior research on the validity of self-reports. It also provides a more complete review of the science of urine and hair testing. Appendix B provides more detailed information about the design of the Validity Study, including the questionnaires, collection procedures for the biological specimens, and incentive payments. The cutoff concentrations for determining positive specimens and laboratory selection and performance are discussed. Appendix C provides details about the sampling design, including respondent selection and the creation of variance estimation strata. Appendix D provides a comparison of the data from the Validity Study using different test cutoffs (i.e., those established by SAMHSA for Federal workplace drug testing). Appendix E discusses technical issues concerning hair drug test data. Appendix F provides more detailed tables from the Validity Study. The tables follow the order of presentation of materials in the chapters. Lastly, Appendix G provides definitions for key terms used in the report.

For each drug type, the report presents two-by-two tables of self-reports compared with urine drug test results. Although the data are weighted, raw counts (number of respondents) are presented in parentheses in the tables. This strategy was used because the NHSDA uses population weights, which in this case would include the estimated population of 12 to 25 year olds in the coterminous United States. The tables present the unweighted *n*'s, which are the number of respondents in respective categories. Due to rounding, the percentages in the tables do not always sum to 100 percent.

In the two-by-two contingency tables, concordance between self-report and biological specimen results is assessed by the use of two goodness-of-fit measures, chi square and kappa. The Cochran-Mantel-Haenszel (CMH) chi-square and kappa values in this report are based on weighted data. The kappa values were computed in SAS. The CMH chi-square values have been calculated using SUDAAN® (Shah, Barnwell, & Bieler, 1997). The SUDAAN® program controls for design effects in the multistage sample design. Chi-square values are influenced by sample size and are generally larger than kappa values. Kappa is a measure of the degree of agreement between two items, adjusting for any agreement occurring by chance. Perfect agreement is indicated by $\kappa = 1$, and chance agreement by $\kappa = 0$. A kappa value greater than 0.75 is considered "good." A kappa value of 0.40 to 0.75 indicates moderate agreement, and a kappa value less than 0.40 indicates poor agreement (Murphy, Durako, Muenz, & Wilson, 2000; Riley, Lu, & Taylor, 2000).

Researchers often use sensitivity and specificity to describe interview validity or the agreement between self-reports and drug tests. The results of the drug tests in these analyses are considered completely accurate and are the standard against which self-reports are judged. The most common method for interpreting the agreement of urinalysis and self-report is to focus on those with positive urine test results and determine the percentage accurately reporting their drug use. This is expressed as the "sensitivity," or the proportion positive on the drug test who report use. The corollary of sensitivity is "specificity," which refers to the proportion negative on the

drug test who deny use. These terms were first defined by Yerushalmy (1947) and result from a diagnostic test being compared with a known standard test. Sensitivity and specificity have been defined by researchers in various fields. In medical research, sensitivity of a test is the probability that a diagnostic test correctly classifies diseased subjects. Similarly, specificity is the ability to correctly identify subjects who do not have a specific disease.

Because this report is on the validity of self-report data on drug use, we examine both underreporters (those who report no use but test positive) and overreporters (those who report use but do not test positive).

Logistic regression models were developed to determine the correlates of overreporting and underreporting. Logistic regression allows other variables that may affect the relationship between self-reports and drug test results to be controlled. That is, we can assess their impact on the "validity of self-report" based on controlling for multiple potential covariates. After extensive bivariate and multivariate analyses, a set of variables was derived that was used consistently to examine under- and overreporting. These are gender, race (white, black, other), region of the country, religiosity, the privacy of the interview, whether the respondent received the experimental appeal to be truthful, difficulty remembering or understanding drug questions, truthfulness, friends smoking, and passive exposure to the drug.

2. Validity Study Methodology

2.1. Response Rates in the Validity Study and the NHSDA

The Validity Study was fielded in 2000 and 2001, and comparisons are with the corresponding 2000 and 2001 National Household Surveys on Drug Abuse (NHSDAs). Because Alaska and Hawaii were not included in the Validity Study, the NHSDA estimates were recomputed excluding those two States. Table 2.1 shows that 27,463 sample dwelling units (SDUs)—households or group quarters—were selected to be screened in the Validity Study, and 409,342 SDUs were selected for screening in the NHSDA. Approximately 15 times as many households were screened in the NHSDA, but about 19 times as many persons were selected in the NHSDA than in the Validity Study. One eligible respondent was found in every 3.5 SDUs in the NHSDA compared with 1 in 4.6 SDUs in the Validity Study. With the eligible age range limited to 12 to 25 years in the Validity Study, relatively more SDUs had to be screened in order to yield the desired number of respondents.

Table 2.1 shows that the screening response rate of the Validity Study is similar to that of the NHSDA. The NHSDA had a slightly higher rate of 92.3 percent, while the rate in the Validity Study was 89.9 percent. In addition to refusals, the screening nonresponse rate includes households where (1) no one was home after repeated visits, (2) residents were physically or mentally incapable of responding, (3) there was a language barrier, or (4) the interviewer was denied access to the property. The lower response rate for the Validity Study may be due to fewer trained Validity Study interviewers being available to make callbacks or assist in refusal conversion activities. The exclusion of selected persons who speak only Spanish also contributed to the lower response rate.

Table 2.1 Comparison of Sampling and Interview Data between Validity Study and NHSDA

Sampling and Interview Data	Validity Study	NHSDA[1]
Dwelling Unit Level		
Selected Dwelling Units	27,463	409,342
Eligible Dwelling Units	23,724	346,138
Screening Completed	21,334	319,771
Weighted Screening Response Rate	89.9%	92.3%
Person Level		
Selected Persons	5,985	116,200
Total Number of Respondents	4,465	91,706[2]
Weighted Interview Response Rate	74.3%	79.1%[3]

[1] NHSDA screening data include the years 2000 and 2001 and exclude persons in Hawaii and Alaska. Dwelling unit–level data for NHSDA represent all ages, and person-level data represent ages 12 to 25.

[2] The total number of respondents shown in this table is the NHSDA sample size of persons aged 12 to 25 based on final interview age.

[3] The weighted interview response rate is computed using the number of respondents aged 12 to 25 based on the screening age, which differs from the final NHSDA sample size of respondents aged 12 to 25 reported within this table.

The Validity Study's selection procedures resulted in the selection of 5,985 individuals for an interview. Among those aged 12 to 25, the NHSDA selected 116,200 individuals for an interview. Among 12 to 25 year olds, the Validity Study interview response rate of 74.3 percent was lower than the NHSDA rate of 79.1 percent. The nonresponse rate in the Validity Study included 8.4 percent (parental refusal) for youths aged 12 to 17 compared with 8.6 percent from the NHSDA. There were differences in parental consent procedures because at the point of biological specimen collection in the Validity Study, parental consent had to again be obtained. Unlike the NHSDA, the Validity Study interview was not translated into Spanish, thereby affecting the response rate. Screening procedures in the Validity Study could be completed in Spanish, but not the interview. This resulted in a 4.0 percent decrease in the Validity Study interview response rate. Otherwise, the NHSDA and Validity Study protocols were conducted in the same manner, and the response rates between the studies should be similar. However, interviewers were not able to contact 8.0 percent of the selected Validity Study respondents compared with 5.5 percent in the NHSDA. Other types of nonresponse (e.g., refusal, breakoff) totaled 11.0 percent in the Validity Study and 10.8 percent in the NHSDA.

2.2. Biological Specimen Response Rates

Only persons completing the Validity Study interview were given the opportunity to provide hair and urine specimens. Of those, 89.4 percent provided one or both biological specimens. This includes 80.5 percent who provided both urine and hair specimens, 4.7 percent providing only urine specimens, and 4.3 percent providing only hair specimens.

The loss of urine specimens due to insufficient quantity was minimal. Among the urine specimens that were obtained, only 0.9 percent had insufficient quantity for the laboratory to conduct the urine screening tests for marijuana, cocaine, amphetamines, and opiates. Another 0.1 to 0.4 percent had insufficient quantity to conduct confirmatory testing for these drugs. For tobacco, 1.0 percent of specimens had insufficient quantity for screening tests, and there were no confirmation tests.

As noted in Section 1.4 in Chapter 1, unlike the urine specimens, there were many specimens with insufficient quantity for hair testing.

2.3. Comparisons of Drug Use Generated by the Validity Study and the NHSDA

Table 2.2 shows prevalence rates for tobacco, marijuana, cocaine, stimulant, and opiate use generated by the Validity Study in comparison with the NHSDA. The prevalence estimates are presented for substance use in the lifetime, the past year, and the past 30 days. Rates were generally very similar, although there were a few differences. Greater lifetime marijuana prevalence was reported in the Validity Study than in the NHSDA and was attributable to greater lifetime prevalence among 18 to 25 year olds in the Validity Study. The Validity Study also had higher past year and past month marijuana use rates than the NHSDA, but the differences were not significant. The Validity Study also had slightly higher prevalence rates for tobacco use, but none of the differences reached statistical significance. The Validity Study also had higher past year and past month opiate use rates than the NHSDA.

Table 2.2 Estimates for Five Substances in the Validity Study Compared with the 2000-2001 NHSDA: Percentages

Substance	12 to 17 Years		18 to 25 Years		Total 12 to 25 Years	
	Validity Study (*n* = 2,303)	NHSDA (*n* = 47,616)	Validity Study (*n* = 2,162)	NHSDA (*n* = 44,090)	Validity Study (*n* = 4,465)	NHSDA (*n* = 91,706)
Tobacco[1]						
Lifetime	37.1	37.5	74.1	72.9	57.6	57.1
Past Year	24.2	23.9	54.6	52.6	41.1	39.8
Past Month	16.4	15.3	45.5	43.4	32.5	30.9
Marijuana						
Lifetime	20.6	19.0	51.3[a]	47.8	37.6[b]	35.0
Past Year	15.2	14.3	26.2	25.2	21.3	20.3
Past Month	9.0	7.6	15.5	14.8	12.6	11.6
Cocaine						
Lifetime	2.3	2.3	10.7	11.9	6.9	7.6
Past Year	1.3	1.6	4.9	5.1	3.3	3.5
Past Month	0.5	0.5	1.3	1.6	0.9	1.1
Stimulants[2]						
Lifetime	3.9	3.9	9.4	8.6	6.9	6.5
Past Year	2.2	2.3	2.9	2.9	2.6	2.6
Past Month	0.9	0.7	1.0	1.0	1.0	0.9
Any Opiates[3]						
Lifetime	9.7	9.0	17.2	16.8	13.9	13.3
Past Year	7.4[a]	6.0	9.6	8.6	8.6[a]	7.5
Past Month	3.2	2.5	4.1	3.3	3.7[a]	2.9

Note: Alaska and Hawaii are excluded.

[a] Difference between the 2000-2001 Validity Study estimate and the 2000-2001 NHSDA estimate is statistically significant at the 0.05 level.

[b] Difference between the 2000-2001 Validity Study estimate and the 2000-2001 NHSDA estimate is statistically significant at the 0.01 level.

[1] Lifetime use and past month use include users who reported use of any tobacco product. Past year use includes users who reported use of cigarettes, chewing tobacco, snuff, and cigars in the past year (no data available for use of pipes in the past year).

[2] Any stimulant use consists of the use of illicit stimulants (including amphetamine and methamphetamine), the nonmedical use of prescription stimulants (including amphetamine and methamphetamine), and the medical use of prescription diet pills (including amphetamine).

[3] Any opiate use consists of the use of heroin or the use of prescription pain relievers (including codeine and morphine).

Unfortunately, there is no simple explanation for the higher marijuana and opiate prevalence rates in the Validity Study. Because the studies did not poststratify or weight on every possible variable, it is likely that the differences in prevalence rates are due to the random luck of the draw in selecting the samples. The similarity in the prevalence rates is more pronounced than any differences. There was reasonable concordance in most of the prevalence rates, suggesting

the Validity Study results can be generalized to the 12- to 25-year-old population in the coterminous United States.

It is important to point out that, unless otherwise noted, these and all remaining estimates in this report are based on a sample that was weighted to represent the civilian, noninstitutionalized population between the ages of 12 and 25 in the coterminous United States. Both the NHSDA and Validity Study weighted the sample for factors known to correlate with nonresponse. Missing data were minimal in both the NHSDA and Validity Study, as the computer-assisted interviewing (CAI) instrument has built-in prompts and assurances to encourage response to all questions. A predictive mean neighborhood imputation method was used to calculate values for some missing data. This method uses the values of the previous respondent sorted according to responses on a number of related variables.

3. Debriefing Intervention

The debriefing intervention was specific to the Validity Study and was not included in the 2000 or 2001 National Household Survey on Drug Abuse (NHSDA).[4] The debriefing questions were administered using the audio computer-assisted self-interviewing (ACASI) method and were placed after the core drug questions. They occurred after the core demographic section in which the interviewer administered questions on household composition using computer-assisted personal interviewing (CAPI) methods. The interviewer then turned the computer back to the respondent and asked the respondent to put on his or her headphones and listen carefully to the next section. At this point, about half of the respondents (n = 2,197) received the persuasion experiment, which was an appeal designed to increase their willingness to respond accurately to drug questions. Those not receiving the appeal (n = 2,268) were told that they would be asked a few questions about their opinions concerning the Validity Study. This was followed by the debriefing questions, then the repeat questions about the drugs.

All respondents completed the debriefing questions, which asked how they thought "most people" would respond to providing the information on their drug use requested in the questionnaire. The same series of questions then asked the respondents' own reaction to answering the drug-related questions. The following tables present information from each of the debriefing questions showing how respondents thought "most people" would react, alongside their reports of their "own experience."

3.1. Memory and Comprehension

Table 3.1 includes the number of respondents and the prevalence (weighted percentage) of respondents who thought most people would have difficulty understanding or remembering the drug-related information requested in the questionnaire. The table also includes the respondents' reports of their own difficulty understanding and remembering the types of drug-related information requested. Approximately 1 in 20 (4.8 percent) thought "most people" would have a "lot of difficulty" understanding the drug-related questions, but less than half as many, 2.2 percent, said they had a lot of difficulty understanding. Nearly three quarters (72.5 percent) said they had "no difficulty" understanding the drug-related questions in the study compared with 42.1 percent who thought most people would have no difficulty understanding. (Note that percentages may not equal 100 percent due to rounding.)

Respondents thought "most people" would have more difficulty comprehending questions than themselves. A quarter of those interviewed (25.0 percent) thought that most people would have "no difficulty" remembering the types of drug-related information asked about, but over twice as many (57.7 percent) said they had no difficulty remembering. Another 25.0 percent said they had "just a little difficulty" remembering. Less than 10 percent of respondents reported they had "some" or a "lot of difficulty" understanding the drug-related questions, but about 17 percent had some to a lot of difficulty remembering the types of drug-related information requested in the questionnaire. This 17 percent translates to about one in six

[4] The debriefing questions were developed by Timothy Johnson and Michael Fendrich at the University of Illinois at Chicago, who served as consultants to the Validity Study.

reporting they had some or a lot of difficulty remembering drug-related information. Further, it appears that the respondents had much more faith in themselves remembering and understanding than they did in "most people."

Table 3.1 *Debriefing* Questions: Reported Difficulty Understanding and Remembering Drug-Related Information from the Point of View of Most People or from the Respondent's Own Experience: Percentages

	"Most People"		"Own Experience"	
***n* = 4,465**	**Difficulty Understanding**	**Difficulty Remembering**	**Difficulty Understanding**	**Difficulty Remembering**
Lot of Difficulty	4.8	7.4	2.2	4.3
Some Difficulty	19.7	31.9	7.3	12.5
Just a Little Difficulty	32.5	35.0	17.6	25.0
No Difficulty	42.1	25.0	72.5	57.7
Don't Know/Refusal	0.8	0.8	0.4	0.4

Table 3.2 provides the weighted percentage of respondents who made a "best guess" when answering questions about drug use compared with what they thought was true for most people. Half (50.4 percent) of the respondents said they "never" had to make a best guess. However, only 13.5 percent felt most people never had to make a best guess. The modal response for "most people" having to make a best guess was "sometimes" (38.3 percent). Very few of the respondents, only 6.1 percent, said they "frequently" had to make a best guess when answering the drug-related questions. The respondents expected twice as many of "most people" (12.2 percent) "frequently" had to make a best guess. Overall, respondents expected most people to make best guesses at least "sometimes" (50.5 percent), while only 18.3 percent of respondents said that they had to resort to a best guess at least sometimes. By some standards, these are reasonably high rates of reports of making best guesses, with about one in six saying they "sometimes" or "frequently" made a best guess when answering the drug-related questions.

Table 3.3 shows responses to two debriefing questions that only asked for the respondent's own experiences. When asked about clearness of their memory, just over half, or 57.6 percent, indicated their memory was "very clear." Another 28.9 percent indicated it was "mostly clear," allowing for some degree of error. Only 2.4 percent said their memory was "not at all clear." Respondents also were asked how certain they were about the accuracy of their answers to the drug-related questions. About 71.4 percent said they were "very certain," and 23.1 percent said they were "mostly certain." Over a quarter of the respondents reported uncertainty regarding the accuracy of their responses to the questions asking about their drug use.

Table 3.2 *Debriefing* Questions: Reported Frequency of Making a "Best Guess" When Answering Drug-Related Questions: Percentages

	"Best Guess" Answers	
***n* = 4,465**	**"Most People"**	**"Own Experience"**
Frequently	12.2	6.1
Sometimes	38.3	12.2
Seldom	34.6	30.7
Never	13.5	50.4
Don't Know/Refusal	1.3	0.7

Table 3.3 *Debriefing* Questions: Respondent's Clearness of Memory and the Degree to Which the Respondent Was Certain of the Accuracy of His or Her Answers: Percentages

	"Own Experience"		**"Own Experience"**
***n* = 4,465**	**Clearness of Memory**	***n* = 4,465**	**Certain of Accuracy of Answers**
Not at All Clear	2.4	**Not at All Certain**	1.4
Somewhat Very Clear	10.6	**Somewhat Certain**	3.8
Mostly Clear	28.9	**Mostly Certain**	23.1
Very Clear	57.6	**Very Certain**	71.4
Don't Know/Refusal	0.5	**Don't Know/Refusal**	0.4

3.2. Embarrassment and Confidentiality Concerns

Table 3.4 shows the number and weighted percentage of respondents who found the drug-related questions embarrassing to answer, as well as their assessment of most people's level of embarrassment. It also shows the respondents' concern about others having access to their answers and how that compares with the concern they think most people would have about confidentiality. Embarrassment was not a big concern. However, there were surprisingly high levels of concern expressed that others would have access to their answers. Only 2.5 percent indicated they were "very" embarrassed to answer the questions, and 74.8 percent said they were "not at all" embarrassed. However, that pattern is different when respondents were asked how they thought most people would feel. Only 24.0 percent felt most people would not be embarrassed at all to answer the drug-related questions, with 9.2 percent saying most people would find it very embarrassing to answer the questions and 31.1 percent saying most people would find it somewhat embarrassing.

Table 3.4 *Debriefing* Questions: Respondent's Level of Concern and Embarrassment When Answering Drug-Related Questions from the Point of View of Most People or from His or Her Own Experience: Percentages

	"Most People"		"Own Experience"	
***n* = 4,465**	**Embarrassing to Answer**	**Concern Others Would Have Access to Answers**	**Embarrassing to Answer**	**Concern Others Would Have Access to Answers**
Very	9.2	27.9	2.5	12.4
Somewhat	31.1	38.6	5.5	15.2
Not Very	34.8	22.0	16.8	22.6
Not at All	24.0	10.3	74.8	49.2
Don't Know/Refusal	0.8	1.2	0.4	0.5

The NHSDA goes to great lengths to assure respondents of the confidentiality of their responses and that no one will have access to their answers other than those working on the survey. Table 3.4 shows about half (49.2 percent) said they were "not at all" concerned that others would have access to their answers. A much smaller proportion (10.3 percent) felt the same was true for most people. Overall, about 72 percent of the respondents indicated they were "not very" or "not at all" concerned about others having access to their questionnaire responses, but only about a third believed most people would feel the same way. There is a clear pattern of respondents suggesting that they have fewer concerns than they feel most people would. It is encouraging to learn that the majority of respondents believed the assurances of confidentiality provided by the survey, although 12.4 percent were still very concerned and an additional 15.2 percent were somewhat concerned that others would have access to their survey responses. This must surely affect their willingness to disclose sensitive information truthfully.

3.3. Truthfulness

Perhaps the biggest contrast in "own experience" and beliefs about "most people" were in the questions asking about the truthfulness of answers. Table 3.5 shows that 90.0 percent of respondents indicated they were "completely" truthful when responding to the drug-related questions. Another 7.9 percent said they were "mostly" truthful, which accounts for about 98 percent of the respondents. Yet only 15.9 percent believed most people were completely truthful when answering the drug-related survey questions, and 46.7 percent thought most people were mostly truthful. Truthfulness did not differ significantly by age. The fact that such a large percentage of respondents indicated they were mostly or completely truthful in responding to the drug-related questions is reassuring, although their lack of trust in their fellow respondents is an interesting comment on their perception of others. A cynic or cognitive psychologist might even suggest their "own experience" may be a socially desirable response and their response for "most people" might be closer to the truth.

Table 3.5 *Debriefing* Questions: Truthfulness of the Respondent and the Truthfulness of Most People from the Respondent's Point of View When Answering Drug-Related Questions: Percentages

	Truthfulness of Answers	
***n* = 4,465**	**"Most People"**	**"Own Experience"**
Not at All Truthful	5.8	0.6
Somewhat Truthful	30.6	1.0
Mostly Truthful	46.7	7.9
Completely	15.9	90.0
Don't Know/Refusal	0.9	0.4

Table 3.6 provides responses to a parallel question about the truthfulness of other respondents. It is the only question that asked perceptions about other people, but not the respondent's own experience. Approximately a third of the respondents (34.4 percent) thought that most people would answer the questions about the frequency of their drug use truthfully—or "about as often as they really did." Over half (57.5 percent) thought that most people would report using drugs "less often than they really did," and 6.7 percent thought that most people would report using "more often than they really did." This is an interesting thought, and perhaps surprising that 6.7 percent thought that most people would exaggerate their drug use. This percentage is likely higher among youths and young adults than among older adults. Less encouraging is the finding that most respondents thought that most people would report using drugs less often than they really did.

Table 3.6 *Debriefing* Questions: Respondent's Point of View of the Accuracy with Which Most People Report Their Frequency of Drug Use: Percentages

	"Most People"
***n* = 4,465**	**Accuracy of Reporting Drug Use**
Report Taking More Often Than Really Did	6.7
Report Taking Less Often Than Really Did	57.5
Report Taking about as Often as Really Did	34.4
Don't Know/Refusal	1.4

3.4. Privacy

Although not part of the debriefing intervention, another potential mitigator of the accuracy and truthfulness of self-reported drug use was the privacy in which the interview is conducted. Several methodological studies conducted in conjunction with the NHSDA have found that youths aged 12 to 17 were more affected by lack of privacy than adults (Turner, Lessler, & Devore, 1992). In the 2000 questionnaire, the Validity Study asked the interviewers to rate how private the interview was on a scale of 1 to 9. This was reduced to a scale of 1 to 5 in 2001. Due to the small number of cases in some of these categories, responses were divided into a three-level scale of totally private, privacy interrupted up to one third of the time with someone or others around, and privacy interrupted one third of the time or more. In the NHSDA and the

Validity Study, as would be expected, the parents(s) of younger children were more likely to be nearby. Once they understood their child was comfortable, and they had a level of comfort with the interview situation, they often left the room. Interviewers tried to conduct the interview with as much privacy as possible. Unfortunately, the use of a laptop oftentimes means the interviews were conducted on a kitchen or dining room table—busy rooms in a household. The computer-assisted interviewing (CAI) procedures were designed to maximize privacy because the respondents listened to the questionnaire via earphones and entered their answers directly using a keyboard. Yet family members coming into the room can interrupt concentration and also may give the respondent concern that a family member is looking over his or her shoulder, even if the person cannot hear the questions or answers.

Table 3.7 shows that more than three quarters (76.1 percent) of the interviews among the 18 to 25 age group were conducted under conditions of total privacy. Among the 12 to 17 year olds, this percentage was lower, 67.7 percent. The 18 to 25 year olds had fewer interruptions. So, overall, although it is reassuring that 72.3 percent of the interviews were conducted in total privacy, this means nearly 30 percent were conducted with less than total privacy. Females had slightly more privacy on average than males, although the differences were not significant.

Table 3.7 *Debriefing* Questions: Privacy of the Interview Based on Interviewer's Assessment, by Age Group: Percentages

	Privacy of Interview		
***n* = 4,465**	**12 to 17 Years**	**18 to 25 Years**	**Total 12 to 25 Years**
Totally Private	67.7	76.1	72.3
Not Private ≤1/3 of Interview	27.4	20.5	23.6
Not Private > 1/3 of Interview	4.7	3.1	3.8
Missing	0.3	0.3	0.3

3.5. Summary

In summarizing the results of the debriefing questions, the majority of respondents felt very certain about the accuracy of their answers to the drug-related questions. A robust 90 percent reported that they were completely truthful in answering the drug-related questions. Seventy-five percent said they were not at all embarrassed by answering the questions. Most felt their memory was good, and few reported any difficulty in understanding the drug-related questions. However, over a quarter expressed some concern that others would have access to their answers. This is worrisome as these respondents may be less likely to report sensitive information that they would not want known. Perhaps what is also surprising about this number is that most of the questions were delivered via ACASI procedures, which were developed to maximize privacy. But enhancing privacy may be only part of the concern expressed by the quarter of respondents who were concerned about others having access to their answers.

It also is surprising that, although respondents did not perceive much difficulty remembering, and indicated they were accurate and truthful in responding to the drug-related questions, they had much less faith in other people. The respondents did not rate "most people"

as having the same ability to comprehend questions, remember the requested information, or be as truthful as they were. Over half thought most people would report using drugs less often than they really did. Respondents expressed concern that others would have access to their answers, but over a quarter thought that most people would be very concerned about others having access to their answers. Perhaps these respondents were projecting their own feelings onto others. Most likely, the "truth" is somewhere in between what they recorded for their "own experience" and what they recorded for "most people" due to issues of social desirability.

4. Tobacco

4.1. Urine Testing for Tobacco

The Validity Study analyzed urine for the presence of cotinine, the principal metabolite of nicotine. There is general agreement among researchers that testing for cotinine in urine is an accurate means of assessing exposure to cigarette smoke, measuring nicotine intake, and, consequently, verifying self-report data (Haufroid & Lison, 1998). Some studies are based on semiquantitative immunoassay data (Bono et al., 1994; Pokorski, Chen, & Bertholf, 1994), and others report quantitative data derived from chromatography-based analysis (Jacob, Yu, Shulgin, & Benowitz, 1999; Jarvis, Tunstall-Pedoe, Feyerabend, Vesey, & Saloojee, 1987). Cotinine has a urinary half-life of 15 to 20 hours (Herzig, Callaway, Halliday, Naylor, & Benowitz, 1998; Jacob et al., 1999; Pokorski et al., 1994). Regular smokers maintain a higher steady-state level of cotinine in their bodies (Herzig et al., 1998).

Nicotine ingestion occurs through the actual use of tobacco products and from environmental tobacco smoke (ETS), or "secondhand smoke." More than 50 epidemiological studies have addressed the association between secondhand smoke exposure and the risk of lung cancer among lifetime nonsmokers. Also noted was that nearly 40 percent of working people who were nontobacco users reported secondhand smoke exposure in the workplace (U.S. Department of Health and Human Services [DHHS], 2006).[5] Researchers have found that nonsmokers who are exposed to ETS have detectable cotinine concentrations, and these concentrations increase with the amount of passive exposure (Bono et al., 1994; Hecht et al., 1993; Knight, Eliopoulos, Klein, Greenwald, & Koren, 1996; Wall, Johnson, Jacob, & Benowitz, 1988).

This Validity Study utilized an enzyme-linked immunosorbent assay (ELISA) manufactured by International Diagnostic Systems to screen urine specimens for cotinine. All cotinine results are semiquantitative because the immunoassay results reflect the combination of cross-reacting compounds and metabolites. The extent of the cross-reactivity for this ELISA test may not be comparable with other immunoassays reported in the literature. Additionally, without a confirmatory test (i.e., using a more specific and quantitative test method, such as gas chromatography/mass spectrometry [GC/MS]), specimens that produced positive results on the immunoassay may not have contained cotinine or another nicotine metabolite. Although it is a limitation that cotinine positive tests were not confirmed by a more specific chromatographic method, good confirmatory assays were not readily available and few interferences with the immunoassay method have been described. Some caution should be used in the interpretation of specimens in which cotinine was detected and the respondent reported no tobacco use or tobacco use more than 3 days prior to the collection of the specimen.

There is no consensus in the research literature on appropriate cutoff concentrations for the purpose of distinguishing smokers, nonsmokers exposed to ETS, and nonsmokers not exposed to ETS. One major difficulty is that urinary cotinine only gives information about recent tobacco smoke exposure. Therefore, for cotinine concentrations to give an accurate account,

[5] For further information, see http://www.surgeongeneral.gov/library/secondhandsmoke/.

tobacco consumption during the past 48 hour period should represent the normal pattern of use. Other research has shown that the presence of cotinine indicates exposure to cigarette smoke from a minimum of 2 to 3 days (Jacob et al., 1999; Vine et al., 1993) to a maximum of 7 days (Jarvis et al., 1987). Among researchers, inconsistent definitions of the term "smokers" (occasional, light, typical, active, regular) present another problem related to establishing cutoff concentrations. According to Benuck, Gidding, and Binns (2001), a frequency of more than seven cigarettes a day is necessary to produce concentrations greater than 100 nanograms per milliliter (ng/mL). They found that nonsmokers heavily exposed to ETS have a mean urine cotinine concentration of approximately 30 ng/mL.

Holl et al. (1998) investigated the disclosure of smoking status by 238 German adolescents and young adults with type 1 diabetes. In this study, urinary cotinine was measured by double-antibody radioimmunoassay. The researchers defined nonsmokers living in a smoke-free environment as having cotinine concentrations of less than 100 ng/mL. However, among the 166 patients classified as nonsmokers, 3 patients reported smoking between one and five cigarettes per day. Holl and colleagues classified passive or occasional smokers as those with urinary cotinine concentrations between 100 and 500 ng/mL. Among the 26 classified as passive or occasional smokers, 3 reported smoking on average one or two cigarettes per day. Regular smokers excreted more than 500 ng/mL of cotinine in their urine. Among the 46 classified as regular smokers, 12 did not admit to smoking at all, and 32 reported smoking between 1 and 20 cigarettes per day. Other studies have found that cotinine concentrations in urine increased as the number of cigarettes smoked increased (Knight et al., 1996; Vine et al., 1993).

As it is not possible to definitively establish a cutoff that will distinguish ETS exposure compared with tobacco use, an analysis of the distribution of semiquantitative values obtained in this study is presented in Section 4.6.

Because there is no clear consensus in the research literature on the proper cotinine cutoff concentration for determining whether a person has used tobacco, this made the choice of a cutoff concentration for the Validity Study especially difficult to determine. Some studies have suggested that cotinine urine concentrations of active smokers are generally greater than 75 ng/mL, but the study reported by Holl et al. (1998) utilized a concentration of 500 ng/mL with the finding that concentrations between 100 and 500 ng/mL may be consistent with infrequent smokers or those environmentally exposed. Because the Validity Study sample includes youths who may be occasional tobacco users, a cutoff concentration was needed that would identify "light" smokers as well. Moreover, it was also necessary to try to distinguish nonsmokers exposed to tobacco smoke from active tobacco users. As a result, the following criterion was used in the results reported here:

A cutoff concentration of 100 ng/mL was used to identify a specimen as positive for cotinine.

4.2. Self-Reported Tobacco Use: Comparison of Responses to Core Questions with Urinalysis

The urine test for cotinine cannot distinguish among different tobacco products. Therefore, comparisons of self-report and urinalysis results used the combined self-report measure of "any tobacco use." The core tobacco section asked separate questions about the use

of cigarettes, smokeless tobacco, cigars, and pipes. The answers to these questions were combined to produce the category of "any tobacco use." Table 4.1 presents information on the use of the various tobacco products and the overlap among tobacco products among persons aged 12 to 25. Most users of tobacco products in this age group smoked cigarettes. About two thirds smoked only cigarettes. One percent or less of the 12 to 17 year olds were exclusive users of smokeless tobacco, cigars, or pipe tobacco. These percentages were slightly higher among 18 to 25 year olds except for use of only pipe tobacco, which remained equally rare among both youths and young adults. Approximately one quarter of youths and young adults aged 12 to 25 used only one tobacco product in the past 30 days. Of those who used more than one tobacco product in the past 30 days, the most popular combination was cigarettes and cigars.

Table 4.1 Tobacco Use: Self-Reported Use of Various Tobacco Products in Past 30 Days Based on *Core* Questions, by Age Group: Percentages

Tobacco Product	Age Group in Years		
	12 to 17 (*n* = 2,303)	**18 to 25 (*n* = 2,162)**	**12 to 25 (*n* = 4,465)**
Cigarettes Only	10.4	31.0	21.8
Smokeless Tobacco Only[1]	1.0	1.4	1.2
Cigars Only	0.9	2.6	1.8
Pipe Only	0.1	0.2	0.1
Used Only 1 Tobacco Product	12.4	35.1	25.0
Used 2 Tobacco Products	2.9	8.3	5.9
Cigarettes and Cigars Only	2.2	6.2	4.4
Cigarettes and Smokeless Tobacco Only[1]	0.6	1.6	1.2
Cigars and Pipe Only	0.0	0.1	0.1
Used 3 or More Tobacco Products	1.1	2.1	1.6
Any Tobacco Use	16.4	45.5	32.5

Note 1: *Core* questions ask about drug use and replicate the NHSDA format.

Note 2: Percentages are based on those reporting use on *core* questions.

[1] Smokeless tobacco is defined as chewing tobacco or snuff.

Table 4.2 presents the results of self-reported use of tobacco products in the past 30 days for the core questions compared with the results of the urine screening test for cotinine. Findings in this table are based on data from those respondents with a valid urine specimen who responded to the core questions on their use of tobacco products in the past 30 days. Only respondents who reported lifetime use were prompted to respond to questions about their use in the past 30 days. In addition, Table 4.2 excludes data from respondents who reported using nicotine replacement products in the past 30 days to quit smoking. Percentages shown in Table 4.2 are estimates for persons aged 12 to 25 in the civilian, noninstitutionalized population of the coterminous United States as a whole.

The table shows the self-report no (-) and yes (+) along the vertical axis and the urinalysis results negative (-) and positive (+) along the horizontal axis. The table shows that 60.9 percent

of persons in this age group were estimated to report that they had not used any tobacco product in the past 30 days and would have tested negative (cell A); 5.8 percent were estimated to report that they had not used a tobacco product in the past 30 days but would have tested positive (cell B); and 9.6 percent were estimated to have a self-report of use of a tobacco product in the past 30 days and a negative urine test result (cell C). Finally, 23.7 percent were estimated to have a self-report of use of a tobacco product in the past 30 days and to have a positive test result (cell D). This table shows congruence between self-report and urinalysis to be 84.6 percent (cell A + cell D). Therefore, 15.4 percent (cell B + cell C) of persons aged 12 to 25 would be estimated to have discordant urinalysis and self-report results.

Table 4.2 Tobacco Use: Comparison of Responses to 30-Day Self-Report *Core* Questions and Urinalysis

(*n* = 3,759)

		Urinalysis (-)	**Urinalysis** (+)	
	A	(-)	(+)	**B**
30-Day Self-Report (Core Questions)	(-)	60.9%	5.8%	(2,602)
	(+)	9.6%	23.7%	(1,157)
	C	(2,674)	(1,085)	**D**

CMH	*df*	*P*	Sensitivity	Specificity	Kappa
565.86	1	<0.0001	0.804	0.864	0.643

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: *Core* questions ask about drug use and replicate the NHSDA format.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of any tobacco product in the core questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

The estimate in cell B is classified as "underreporting" because it is based on data from respondents who self-reported no tobacco use in the past 30 days but tested positive. The estimate in cell C is classified as "overreporting" because it is based on data from respondents who reported use but tested negative. Some of these latter respondents may not actually have overreported their tobacco use, however. In particular, past month users of tobacco products could have fallen into cell C if they did not use within the 2- to 7-day timeframe for which cotinine generally can be detected. In addition, some users may have used tobacco infrequently or at a light level in the past 30 days. Infrequent users may not maintain high enough concentrations of cotinine in the urine for the specimen to be positive at the cutoff used in the

Validity Study (i.e., 100 ng/mL for the screening test). Thus, an infrequent user's cotinine concentration could have fallen below the assay limit of detection at the time their specimen was collected. Light users may not produce high enough concentrations of cotinine to be detected or to test positive.

4.2.1 Assessing Concordance

In Table 4.2, the overall sensitivity rate is 0.80 (i.e., proportion of self-reports of past month tobacco use among those testing positive), and the specificity rate is 0.86 (i.e., proportion of self-reports of no tobacco use in the past month among those testing negative). Both numbers suggest reasonably high levels of agreement between self-report and the results of the urine test for cotinine.

The chi-square value and probability of less than 0.0001 in Table 4.2 indicate that the cell values are significantly different from one another. Perfect agreement is indicated by $\kappa = 1$, and chance agreement only by $\kappa = 0$. The kappa value in Table 4.2 of 0.64 indicates moderate agreement between the self-report and urinalysis results.

Taken together, these results from this two-by-two table suggest that most recent users of tobacco products would report use on the questionnaire. However, the sensitivity data suggest that about 20 percent of persons with a positive urine test result for cotinine might not report use in the past 30 days in a survey.

4.2.2 Other Information on Concordant and Discordant Cases

Among persons aged 12 to 25 who tested positive for cotinine in their urine, 80.4 percent reported use of some tobacco product in the past month (i.e., the sensitivity rate). In addition, some persons in this age group may be willing to report past use, but not in the past month: 3.7 percent of persons in this age group who tested positive reported using a tobacco product more than 1 month ago but in the past year, and 5.2 percent reported using tobacco products more than a year ago. Further, 10.8 percent of persons in this age group who tested positive indicated they had never used tobacco products. According to the best available research, cotinine should not be detected in urine following tobacco use that occurred more than 7 days in the past. This suggests that about half of persons aged 12 to 25 who would underreport their tobacco use in the survey (but who would have a positive urine test result) may not be willing to report that they have ever used.

Other possible explanations for the potential rate of underreporting in cell B are that the cutoff concentration of 100 ng/mL was too low or that ETS may be influencing cotinine concentrations; this latter issue is discussed further in Section 4.4.1. Also, the research literature suggests that the brand of cigarette and amount of tar it contains can have an impact on cotinine concentrations. Some types of chewing tobacco and snuff have been found to elevate concentrations. One study found that excretion rates for pipe smokers were higher than for cigarette and cigar smokers (Jacob et al., 1999). So, the impact of using various tobacco products may need to be considered. Also, the Validity Study did not confirm the immunoassay screening results. Nevertheless, the conclusion suggested from this research is that about one in five recent

tobacco users aged 12 to 25 may underreport their use on national surveys to varying degrees, including about 10 percent who would report that they never used any tobacco.

Those who self-reported use of a specific tobacco product in the core questions in the past 30 days were asked to report the number of days they had used in that period. Due to the variation in overlap of use, however, it was not possible to combine categories to generate information on the number of days that persons used any tobacco product in the past 30 days. Therefore, information on a specific tobacco product has to be interpreted in light of the fact that nearly a quarter of current tobacco users were users of more than one type of tobacco product.

Focusing on cigarette users specifically, only about half (49.6 percent) smoked cigarettes every day. This is undoubtedly due to the age of this population. In particular, youths aged 12 to 17 were much less likely to be daily smokers than young adults aged 18 to 25. In addition, approximately a quarter (27.0 percent) of smokeless tobacco users and 2.2 percent of cigar smokers were daily users. These findings present difficulties for cotinine testing because many youths and young adults were not regular tobacco users.

More than 70 percent of youths and young adults would be estimated to have negative test results for cotinine in their urine (cell A + cell C). However, nearly one out of seven persons in this age group who would have a negative test result also were estimated to have reported using tobacco products in the past month. As noted previously, some of these persons could have used in the past 30 days but earlier than the 7-day timeframe for which cotinine is maximally thought to be detectable.

This issue can be examined by looking at the follow-up questions, which ask about use in the past 3 days among those who reported past month tobacco use. Of persons aged 12 to 25 who were estimated to have reported use in the past 30 days and to test negative, 43.3 percent reported that use occurred in the past 3 days. This finding suggests that the cutoff concentration of 100 ng/mL is too high.

Another explanation suggested above is that these individuals may be using at very low levels (one or two cigarettes per day), such that they are not testing positive. Focusing on cigarette smokers, those who smoked one cigarette or less per day on average over the past month had a median cotinine concentration of 30 ng/mL, which is below the Validity Study's cutoff concentration for detection of tobacco use. Further, those who reported smoking cigarettes between 1 and 9 days in the past month had a median cotinine concentration of 34 ng/mL. These findings suggest that occasional and very light smokers will not meet the cutoff concentration of 100 ng/mL, even if they smoked in the past 3 days and reported this behavior in a survey.

4.3. Self-Reported Tobacco Use: Comparison of Responses to Core and Repeat Questions

Table 4.3 presents a two-by-two table showing the relationship between self-reported 30-day cigarette and cigar use in the core and repeat questions. Recall that the Validity Study's core questions are modeled precisely on the National Household Survey on Drug Abuse (NHSDA), while the repeat questions occurred after the "debriefing intervention" near the end of the interview. Another difference was that missing data in the core tobacco variables were removed

through statistical imputation according to standard practices with the NHSDA; variables corresponding to the repeat questions were not imputed.

The introduction to the repeat questions was randomized. Approximately half of the sample received the long introduction, which consisted of an appeal to "please report truthfully." The other half received the short introduction, which instructed them that they would be asked a few questions "about what you think about our survey, followed by a repeat of some of the questions about drugs." The time periods for the repeat questions were designed to correspond more closely to the timeframes for the biological specimens and testing methods, with timeframes of 3 days, 7 days, 30 days, and 6 months. The repeat questions asked about the recency of use, following a positive response to a question on lifetime use.

Table 4.3 Cigarette or Cigar Use: Comparison of Responses to 30-Day Self-Report *Core* and *Repeat* Questions

(*n* = 4,431)

		30-Day Self-Report of Cigarettes or Cigars (Core Questions)		
	A	(-)	(+)	**B**
30-Day Self-Report of Cigarettes or Cigars (Repeat Questions)	(-)	67.6%	2.8%	(3,216)
	(+)	1.3%	28.4%	(1,215)
	C	(3,162)	(1,269)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
840.50	1	<0.0001	0.912	0.981	0.904

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those reporting past 30 day use of cigarettes and cigars in the core and/or *repeat* questions.

Due to space and timing constraints, however, the repeat questions asked only about cigarettes and cigars and excluded other types of tobacco products. Questions about pipes and smokeless tobacco were not included because prior administrations of the NHSDA identified few current users of these tobacco products, especially among 12 to 25 year olds.

Table 4.3 shows the concordance between self-reported use of cigarettes or cigars in the core and repeat questions. The table shows that 67.6 percent were self-reported nonusers of cigarettes or cigars in either the core or repeat questions (cell A), and 28.4 percent were self-reported users of cigarettes or cigars in both the core and repeat questions (cell D). Therefore, 96.0 percent of persons aged 12 to 25 would be estimated to answer the same on both the core and repeat questions. Cell B shows that an estimated 2.8 percent would answer affirmatively in the core but negatively in the repeat questions. Cell C indicates that an estimated 1.3 percent would answer negatively in the core but affirmatively in the repeat questions.

Nearly 200 respondents (about 1 in 20) changed their answers between the core and repeat questions for use of cigarettes or cigars in the past 30 days. It was anticipated that more respondents would report use in the repeat than the core questions following the introduction of the appeal to half the respondents. However, it was not anticipated that a larger number (2.8 percent) would change their responses to "no use" in the repeat questions after reporting use in the core questions.

Table 4.4 compares self-reported use of any tobacco product in the core with self-reported use of cigarettes or cigars in the repeat questions. Note that 4.0 percent reported use of any tobacco product in the core, but reported no use of cigars and cigarettes in the repeat questions (cell B). At least some of the discordant results in cell B can be attributed to the different core and repeat questions. That is, those reporting use of tobacco products other than cigars and cigarettes in the core questions could honestly deny use of cigars or cigarettes in the repeat questions, yet they would fall in the "underreporting" category (cell B). As shown in cell C, 1.2 percent reported no use of any tobacco product in the core, but reported use of cigarettes or cigars in the repeat questions. The difference in cell B between Tables 4.3 and 4.4 has increased (1.2 percent), but there was less reporting of use in the repeat questions.

4.4. Self-Reported Tobacco Use: Comparison of Responses to Repeat Questions with Urinalysis

Although higher prevalence rates were generated in the core than in the repeat questions, the real issue is whether the core or repeat questions produced more accurate results. Table 4.5 provides a two-by-two table comparing self-reports in the past 30 days in the repeat questions with the urinalysis results. In the repeat questions, 21.4 percent of persons aged 12 to 25 were estimated to have self-reported that they used cigarettes or cigars and tested positive (cell D). Looking at just the positive responses to cigarettes and cigars in the core questions, 22.5 percent reported smoking in the past 30 days. The percentage classified as underreporting (cell B) in the repeat questions (8.0 percent) was larger than in the core questions (7.0 percent); however, in the repeat questions, 8.8 percent were classified as overreporting (cell C) compared with 9.4 percent in the core questions. Overall, 30.2 percent self-reported cigarette or cigar use in the repeat questions compared with 31.9 percent in the core questions.

Table 4.4 Tobacco Use: Comparison of Responses to 30-Day Self-Report *Core* and *Repeat* Questions

(*n* = 4,431)

		30-Day Self-Report of Any Tobacco Product (Core Questions)		
	A	(-)	(+)	**B**
30-Day Self-Report of Cigarettes or Cigars (Repeat Questions)	(-)	66.3%	4.0%	(3,216)
	(+)	1.2%	28.5%	(1,215)
	C	(3,105)	(1,326)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
873.99	1	<0.0001	0.876	0.982	0.878

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those reporting past 30 day use of cigarettes and cigars in the repeat questions any tobacco use in the core questions.

The sensitivity for the repeat questions, or the proportion of those who were classified as reporting use and testing positive, was 0.73, which was reduced relative to the core questions for any tobacco (sensitivity was 0.80). The specificity for the repeat questions, or the proportion of those who were classified as reporting no use and testing negative, was 0.88 for the repeat questions and 0.86 for the core questions for any tobacco. Thus, the specificity was about the same for the repeat and core questions. Overall, there was somewhat less agreement between self-report and urinalysis on the repeat questions than on the core questions, as indicated by the kappa values (0.60 for the repeat questions vs. 0.64 for any tobacco use from the core), but both the repeat and core questions showed sensitivity rates of at least 0.73 and specificity rates of at least 0.86.

Because the repeat questions were preceded by the appeal, it is important to determine the impact of the appeal on the concordance between self-report and urinalysis. Table 4.6 presents a pair of two-by-two tables—the first table for those who received the appeal, and the second for those who did not. The sensitivity is higher for those who received the appeal (0.77) compared with 0.69 for those who did not receive the appeal. The rate of underreporting (cell B) was lower among those who received the appeal (7.1 percent) compared with 8.9 percent for those who did not receive the appeal, which was precisely what the appeal was designed to do—

Table 4.5 Cigarette or Cigar Use: Comparison of Responses to 30-Day Self-Report *Repeat* Questions and Urinalysis

(*n* = 3,738)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Repeat Questions)	(-)	61.7%	8.0%	(2,683)
	(+)	8.8%	21.4%	(1,055)
	C	(2,660)	(1,078)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
479.19	1	<0.0001	0.728	0.875	0.598

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of cigarettes and cigars in the repeat questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration level of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

appeal to respondents to be more truthful. Also, the percentage in cell C who reported use despite having a negative urine test result was reduced among those receiving the appeal (7.9 vs. 9.7 percent for those not receiving the appeal). The CMH chi-square values on both tables were significant, and the kappa values showed moderate agreement between the self-report and urinalysis results. In addition, the kappa value for those who received the appeal (0.65) was improved over the kappa value for those who did not receive the appeal (0.55). Thus, the appeal created more congruent reporting.

Tables 4.7 and 4.8 present another way to examine the relationship between self-reported 30-day use in the core and repeat questions. In Table 4.7, respondents had to report use on both the core and repeat questions to be considered a user. Any discrepancy was recoded to "no use." In contrast, Table 4.8 combines the answers to the core or repeat questions, with any affirmative report of use taking precedence; that is, a "yes" on either the core or repeat was recoded as "yes."

Self-report responses were compared with urinalysis results under the two scenarios described above. To the extent that the pattern of reporting was the same in the core and repeat questions, data in Tables 4.7 and 4.8 should be identical. However, they were not. There was less

Table 4.6 Cigarette or Cigar Use: Comparison of Responses to 30-Day Self-Report *Repeat* Questions and Urinalysis, by Receipt of Appeal

Respondent Received Appeal

(*n* = 1,830)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Repeat Questions)	(-)	61.6%	7.1%	(1,286)
	(+)	7.9%	23.4%	(544)
	C	(1,286)	(544)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
336.84	1	<0.0001	0.768	0.886	0.649

Respondent Did Not Receive Appeal

(*n* = 1,908)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Repeat Questions)	(-)	61.8%	8.9%	(1,397)
	(+)	9.7%	19.6%	(511)
	C	(1,374)	(534)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
271.93	1	<0.0001	0.688	0.864	0.547

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use of tobacco products. Respondents then listen to one of two possible introductions to the next series of questions in which respondents are asked again to report their recency of use. One scenario very broadly introduces the next series of questions. This is defined as *repeat questions without appeal*. The second scenario gives a broad overview of the study and emphasizes the importance of the respondent's responses. An appeal is made to the respondent to answer the questions as honestly as he or she can. This is defined as *repeat questions with appeal*. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of cigarettes and cigars in the repeat questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

Table 4.7 Tobacco Use: Comparison of Responses to 30-Day Self-Report *Core and Repeat* Questions and Urinalysis

(*n* = 3,738)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Core and Repeat Questions)	(-)	62.6%	8.4%	(2,737)
	(+)	8.0%	21.1%	(1,001)
	C	(2,660)	(1,078)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
469.44	1	<0.0001	0.716	0.887	0.605

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of cigarettes and cigars in the *core* and *repeat* questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

sensitivity when an affirmative response to past 30 day use was required on both core and repeat questions (0.72) than when it was only required on either one (0.82). The specificity moved in the opposite direction, with more specificity when reports of past 30 day use on both the core and repeat questions were required (0.89) and less specificity when reported use on only one set of questions was sufficient (0.85). The CMH chi squares for both tables were significant, and the kappa values indicated moderate agreement (0.61 in Table 4.7 and 0.64 in Table 4.8).

Results of these tables show that self-report was more sensitive when an affirmative response to either the core or repeat questions was sufficient to classify a person as a user. However, a higher proportion would be classified as users in the absence of a positive urine test result (cell C) when this strategy is adopted (10.5 percent in Table 4.8 vs. 8.0 percent in Table 4.7), so specificity was reduced.

Table 4.8 Tobacco Use: Comparison of Responses to 30-Day Self-Report *Core or Repeat* Questions and Urinalysis

(*n* = 3,746)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Core or Repeat Questions)	(-)	60.0%	5.4%	(2,535)
	(+)	10.5%	24.1%	(1,211)
	C	(2,663)	(1,083)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
590.88	1	<0.0001	0.818	0.851	0.637

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of cigarettes and cigars in the *core* or *repeat* questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

4.4.1 Passive Exposure

Passive exposure is a large concern with smoked drugs. Recall that in epidemiological studies the exposure to secondhand smoke has been repetitively linked to the elevation of biomarker levels in nonsmokers, including the tobacco-specific biomarkers nicotine, cotinine, and 4-(methylnitrosamino)1-(3-pryidyl)-1-butanone (U.S. DHHS, 2006). The Validity Study used a cutoff of 100 ng/mL; however, there was no clear consensus in the literature that the cutoff concentration for cotinine should be 100 ng/mL to distinguish ETS exposure from tobacco use. At least one study suggested it should be 500 ng/mL, although another study used a concentration of 75 ng/mL to separate ETS exposure from tobacco use. ETS exposure is far from insignificant among those who would be classified as nonusers of tobacco products in the past 30 days based on self-reports for either the core or repeat questions, but who nevertheless tested positive. Approximately half (45.0 percent) reported being in a room with smokers on a daily basis. Their median cotinine screening test concentration was 977 ng/mL, which is well above the 100 ng/mL concentration that the Validity Study used as a cutoff. Another 28.8 percent reported frequently being exposed to cigarette smoke, but not every day. Therefore, it appears that the majority of those who did not report tobacco use but had a positive urine test result were

exposed regularly to ETS. Consequently, discordance between self-reports of nonuse in the past 30 days and positive urine test results may not discredit all self-reports of nonuse of tobacco products.

4.4.2 Comparisons of Self-Report and Urinalysis Results, by Age

Because it is illegal for youths under the age of 18 to purchase tobacco products in the United States, youths who are tobacco users may be more reluctant than tobacco users aged 18 or older to report use in a survey. Research literature, primarily derived from the methodological studies of the NHSDA, has shown, for example, that youths were more likely to report tobacco use when they answered questions on a self-administered and confidential answer sheet than they were if they had to report use aloud to an interviewer (Office of Applied Studies [OAS], 1996). To examine this issue in the Validity Study, Table 4.9 shows a pair of two-by-two tables of the self-report from the core or repeat questions compared with urinalysis results. The first table shows results for youths aged 12 to 17, and the second shows results for young adults aged 18 to 25.

Consistent with the hypothesis that youths may underreport tobacco use to avoid adverse social or legal consequences, the sensitivity was higher among young adults (0.89) than among youths (0.60). Specificity also was affected and was lower among young adults (0.79) than among youths (0.91). This latter finding is due to a higher proportion of young adults in cell C who self-reported use but did not test positive (12.8 vs. 7.7 percent for youths).

Although these results suggest that the legal prohibition in the United States against the purchase of tobacco products may influence some youths to underreport use, there also appeared to be some underreporting among young adults aged 18 to 25, who can legally purchase tobacco products. For young adults, 4.2 percent tested positive for cotinine and reported no use of a tobacco product in the past 30 days (cell B). Running these same tables with only the core questions yielded very similar results, with specificity increasing slightly in both age groups. Sensitivity was lower among the 12 to 17 year olds and slightly lower among the 18 to 25 year olds.

4.5. Comparisons of Self-Report and Urinalysis Results for 7-Day and 3-Day Windows

Because the time window for detection of cotinine in the urine is estimated to be as little as 2 days and as much as 7 days, the 30-day timeframe is too long. For many of those who might be considered "overreporters" in the prior tables, that might not actually be the case because urinalysis for cotinine probably should not detect tobacco use that occurred more than 7 days in the past. However, the repeat questions asked about use in the 7-day timeframe, although these questions were restricted to cigarette and cigar use only.

Table 4.10 is a two-by-two table showing self-reported use of cigarettes or cigars in the past 7 days in the repeat questions in comparison with urinalysis results. An estimated 65.2 percent (cell A) reported no smoking of cigarettes or cigars in the past 7 days and had a negative urinalysis for cotinine. The proportion of persons in cell B who reported no use in the past 7 days

Table 4.9 Tobacco Use: Comparison of Responses to 30-Day Self-Report *Core or Repeat* Questions and Urinalysis, by Age Group

12 to 17 Years

(*n* = 1,943)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Core or Repeat Questions)	(-)	75.1%	6.9%	(1,595)
	(+)	7.7%	10.4%	(348)
	C	(1,594)	(349)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
153.27	1	<0.0001	0.601	0.907	0.500

18 to 25 Years

(*n* = 1,803))

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Core or Repeat Questions)	(-)	47.8%	4.2%	(940)
	(+)	12.8%	35.2%	(863)
	C	(1,069)	(734)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
499.05	1	<0.0001	0.894	0.789	0.658

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of any tobacco use in the core questions or past 30 day of cigarettes or cigars in the repeat questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

Table 4.10 Cigarette or Cigar Use: Comparison of Responses to 7-Day Self-Report *Repeat* Questions and Urinalysis

(*n* = 3,738)

		Urinalysis		
	A	(-)	(+)	**B**
7-Day Self-Report (Repeat Questions)	(-)	65.2%	8.6%	(2,834)
	(+)	5.4%	20.8%	(904)
	C	(2,660)	(1,078)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
461.88	1	<0.0001	0.706	0.924	0.651

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 7 day use of cigarettes or cigars in the repeat questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

but who tested positive (8.6 percent) increased relative to the 8.0 percent for the 30-day time referent in the repeat questions (Table 4.5). In addition, 5.4 percent reported use and tested negative (cell C). Compared with the 30-day results in Table 4.5, the proportion in cell C was reduced but not eliminated. However, this latter finding was expected and is reflected in the improved specificity (0.92 for the 7-day period vs. 0.88 for the 30-day period from Table 4.5), although sensitivity was reduced slightly (0.71 vs. 0.73, respectively). The CMH chi square for Table 4.10 was significant, and the kappa value of 0.65 showed moderate agreement between the self-report and urinalysis results.

Taken together, these results suggest that self-reports of use in the more recent time period of 7 days were more specific in terms of not eliciting discordant self-reports of use in the past 7 days among persons whose urine test results were negative, in comparison with the situation for a 30-day reference period. However, this 7-day reference period that is more proximal to the interview date also may elicit a higher rate of underreporting of current use.

Table 4.11 includes a pair of two-by-two tables showing the relationship between the urinalysis results and self-reported cigarette or cigar use in the past 7 days in the repeat questions

Table 4.11 Cigarette or Cigar Use: Comparison of Responses to 7-Day Self-Report *Repeat* Questions and Urinalysis, by Age Group

12 to 17 Years

(*n* = 1,941)

		Urinalysis		
	A	(-)	(+)	**B**
7-Day Self-Report (Repeat Questions)	(-)	79.5%	9.1%	(1,721)
	(+)	3.3%	8.1%	(220)
	C	(1,593)	(348)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
136.59	1	<0.0001	0.472	0.960	0.499

18 to 25 Years

(*n* = 1,797)

		Urinalysis		
	A	(-)	(+)	**B**
7-Day Self-Report (Repeat Questions)	(-)	53.6%	8.3%	(1,113)
	(+)	7.1%	31.0%	(684)
	C	(1.067)	(730)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
397.28	1	<0.0001	0.789	0.884	0.677

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 7 day use of cigarettes or cigars in the repeat questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

by age group. Relative to the 30-day reference period data in Table 4.9 for the core *or* repeat questions, the kappa value for the 7-day reference period was virtually the same for youths (0.50) and increased slightly for young adults (0.68 for the 7-day period vs. 0.66 for the 30-day period). The sensitivity for 12 to 17 year olds was reduced from 0.60 for the 30-day reference period to 0.47 for the 7-day period, while the 7-day reference period sensitivity for young adults remained relatively high at 0.79. For both youths and young adults, specificity improved for the 7-day reference period (for youths: 0.96 for the 7-day period vs. 0.91 for the 30-day period; for young adults: 0.88 vs. 0.79, respectively).

The improvement in specificity for the 7-day reference period, or reduction in the occurrence of "false-positive" self-reports that are inconsistent with urine test results, is consistent with research literature suggesting that the 7-day window is more accurate for the maximum length of time that cotinine may remain in the urine. Conversely, the reduction in sensitivity for youths for the 7-day reference period is consistent with the hypothesis of some youths underreporting tobacco use because of social stigma associated with use, and particularly for use that may be relatively close to the interview date.

Both the follow-up questions following the core module and the repeat questions toward the end of the interview asked about tobacco use in the past 3 days; as noted previously, however, the repeat questions ask only about cigarettes and cigars specifically. Therefore, Table 4.12 presents a two-by-two table using self-reported use of tobacco in the past 3 days in the follow-up questions in comparison with urinalysis results. An estimated 66.9 percent of persons aged 12 to 25 were classified as reporting no tobacco use in the past 3 days and having a negative urinalysis for cotinine (cell A), while 21.8 percent (cell D) reported using tobacco in the past 3 days and tested positive. So, 88.7 percent had congruent self-report and cotinine results. In terms of incongruent patterns, 7.7 percent had a negative self-report but a positive urine test result (cell B), and 3.6 percent had a positive self-report of use but a negative urine test result (cell C).

The sensitivity level for the past 3 days (0.74) decreased relative to the comparable 30-day measure for the core (0.80; Table 4.2), while specificity was enhanced (0.95 for the past 3 days vs. 0.86 for the past 30 days). The CMH chi square was significant, and the kappa value of 0.72 showed moderate agreement between the self-report and urinalysis results. These findings again suggest that there may be a tendency for respondents to underreport recent use that is more proximal to the interview date, both for the more recent time period of 3 days in comparison with a 30-day time period, as well as the 7-day time period.

4.6. Varying Cotinine Cutoff Concentrations and Self-Report Timeframes

The Validity Study obtained information on cotinine concentrations, so it is possible to examine the efficiency of various cutoff concentrations to designate specimens as positive by comparing the distribution of concentrations in relationship to the self-reported use. Table 4.13 examines the recency of use in the repeat questions and various screening cutoff concentrations for cotinine. Most of those with 0 to 75 ng/mL of cotinine in their urine (87.5 percent) reported no use of cigarettes or cigars in the past 30 days. Among those with cotinine concentrations between 76 and 99 ng/mL, 65.7 percent reported no use, but about one fourth (23.4 percent)

Table 4.12 Tobacco Use: Comparison of Responses to 3-Day Self-Report *Follow-Up* Questions and Urinalysis

(*n* = 3,759)

		Urinalysis		
	A	(-)	(+)	**B**
3-Day Self-Report (Follow-Up Questions)	(-)	66.9%	7.7%	(2,873)
	(+)	3.6%	21.8%	(886)
	C	(2,674)	(1,085)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
535.50	1	<0.0001	0.740	0.948	0.717

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 3 day use of tobacco products on the *follow-up* questions.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

Note 4: A cutoff concentration of 100 nanograms per milliliter (ng/mL) was used to identify a specimen as positive for cotinine.

reported use in the past 3 days. Among those who tested positive between 100 and 249 ng/mL, slightly more than half (52.9 percent) reported no cigarette or cigar use, and more than a third (39.5 percent) reported use in the past 3 days. The vast majority (71.7 percent) of those with cotinine concentrations greater than 500 ng/mL reported smoking in the past 3 days. The majority of those with positive results at the 250 to 500 ng/mL screening concentration also reported cigarette or cigar use in the past 3 days (57.6 percent).

Our data generally support 100 ng/mL as an efficient cutoff using the ELISA method to differentiate frequent tobacco users from infrequent users and nonusers exposed to ETS. There were 170 respondents who reported cigarette or cigar use in the past 7 days but were not positive at the 100 ng/mL cutoff concentration, and 320 respondents who reported no use in the past 30 days but were positive at the 100 ng/mL cutoff concentration. Recall that the repeat questions did not ask about the full range of tobacco products, which may explain some of these inconsistencies. If a cutoff concentration of 500 ng/mL is applied, only 11 (23.2 percent, *n* = 214) of those who tested positive reported no use in the past month.

Table 4.13 Cigarette or Cigar Use: Comparison of Responses to Self-Report *Repeat* Questions on Recency of Use at Varying Urinalysis Screening Concentrations: Percentages

Recency of Use	Screening Concentrations				
	0 - 75 ng/mL (*n* = 2,631)	76 - 99 ng/mL (*n* = 43)	100 - 249 ng/mL (*n* = 127)	250 - 500 ng/mL (*n* = 107)	> 500 ng/mL (*n* = 851)
No Past Month Use	87.5	65.7	52.9	30.7	23.2
Past 3 Days	4.7	23.4	39.5	57.6	71.7
More Than 3 Days, but in Past 7 Days	2.5	7.1	4.8	8.5	2.3
More Than 7 Days, but in Past 30 Days	4.9	2.4	2.8	2.9	2.0
Don't Know/Refusal	0.4	1.4	-	0.3	0.7

ng/mL = nanograms per milliliter.

- Zero cases in cell.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions ask only about cigarettes and cigars.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those reporting cigarette or cigar use on *repeat* questions and submitting a valid urine specimen.

Note 3: Those who reported using products to quit smoking in the past 30 days were excluded.

If the cutoff concentration were raised to 500 ng/mL, it still would not account for all the underreporting of tobacco use, although 106 fewer respondents would not be considered underreporters in the past 30 day timeframe. Such a change also would increase the number of overreporters by 106. Generally, attention focuses on those reporting no use but testing positive, but those reporting use and not testing positive are of equal interest to validity research. Those positive at the comparatively lower cutoff concentrations (the majority of those who test positive have cotinine metabolite concentrations over 500 ng/mL) are perhaps light smokers or occasional tobacco users. Fewer respondents who tested positive between 250 and 500 ng/mL and those who tested positive at more than 500 ng/mL reported no use. It would thus appear that a small proportion were not willing to admit to tobacco use, even on a questionnaire with all the built-in safeguards of the NHSDA. It is also possible the screening test for cotinine in the urine may not be sensitive enough to detect infrequent use. The corollary also seems possible that the screening test is not specific enough, and cross-reactivity with other analytes may have caused a positive test. It is important to remember that the Validity Study did not use a confirmatory test for cotinine.

4.7. Overreporters

The goal of the Validity Study was to determine not only the prevalence, but also the correlates of underreporting and overreporting. It was therefore important to ascertain which respondents accurately fit these descriptors. There were 322 respondents who reported using a tobacco product in the past 30 days in the core and who did not test positive and who did not

report using a quit-smoking product in the past 30 days; 311 of these reported cigarette and cigar use. There were 183 respondents who reported use in the past 7 days in the repeat questions (recall that repeat questions asked only about cigarettes and cigars) and/or reported use of any tobacco product in the past 3 days on the follow-up questions. There were 112 who reported use of a tobacco product in the past 3 days on the follow-up questions and who did not test positive. Many of the 30-day overreporters can be dismissed because they used outside the window of detection, but not the 7-day or 3-day overreporters. Of the 7-day overreporters, 98 of the 183 smoked on 10 or fewer days in the past 30 days, which may explain why they did not test positive. Of the 27 who reported smoking cigarettes every day in the past 30 days, 19 said they smoked six or more cigarettes a day. Of these, 16 had less than 34 ng/mL of cotinine in their urine, and 10 had no detectable cotinine.

Turning to the 3-day overreporters, 71 of 112 had a cotinine concentration of less than 30 ng/mL. Of these, 19 reported smoking cigarettes every day on each of the past 30 days (18 reported using other tobacco products). Of the 19 respondents, 6 reported smoking 5 or fewer cigarettes a day, which may not be sufficient tobacco use to produce a positive cotinine screening test. However, the remaining 13 reported smoking at the level of half of a pack a day or more.

Among the 3-day overreporters, 43 reported smoking cigarettes on at least half the days in the past month. Of these, 13 had no detectable cotinine in their urine, and 23 had 10 ng/mL of cotinine or less. Of the 13 with no detectable cotinine who reported smoking in the past 3 days, 11 said they smoked daily in the past 30 days. Of these 11 respondents, 9 reported being "always truthful," 9 had complete privacy during the interview (the remaining 4 had "minor distractions" or less), and all reported being exposed to tobacco smoking on a daily or frequent basis. Therefore, although it appears that about a third of the overreporters were light or occasional smokers (smoked on 3 or fewer days in the past 30 days), the rest were smoking, some as much as several cigarettes a day, every day in the past month and did not test positive at the 100 ng/mL screening cutoff concentration.

Logistic regression models were developed to determine the correlates of 3-day overreporting. Because overreporting varied by age group, separate models were developed for each of the two age groups. The adjusted odds ratio for each variable in the model is reported. The odds ratios reflect the likelihood of a positive response on the dependent variable relative to that for the defined reference group, after controlling for all the other variables included in the model. Adjusted odds ratios greater than 1.0 indicate an increased likelihood of overreporting, and those less than 1.0 indicate a decreased likelihood of overreporting. To help ensure the models were targeted to actual overreporters, the dependent variable was restricted to respondents who reported that they had used any kind of tobacco product on 10 or more days in the past month and used in the past 3 days. This group included 46 respondents: 12 between the ages of 12 and 17 and 34 between the ages of 18 and 25.

The models for overreporting and the models for underreporting that follow contain two types of predictor or independent variables. The first group consists of the categorical variables:

- region (four geographic regions, Northeast, Midwest, South, and West used as dummy variables, with West as the reference category),

- race (white, black, and other, with white as the reference category),
- gender (male and female, with male as the reference group),
- appeal (received appeal or not, with those receiving the appeal as the reference group), and
- truthfulness (completely truthful or not, with completely truthful as the reference category).

The second group of variables are considered as ordinal or continuous measures: privacy (an ordinal measure with higher scores meaning less privacy), passive exposure (a higher score means more frequent passive exposure), and friends' cigarette smoking (a higher score means more friends who smoke cigarettes). In the 2000 survey, respondents were asked about their passive exposure to cigarette smoke; in 2001, however, the corresponding question was generalized to exposure to smoke from cigarettes or any other tobacco product.

Research has found repeatedly that peers are a very important influence on any type of drug use, particularly for youths. Because of this, the variable related to the question "How many of your friends would you say smoke cigarettes?" was included in the models. The goal was to ascertain how peer drug-using behavior affects the validity of self-reports.

Beyond these ordinal measures, two complex variables treated as continuous—religiosity and "difficulties"—are included in the models. Each of these variables is the sum of four related indicators where respondents with unknown information on any of the four questions were excluded. The religiosity measure has a potential range between 1 and 15, while the "difficulties" measure ranges from 1 to 13. "Religiosity" is composed of questions about the number of religious services attended in the past year, the importance of religious beliefs, how much religious beliefs influence decisions, and the importance of friends sharing religious beliefs. For "religiosity," 1 indicates not religious and 15 very religious. The "difficulties" variable is composed of responses to debriefing questions that ask how difficult it was for the respondents to remember; how difficult it was for them to understand; the clearness of their memory; and how often they had to make their "best guess" when answering the drug-related questions. For the "difficulties" variable, 1 means lack of difficulties, and 13 indicates a lot of difficulties answering drug-related questions.

4.7.1 12 to 17 Year Olds

The models for youths aged 12 to 17 contain only 13 overreporters. Race was not included because all 13 of the 3-day overreporters were white. The results for the first logistic model are shown in column 1 of Table 4.14. This model compares the overreporters with the "true nonusers" (those who reported no use and tested negative [cell A]). Four variables were found to be significantly related to overreporting in this model: truthfulness, passive tobacco smoke exposure, having friends who smoke, and religiosity. The variable with the largest odds ratio in the model was friends' cigarette smoking. The likelihood of overreporting increased as the score for the number of friends who smoked cigarettes also increased. In addition, those who admitted to being less than completely truthful in their answers to drug questions were more than 6 times more likely to overreport compared with those who reported that they were completely truthful, controlling for the other variables in the model. The odds ratios also suggest that the

likelihood of overreporting increased as passive exposure increased. Finally, the likelihood of overreporting decreased as religiosity increased.

Having smoking or nonsmoking friends should be correlated with the degree of passive exposure to ETS; when somebody is around smokers, it is normal to be exposed to ETS. The combined effect of the two variables for passive exposure and friends' cigarette smoking was included in a model (not shown here) as an interaction term. The results did not show a significant improvement over the additive model.

In the second model shown in the second column of Table 4.14, overreporters are compared with the "true users" (those who self-reported tobacco use and tested positive for cotinine [cell D in the two-by-two tables]). It should be noted that there were many fewer "true users" (n = 166) than "true nonusers" (n = 1,501) in the 12- to 17-year-old group.

The rationale behind the comparison with true users is that if overreporters really are using tobacco products and are mistakenly classified based on their urine test results (e.g., if they happened to have lower cotinine concentrations at the time of the interview), they should not differ significantly from the group of true users. Indeed, none of the variables in this second model was a significant predictor of overreporting. Thus, the hypothesis that overreporters will resemble true users was supported among youths aged 12 to 17.

4.7.2 18 to 25 Year Olds

The models examining overreporters among participants between 18 and 25 years of age (n = 36) consist of the same independent variables as the models for the younger age group. Results are presented in Table 4.15.

The results of the first model comparing overreporters with true nonusers are presented in the first column of Table 4.15. Three of the four variables that were significant predictors of overreporting for youths also were significant predictors for young adults (truthfulness, passive tobacco smoke exposure, and friends' cigarette smoking). However, religiosity was not a significant predictor in the model for young adults. The odds ratios for the significant predictors in the model for young adults also were more modest than those observed with the younger age group, although the two largest odds ratios for young adults again were truthfulness and having friends who smoke. The most significant independent variable was truthfulness. Not surprisingly, there appeared to be some collinearity between having friends who smoke cigarettes and exposure to ETS.

The second model in Table 4.15 compares 18- to 25-year-old overreporters with true users from the same age group. For these young adults, the number of true users (n = 587) was much greater than it was for the 12- to 17-year-old age group. The results indicate that three variables discriminate between young adults who were overreporters or true users: gender, passive exposure to tobacco smoke, and truthfulness. Among young adults, the likelihood of overreporting was greater for those who admitted to being less than completely truthful in their answers to some drug questions. However, the odds ratio for passive exposure was in the opposite direction than that observed in the previous model comparing overreporters with the true nonusers. Stated another way, the likelihood of overreporting *decreased* with more frequent exposure to ETS. The 18- to 25-year-old overreporters were predominantly females (62.5 percent) compared with 45 percent for the true users.

Table 4.14 Tobacco Use: Logistic Regression Models Predicting 3-Day Overreporting among Youths Aged 12 to 17

Model Covariate	OR Versus TNU		OR Versus TU	
	Odds Ratio	95% CI	Odds Ratio	95% CI
Gender				
Male	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Female	1.50	(0.42-5.36)	1.51	(0.44-5.14)
Region				
West	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Midwest	0.39	(0.04-3.43)	0.80	(0.12-5.37)
South	0.74	(0.12-4.44)	0.52	(0.11-2.53)
Northeast	0.51	(0.07-3.52)	0.42	(0.08-2.40)
Receipt of Appeal				
Received Appeal	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Did Not Receive Appeal	1.41	(0.34-5.88)	2.48	(0.56-10.91)
Truthfulness				
Completely Truthful	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Not at All/Somewhat/ Mostly Truthful	6.26[a]	(1.30-30.14)	2.28	(0.52-9.99)
Privacy of Interview[1]	1.10	(0.88-1.36)	1.07	(0.86-1.34)
Difficulty in Answering Drug-Related Questions[2]	0.93	(0.68-1.28)	0.79	(0.54-1.14)
Passive Tobacco Smoke Exposure in Past 6 Months[3]	4.57[b]	(1.64-12.75)	1.31	(0.58-2.97)
Friends Smoking Cigarettes [4]	7.92[b]	(3.56-17.58)	1.69	(0.74-3.88)
Religiosity[5]	0.78[a]	(0.65-0.94)	0.89	(0.75-1.07)
	n = 13/1,501 χ^2 = 92.78 df = 11		n = 13/166 χ^2 = 17.00 df = 11	

CI = Confidence Interval.

Note: Overreporters (OR) self-report positive and have a negative urine test, true nonusers (TNU) self-report negative and have a negative urine test, and true users (TU) self-report positive and have a positive urine test. The 3-day self-reporting is based on responses to follow-up questions pertaining to any tobacco use. Those who reported using quit-smoking products in the past 30 days were excluded.

[a] Statistically significant at the 0.05 level.

[b] Statistically significant at the 0.01 level.

[1] Higher score indicates less privacy during interview.

[2] Respondents were asked four questions on their difficulties in (1) understanding the drug-related questions and (2) remembering the drug-related information, as well their (3) clarity of memories regarding drug-related information and (4) how often they made a best guess when answering the drug-related questions. Higher scores indicate more difficulty.

[3] Higher score indicates more frequent exposure to persons smoking cigarettes or other tobacco products in the past 6 months.

[4] Higher score indicates more friends who smoke cigarettes.

[5] Respondents were asked four questions regarding their religious beliefs and practices: (1) number of times attended religious services in the past year, (2) importance of religious beliefs, (3) influence of religious beliefs on decisionmaking, and (4) importance of friends sharing religious beliefs. Higher scores indicate more religiosity.

Table 4.15 Tobacco Use: Logistic Regression Models Predicting 3-Day Overreporting among Young Adults Aged 18 to 25

Model Covariate	OR Versus TNU		OR Versus TU	
	Odds Ratio	95% CI	Odds Ratio	95% CI
Gender				
Male	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Female	1.45	(0.68-3.09)	2.33[a]	(1.09-4.98)
Region				
West	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Midwest	1.53	(0.33-7.11)	1.19	(0.27-5.28)
South	1.87	(0.52-6.70)	1.80	(0.58-5.59)
Northeast	1.45	(0.45-4.69)	1.29	(0.39-4.23)
Receipt of Appeal				
Received Appeal	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Did Not Receive Appeal	1.53	(0.64-3.64)	1.70	(0.76-3.79)
Truthfulness				
Completely Truthful	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Not at All/Somewhat/ Mostly Truthful	4.40[a]	(1.39-13.95)	3.28[a]	(1.29-8.33)
Privacy of Interview[1]	0.94	(0.68-1.31)	0.89	(0.60-1.33)
Difficulty in Answering Drug-Related Questions[2]	1.14	(0.99-1.32)	1.06	(0.92-1.22)
Passive Tobacco Smoke Exposure in Past 6 Months[3]	1.93[a]	(1.14-3.28)	0.61[a]	(0.39-0.95)
Friends Smoking Cigarettes[4]	3.97[b]	(2.23-7.07)	1.73	(0.89-3.34)
Religiosity[5]	0.97	(0.87-1.07)	1.00	(0.89-1.13)
	n = 36/979 χ^2 = 88.28 df = 11		n = 36/587 χ^2 = 25.66 df = 11	

CI = confidence interval.

Note: Overreporters (OR) self-report positive and have a negative urine test, true nonusers (TNU) self-report negative and have a negative urine test, and true users (TU) self-report positive and have a positive urine test. The 3-day self-reporting is based on responses to follow-up questions pertaining to any tobacco use. Those who reported using quit-smoking products in the past 30 days were excluded.

[a] Statistically significant at the 0.05 level.

[b] Statistically significant at the 0.01 level.

[1] Higher score indicates less privacy during interview.

[2] Respondents were asked four questions on their difficulties in (1) understanding the drug-related questions and (2) remembering the drug-related information, as well their (3) clarity of memories regarding drug-related information and (4) how often they made a best guess when answering the drug-related questions. Higher scores indicate more difficulty.

[3] Higher score indicates more frequent exposure to persons smoking cigarettes or other tobacco products in the past 6 months.

[4] Higher score indicates more friends who smoke cigarettes.

[5] Respondents were asked four questions regarding their religious beliefs and practices: (1) number of times attended religious services in the past year, (2) importance of religious beliefs, (3) influence of religious beliefs on decisionmaking, and (4) importance of friends sharing religious beliefs. Higher scores indicate more religiosity.

4.8. Underreporters

Among underreporters (i.e., those who self-reported no use but who tested positive for cotinine in their urine), there were 311 who reported no use of any tobacco products in the past 3 days on the follow-up questions, 344 who reported no use of cigarettes or cigars in the past 7 days in the repeat questions, and 339 who reported no use in the past 30 days in the core or repeat questions. The 30-day reporting period is considered the most stringent of any of the time periods because anyone who reported no use of tobacco products in the past 30 days should not have cotinine concentrations above 100 ng/mL. However, because the best available research suggests the window of detection for cotinine is no more than 7 days, the 7-day self-reports formed the basis for the underreporter models examined here.

Among underreporters 18 to 25 years old, 71.5 percent had cotinine concentrations above 500 ng/mL, and about 25 percent had concentrations between 100 and 249 ng/mL. More than 41 percent of the 18- to 25-year-old underreporters indicated daily ETS exposure. In this age group, females were a little more likely to be exposed to ETS, although gender differences were minimal.

An examination of underreporters' responses to the debriefing questions shows that their responses did not differ greatly from the overall responses. A slightly smaller, but still robust, 83.4 percent reported being always truthful. They did not report greater difficulty understanding the drug-related questions than most respondents. However, they were a little more likely to report having difficulty remembering, and more frequently they had to make their "best guess" when answering the drug-related questions. They were not more embarrassed than the norm, which is "not very embarrassed," nor were they more concerned that others would have access to their answers. Almost half (47.4 percent) reported being exposed to ETS daily, and three quarters (75.6 percent) had daily-to-frequent exposure to ETS. The same patterns were evident among 30-day and 3-day underreporters.

As with the overreporters, logistic regression models were estimated separately for each of the two age groups of underreporters: youths aged 12 to 17 and young adults aged 18 to 25. Even more than for overreporters, prior analyses have shown differences in underreporting by age group, necessitating the separate age group models. Responses to both the core and repeat questions were combined for these analyses. Four models were constructed for the purpose of describing the group of underreporters. The same independent variables used in the logistic regressions for overreporters were used in the models predicting underreporters, distinguished from "true nonusers" and from "true users" and defined using two different cutoff concentrations for cotinine (100 and 500 ng/mL). Analyses showed that there were no differences in significant effects using the 500 ng/mL cutoff concentration and an even higher cutoff concentration of 1,000 ng/mL, so the highest cutoff concentration used in these regression model criteria is 500 ng/mL.

4.8.1 12 to 17 Year Olds

Table 4.16 presents the results of logistic regression models predicting 7-day underreporting of tobacco products among respondents aged 12 to 17. As expected, underreporters were more likely to be found in this younger age group (n = 174). As suggested

previously, these youthful smokers may have had more of a reason not to reveal their tobacco use because smoking is illegal for those under age 18.

The first of the three models compares underreporters with true nonusers. This is the first and only time in the tobacco models that a significant effect for region was found. The results suggest that respondents aged 12 to 17 from the South and Northeast were more likely than those in the West to underreport, with odds ratios of 2.9 and 2.7, respectively. Race also was a significant predictor of underreporting. Persons in the "other" race category were much less likely than whites to underreport, but the "other" category was very small and diverse, rendering the results not easily interpretable. However, the likelihood of being an underreporter was not significantly different for blacks and whites. In addition, those who admitted to being less than completely truthful in their answers to drug questions were more likely to be underreporters compared with persons who reported being completely truthful.

The three noncategorical variables that were significant predictors of underreporting relative to true nonusers were passive tobacco smoke exposure, friends' cigarette smoking, and religiosity. For passive tobacco smoke exposure and friends' smoking, the odds of underreporting increased with more frequent exposure to ETS or increased numbers of friends who smoked. The odds ratio for religiosity was in the opposite direction, indicating that the likelihood of underreporting decreased somewhat as religiosity increased. Separate analyses of the levels of exposure (not shown in Table 4.16) revealed that respondents with daily exposure were twice as likely to underreport than those not exposed.

The next models shown in Table 4.16 compare the underreporters with the group of true users using two cutoff concentrations to define true users. The first of these models is presented in the second column of Table 4.16 using the 100 ng/mL cutoff concentration. Again, the race variable was significant for persons in the "other" category compared with whites. Although blacks appeared to have increased odds of being underreporters compared with whites, the odds ratio for blacks was not significantly different from 1.0.

The other significant predictors in this model were passive tobacco smoke exposure and friends' cigarette smoking. The effect for the latter was quite strong, with an odds ratio much less than 1.0, suggesting that the more friends that the respondent had who smoked cigarettes, the less likely the respondent was to underreport.

The model shown in column 3 of Table 4.16 again compares underreporters with true users. For this model, however, only respondents with cotinine concentrations of 500 ng/mL and above were included in the true user group. The rationale was to include only those cases that were "definitely positive" and could not reach such cotinine concentrations due to ETS. The 500 ng/mL reading can be considered a relatively safe cutoff concentration for this purpose. Interestingly, raising the threshold to 500 ng/mL for classifying a person as a true user did not remove passive tobacco smoke exposure as a significant predictor of underreporting. The likelihood of underreporting decreased as the reported frequency of passive exposure increased.

Table 4.16 Tobacco Use: Logistic Regression Models Predicting 7-Day Underreporting among Youths Aged 12 to 17 at Varying Cotinine Cutoff Concentrations

Model Covariate	UR Versus TNU		UR Versus TU (Cotinine Cutoff 100 ng/mL)		UR Versus TU (Cotinine Cutoff 500 ng/mL)	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Race						
White	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Black	1.50	(0.89-2.53)	2.71	(0.92-7.97)	2.86	(0.87-9.44)
Other	0.04[b]	(0.00-0.31)	0.03[a]	(0.00-0.46)	0.08	(0.01-1.09)
Gender						
Male	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Female	0.88	(0.59-1.30)	1.07	(0.55-2.05)	0.93	(0.47-1.87)
Region						
West	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Midwest	1.72	(0.78-3.77)	1.42	(0.59-3.41)	1.43	(0.64-3.17)
South	2.88[b]	(1.50-5.51)	1.61	(0.66-3.89)	1.58	(0.69-3.62)
Northeast	2.71[b]	(1.40-5.26)	1.19	(0.49-2.91)	1.34	(0.57-3.18)
Receipt of Appeal						
Received Appeal	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Did Not Receive Appeal	1.01	(0.65-1.58)	1.09	(0.59-2.02)	1.10	(0.58-2.09)
Truthfulness						
Completely Truthful	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Not at All/Somewhat/ Mostly Truthful	2.13[a]	(1.14-3.96)	1.24	(0.51-3.02)	1.11	(0.43-2.86)
Privacy of Interview[1]	1.07	(0.97-1.18)	1.26	(0.88-1.79)	1.25	(0.86-1.81)
Difficulty in Answering Drug-Related Questions[2]	1.02	(0.95-1.11)	1.03	(0.92-1.16)	1.05	(0.93-1.18)
Passive Tobacco Smoke Exposure in Past 6 Months[3]	1.69[b]	(1.28-2.23)	0.44[b]	(0.31-0.62)	0.44[b]	(0.31-0.63)
Friends Smoking Cigarettes[4]	1.48[b]	(1.13-1.93)	0.16[b]	(0.09-0.29)	0.17[b]	(0.09-0.30)
Religiosity[5]	0.93[a]	(0.88-0.99)	1.09	(1.00-1.20)	1.09	(0.99-1.20)
	n = 174/1,469 χ^2 = 126.54 df = 13		n = 174/156 χ^2 = 164.23 df = 13		n = 174/133 χ^2 = 149.82 df = 13	

CI = confidence interval.

ng/m/L = nanograms per milliliter.

Note: Underreporters (UR) self-report negative and have a positive urine test, true nonusers (TNU) self-report negative and have a negative urine test, and true users (TU) self-report positive and have a positive urine test. The 7-day self-reporting is based on responses to repeat questions pertaining to cigarette or cigar use only. Those who reported using quit-smoking products in the past 30 days were excluded.

[a] Statistically significant at the 0.05 level.

[b] Statistically significant at the 0.01 level.

[1] Higher score indicates less privacy during interview.

[2] Respondents were asked four questions on their difficulties in (1) understanding the drug-related questions and (2) remembering the drug-related information, as well their (3) clarity of memories regarding drug-related information and (4) how often they made a best guess when answering the drug-related questions. Higher scores indicate more difficulty.

[3] Higher score indicates more frequent exposure to persons smoking cigarettes or other tobacco products in the past 6 months.

[4] Higher score indicates more friends who smoke cigarettes.

[5] Respondents were asked four questions regarding their religious beliefs and practices: (1) number of times attended religious services in the past year, (2) importance of religious beliefs, (3) influence of religious beliefs on decisionmaking, and (4) importance of friends sharing religious beliefs. Higher scores indicate more religiosity.

The results from this model differed from the results shown in column 2 in that race/ethnicity in the model in column 3 was no longer a predictor of underreporting, even for persons in the "other" category. Friends' cigarette smoking and passive tobacco smoke exposure were significant predictors in both models, and the values of the odds ratios were also very close. This suggests that the composition of the groups of underreporters and true users in the models in the second and third columns of Table 4.16 was very similar.

4.8.2 18 to 25 Year Olds

The 18- to 25-year-old age group had somewhat fewer 7-day underreporters (n = 150) than did the 12- to 17-year-old age group. The same models were used again to compare underreporters with true nonusers and true users at the two cutoff concentrations (i.e., 100 ng/mL and 500 ng/mL). The first column of Table 4.17 compares the underreporters with the true nonusers. The five significant predictors were gender, receipt of the appeal, truthfulness, passive tobacco smoke exposure, and friends' cigarette smoking. Females were less likely than males to underreport.

In addition, those who did not receive the appeal to be truthful were more likely to underreport compared with those who received the appeal. Similarly, respondents who admitted being less than completely truthful in some answers to drug-related questions were about 2.5 times more likely to underreport compared with those who reported being completely truthful in answering the drug questions.

Passive exposure and friends' cigarette smoking also made a difference for this age group, suggesting that the likelihood of underreporting increased with more frequent exposure to ETS or greater numbers of friends who smoked cigarettes. In contrast to the youth group, the 18 to 25 year olds did not show differences between underreporters and true nonusers by region or religiosity.

The second model compares underreporters with true users using the 100 ng/mL cutoff. The results are presented in the second column of Table 4.17. Unlike the model for true nonusers, blacks were more likely than whites to underreport in this model comparing underreporters with true users at the 100 ng/mL cutoff concentration. As in the model comparing underreporters with true nonusers, receipt of the appeal also was a significant predictor of underreporting in this second model. However, truthfulness in answering the drug questions was not a significant predictor in the second model. The odds ratio of the religiosity variable indicated that increased religiosity slightly increased the likelihood of underreporting.

The variables for passive tobacco smoke exposure and friends' cigarette smoking also were significant, but their direction was reversed from the pattern seen among true nonusers in column 1. This was consistent with the youth group. In the column 2 model, more frequent exposure to ETS and increased numbers of friends who smoked cigarettes reduced the likelihood of underreporting. Stated another way, among participants who tested positive for cotinine at the 100 ng/mL cutoff concentration, those with more frequent exposure to ETS and more friends who smoked cigarettes were less likely to underreport. The third model, shown in column 3 of Table 4.17, compares underreporters with true users at the 500 ng/mL cutoff concentration.

Table 4.17 Tobacco Use: Logistic Regression Models Predicting 7-Day Underreporting among Young Adults Aged 18 to 25 at Varying Cotinine Cutoff Concentrations

Model Covariate	UR Versus TNU		UR Versus TU (Cotinine Cutoff 100 ng/mL)		UR Versus TU (Cotinine Cutoff 500 ng/mL)	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Race						
White	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Black	1.59	(0.97-2.61)	1.93[a]	(1.08-3.45)	1.80	(0.95-3.39)
Other	0.69	(0.25-1.94)	0.92	(0.31-2.76)	0.89	(0.30-2.63)
Gender						
Male	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Female	0.57[b]	(0.38-0.87)	0.84	(0.55-1.29)	0.77	(0.49-1.22)
Region						
West	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Midwest	1.03	(0.48-2.20)	0.65	(0.27-1.57)	0.61	(0.24-1.55)
South	1.36	(0.66-2.81)	0.80	(0.40-1.58)	0.71	(0.35-1.44)
Northeast	1.93	(0.95-3.90)	1.35	(0.69-2.65)	1.29	(0.64-2.58)
Receipt of Appeal						
Received Appeal	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Did Not Receive Appeal	1.62[a]	(1.08-2.43)	1.63[a]	(1.06-2.51)	1.98[b]	(1.24-3.14)
Truthfulness						
Completely Truthful	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Not at All/Somewhat/ Mostly Truthful	2.46[a]	(1.16-5.21)	1.44	(0.74-2.81)	1.31	(0.66-2.60)
Privacy of Interview[1]	0.96	(0.85-1.09)	0.99	(0.85-1.15)	1.01	(0.84-1.21)
Difficulty in Answering Drug-Related Questions[2]	1.07	(0.98-1.17)	0.98	(0.89-1.07)	0.97	(0.88-1.06)
Passive Tobacco Smoke Exposure in Past 6 Months[3]	1.49[b]	(1.16-1.92)	0.51[b]	(0.40-0.65)	0.47[b]	(0.36-0.62)
Friends Smoking Cigarettes[4]	1.61[b]	(1.13-2.30)	0.61[a]	(0.40-0.93)	0.58[a]	(0.37-0.90)
Religiosity[5]	0.99	(0.93-1.07)	1.09[a]	(1.01-1.17)	1.08[a]	(1.01-1.16)
	n = 150/943 χ^2 = 78.76 df = 13		n = 150/561 χ^2 = 98.44 df = 13		n = 150/465 χ^2 = 107.84 df = 13	

CI = confidence interval.

ng/mL = nanograms per milliliter.

Note: Underreporters (UR) self-report negative and have a positive urine test, true nonusers (TNU) self-report negative and have a negative urine test, and true users (TU) self-report positive and have a positive urine test. The 7-day self-reporting is based on responses to repeat questions pertaining to cigarette or cigar use only. Those who reported using quit-smoking products in the past 30 days were excluded.

[a] Statistically significant at the 0.05 level.

[b] Statistically significant at the 0.01 level.

[1] Higher score indicates less privacy during interview.

[2] Respondents were asked four questions on their difficulties in (1) understanding the drug-related questions and (2) remembering the drug-related information, as well their (3) clarity of memories regarding drug-related information and (4) how often they made a best guess when answering the drug-related questions. Higher scores indicate more difficulty.

[3] Higher score indicates more frequent exposure to persons smoking cigarettes or other tobacco products in the past 6 months.

[4] Higher score indicates more friends who smoke cigarettes.

[5] Respondents were asked four questions regarding their religious beliefs and practices: (1) number of times attended religious services in the past year, (2) importance of religious beliefs, (3) influence of religious beliefs on decisionmaking, and (4) importance of friends sharing religious beliefs. Higher scores indicate more religiosity.

Results were almost identical to those for the second model with the 100 ng/mL cutoff. Again, receipt of appeal, passive exposure to ETS, friends' cigarette smoking, and religiosity all were significant, with odds ratios similar to those in the 100 ng/mL cutoff model shown in column 2.

Here, for the first time in the multivariate context, we find evidence for the effect of the "appeal." Young adults who received the (long) appeal explaining the importance of the study were less likely to underreport than those who did not receive the appeal. In all three models, those who did not receive the appeal were 1.6 to 1.9 times more likely to underreport compared with those who received the appeal.

4.8.3 Summary of Tobacco Regression Model Results for Underreporters

These analyses provide important information on the correlates of tobacco underreporting. Most notable is the fact that underreporters did not differ greatly from true users on these predictor variables, and, surprisingly, they were not even all that much different on these predictors than the true nonusers. However, as noted previously, some caution should be used in interpreting these results because a confirmation test was not used to confirm specimens testing positive on the immunoassay screening test for cotinine. Further, based on a review of the literature and the results from urine specimens and self-report in this study, it is not clear what cutoff concentration should be used to identify a specimen as positive.

For the comparison of underreporters with true nonusers, the logistic regression models for both the youth and young adult groups showed that underreporting of tobacco use was predicted by ETS. There were some differences as well between the two age groups. Logistic regression models showed that underreporting was lower in the West than in the South and Northeast among the 12 to 17 year olds. The receipt of the appeal did not affect underreporting in the youth group.

5. Marijuana

5.1. Urine Testing for Marijuana and Its Window of Detection

This chapter examines the validity of self-reported use of marijuana, the most widely used illicit drug in the United States. Marijuana is defined as any product made from the plant cannabis sativa and containing the active ingredient delta-9-tetrahydrocannabinol (THC). The most common forms of marijuana used in the United States are the plant leaves and hashish, the concentrated plant resin. The urine test for marijuana screens for cannabinoids and confirms for a metabolite of THC, delta-9-tetrahydrocannabinol-9-carboxylic acid (carboxy-THC). Urine testing is unable to distinguish between the use of marijuana and hashish.

Marijuana contains over 60 cannabinoids. One cannabinoid, THC, is the major source of the psychoactive effects experienced after marijuana use (Jones, 1980; Turner, 1980). THC is metabolized by the body to carboxy-THC (Wall, Brine, Pitt, & Perez-Reyes, 1972), which is a viscous, noncrystalline, water-insoluble, but highly fat-soluble compound. Because THC and carboxy-THC are stored in the "fatty" tissues, they may be detected in the urine longer than other drugs of abuse. Smoking is almost exclusively the route by which marijuana is used in the United States. As is true with tobacco, the amount of active substance reaching the bloodstream is dependent, in very large measure, upon dose, the smoking technique employed, and the amount of substance destroyed or decomposed by the high temperature associated with smoking (Jones, 1980).

It is estimated that, when smoked with maximum efficiency, no more than 50 percent of the THC in a marijuana cigarette will be absorbed into the lungs (Davis, McDaniel, Cadwell, & Moody, 1984; Lemberger & Rubin, 1976). The pharmacological effects of marijuana begin almost immediately after smoking, often within minutes, and blood plasma concentrations of THC peak 8 to 10 minutes after smoking begins (Huestis, Henningfield, & Cone, 1992). Marijuana is metabolized rapidly by the body, and THC metabolites are excreted in the urine and feces. Plasma concentrations of THC peak as a marijuana cigarette is smoked and THC is absorbed and distributed to body tissues; the concentrations then drop quickly (half-time of minutes) as smoking ceases and THC continues to be distributed and metabolized; and finally the concentrations slowly decline as the THC is released from tissues back into the plasma and metabolized. Carboxy-THC, the main THC metabolite, is detectable in the plasma within 4 minutes of the beginning of smoking (Huestis et al., 1992; Hunt & Jones, 1980). Traces of carboxy-THC may exist for several days in human plasma and also in the fat and brain after a single administration (Bronson, Latour, & Nahas, 1984; Kelly & Jones, 1992).

Research shows that the excretion rate of cannabinoids is quite variable. This can result in dramatic fluctuations in the amount of marijuana metabolites in the urine. A review of the literature illustrates that researchers use various consumption levels and time windows to define use (Akinci, Tarter, & Kirisci, 2001; Katz, Webb, Gartin, & Marshall, 1997; Murphy et al., 2000). Carboxy-THC is detectable in the urine for 1 to 3 days after casual use and for 7 or more days after moderate use (Cone, 1997). Higher doses through either increased frequency of use or increased potency of the marijuana may extend the window of detection. The elimination phase of urinary THC metabolites may be extended for up to 30 days after cessation of use among

heavy marijuana smokers (Cone, 1997; Johannson & Halldin, 1989). There also have been cases of marijuana metabolites appearing months after the cessation of active use (Dackis, Pottash, Annitto, & Gold, 1982; Ellis, Mann, Judson, Schramm, & Tashchian, 1985). In addition, the cutoff concentration will affect the length of time that marijuana use may be detected in urine.

In a study that compared self-report data gathered through audio computer-assisted self-interviewing (ACASI) methods and urinalysis among human immunodeficiency virus (HIV)-positive and high-risk adolescents, Murphy et al. (2000) used a screening cutoff concentration of 100 nanograms per milliliter (ng/mL) and compared the urinalysis results with past 2 day, 5 day, and 7 day self-reports. Concordance of urinalysis and self-reports was greater at 5 days and 7 days than at 2 days, indicating that the window of detection for the test was greater than 2 days or that individuals were underreporting their past 2 day use. To avoid recall bias, Akinci et al. (2001) compared past 48 hour self-reports among 15- to 18-year-old males and their parents with urine screens. The detection level was set at 41 ng/mL using an E-Z screen test kit. Of the 200 respondents, a scant 19 tested positive. Of these, 13 self-reported use in the past 48 hours. Among those testing negative, 20 reported use in the past 48 hours. In summary, 13 percent of the respondents had incongruent self-reports and urinalysis results, leading the researchers to conclude that caution is needed in interpreting youths' self-reports of marijuana use.

There have been several studies examining the effect of passive inhalation of marijuana on urine test outcomes (Cone & Johnson, 1986; Cone et al., 1987; Law, Mason, Moffat, & King, 1984). These studies have shown that if an individual is exposed to marijuana smoke under extreme conditions, a positive urinalysis result may be obtained if the cutoff for the urinalysis is sufficiently sensitive.

Dronabinol, a synthetic THC (Marinol®), is available by prescription and is used as an appetite stimulant or antiemetic. As noted in the tables in this chapter, respondents reporting legitimate Marinol® use were excluded from the marijuana study results.

5.2. Self-Reported Marijuana Use: Comparison of Responses to Core Questions and Urinalysis

Table 5.1 presents the results of a comparison of self-reported use of marijuana in the past 30 days in the core questions with the urine test results. Findings in this table are based on data from those respondents with a valid urine specimen who responded to the core questions on marijuana use. Percentages are for persons aged 12 to 25 in the civilian, noninstitutionalized population of the coterminous United States as a whole.

The two-by-two table shows that the vast majority of persons in this age group—82.9 percent (cell A)—would be estimated not to have used marijuana in the past 30 days and to have tested negative. An additional 4.4 percent were estimated to have self-reported not using marijuana but would have tested positive (cell B). Cell C shows that 5.8 percent reported using marijuana in the past 30 days but did not test positive, and cell D indicates that 6.9 percent self-reported using marijuana and tested positive. The overall congruence between self-report and urinalysis was 89.8 percent (cells A + D).

Table 5.1 Marijuana Use: Comparison of Responses to 30-Day Self-Report *Core* Questions and Urinalysis

(*n* = 3,760)	A	Urinalysis (-)	Urinalysis (+)	B
30-Day Self-Report (Core Questions)	(-)	82.9%	4.4%	(3,285)
	(+)	5.8%	6.9%	(475)
	C	(3,342)	(418)	D

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
163.60	1	<0.0001	0.609	0.935	0.517

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of marijuana in the core questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: A screening cutoff concentration of 30 nanograms per milliliter (ng/mL) for cannabinoids and a confirmatory cutoff concentration of 2 ng/mL for carboxy-THC were used to identify a specimen as positive for marijuana.

Based on the available self-report data, the prevalence rate for past month marijuana use was 12.7 percent (cells C + D). Based on the urine data (i.e., if only the biological data were available), however, the prevalence rate would have been 11.3 percent (cells B + D). The respondents in cell C may have reported their use correctly if they used marijuana in the past 30 days but outside of the window of detection for urine testing, depending on their frequency of marijuana use or the potency of the marijuana that they used. Therefore, adding the positive self-reports and positive urinalyses would suggest that the overall prevalence rate for past 30 day marijuana use among young people aged 12 to 25 in the coterminous United States could have been as high as 17.1 percent, although some respondents in cell C may have overreported their use.

The sensitivity level, or the proportion of those testing positive for THC metabolites who reported using marijuana in the past 30 days, was 0.61. This was substantially lower than the comparable level for tobacco, which was 0.80 (see Table 4.2). The specificity, which measures the proportion of those testing negative who reported no marijuana use in the past 30 days, was 0.94, which was somewhat higher than that for tobacco (0.86). Thus, the specificity was quite good, but the sensitivity was comparatively low. The Cochran-Mantel-Haenszel (CMH) chi-square test was significant, and the kappa value of 0.52 showed moderate agreement between self-report and urinalysis for marijuana.

Focusing on those persons aged 12 to 25 who tested positive for marijuana in their urine, 60.9 percent reported use of marijuana in the past month (which is the sensitivity level reported above). Another 9.6 percent reported marijuana use more than 1 month ago, but in the past year; and 11.6 percent reported using marijuana more than a year ago. These data suggest that about 20 percent of recent marijuana users may be willing to report less recent use in a survey, but not use within the past month. The remaining 17.9 percent of those testing positive indicated they had never used marijuana. This suggests that about one in six recent marijuana users may be unwilling to report their recent use. Among those who might underreport their use, about half would indicate that they never used. These findings were similar to those for tobacco use (Section 4.2.2).

Those who self-reported use in the past month were asked to report the number of days that they used marijuana in that period. Among past month marijuana users, 11.8 percent used on 1 to 2 days. The average carboxy-THC concentration generally increased as the number of days of use increased. The highest average carboxy-THC concentrations were found among those reporting smoking marijuana on all 30 days (90 ng/mL), followed by those who reported smoking on 20 to 29 days (89 ng/mL). Those who reported smoking on only 1 to 2 days in the past month had an average carboxy-THC concentration of 78 ng/mL, although carboxy-THC was not detected in most of those specimens.

5.3. Self-Reported Marijuana Use: Comparison of Responses to Core and Repeat Questions

Table 5.2 presents the results of self-reported marijuana use in the past 30 days in the core versus the repeat questions toward the end of the interview. The two-by-two table shows that 86.0 percent were classified as not using marijuana in the past 30 days in both the core and repeat questions (cell A), and 11.3 percent reported past month use in both questions (cell D). However, 1.2 percent reported past month use in the core but not in the repeat questions (cell B), and 1.5 percent reported use in the repeat questions but not in the core (cell C). In total, then, nearly 3 percent had discrepant data between the core and repeat questions. Some of the 1.5 percent who did not report past month use in the core questions and reported use in the repeat questions might have been influenced by the appeal. It is less easy to explain the 1.2 percent who reported use in the core questions, and changed their responses to deny past month use in the repeat questions.

5.4. Self-Reported Marijuana Use: Comparison of Responses to Repeat Questions and Urinalysis

Table 5.3 compares self-reported use in the past 30 days in the repeat questions with the urinalysis results. The table looks fairly similar to that comparing the core questions and urinalysis. Overall congruence (cells A + D) was nearly the same in both the repeat and core questions (90.4 and 89.8 percent, respectively). The percentage in cell B that did not report use in the past month despite having a positive urine test result (i.e., potential underreporters) was slightly higher in the core (4.4 percent) than in the repeat questions (4.1 percent). Similarly, the percentage in cell C that reported use in the past month despite having a negative urine test result (i.e., potential overreporters) was somewhat higher in the core questions (5.8 percent) than in the

Table 5.2 Marijuana Use: Comparison of Responses to 30-Day Self-Report *Core* and *Repeat* Questions

(*n* = 4,439)	A	30-Day Self-Report (Core Questions) (-)	30-Day Self-Report (Core Questions) (+)	B
30-Day Self-Report (Repeat Questions) (-)		86.0%	1.2%	(3,900)
30-Day Self-Report (Repeat Questions) (+)		1.5%	11.3%	(539)
	C	(3,903)	(536)	D

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
272.40	1	<0.0001	0.904	0.983	0.877

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those reporting past 30 day use of marijuana in the *core* or *repeat* questions.

repeat questions (5.5 percent). The repeat questions had similar sensitivity (0.64) and kappa (0.55) to the core questions, suggesting little difference in accuracy between the two.

The repeat questions also asked about use in more recent time periods than did the core questions. Of those testing positive, about half (50.6 percent) reported using marijuana in the past 3 days. Another 6.7 percent reported using more than 3 days ago but within the past 7 days. There was no correlation between carboxy-THC concentrations and recency of use.

The highest average carboxy-THC concentrations were observed among those who reported using marijuana more than 30 days ago but within the past 6 months. Interestingly, the highest median carboxy-THC concentrations were observed among persons who reported that their most recent use was more than 30 days ago but within the past 6 months (100 ng/mL) and among those who reported using more than a year ago (98 ng/mL). Approximately one in five of those who tested positive reported never using marijuana. Their average carboxy-THC concentration was 84.8 ng/mL.

Because the repeat questions were preceded by the appeal, it was important to determine the impact of the appeal on the congruence between self-report and urinalysis. Table 5.4 presents a pair of two-by-two tables: the first presents data based on those respondents who received the appeal, and the second presents data based on those who did not receive the appeal. The sensitivity was much higher for those receiving the appeal (0.71) than for those not receiving the appeal (0.57), showing that those who tested positive were more likely to report their use when

Table 5.3 Marijuana Use: Comparison of Responses to 30-Day Self-Report *Repeat* Questions and Urinalysis

(*n* = 3,748)		Urinalysis (-)	Urinalysis (+)	
		A		**B**
30-Day Self-Report (Repeat Questions)	(-)	83.2%	4.1%	(3,280)
	(+)	5.5%	7.2%	(468)
		C (3,334)	(414)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
188.77	1	<0.0001	0.639	0.938	0.546

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of marijuana in the repeat questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: A screening cutoff concentration of 30 nanograms per milliliter (ng/mL) for cannabinoids and a confirmatory cutoff concentration of 2 ng/mL for carboxy-THC were used to identify a specimen as positive for marijuana.

they received the appeal. The percentage of potential underreporters in cell B was lower among those who received the appeal (3.1 percent) compared with 5.0 percent for those who did not receive the appeal. The proportion of potential overreporters in cell C also was smaller for those who received the appeal (5.2 vs. 5.9 percent, respectively). Specificity was unaffected by the appeal (0.94 for the appeal vs. 0.93 for those who did not receive the appeal). The CMH chi squares on both tables were significant, and the kappa values for both conditions showed moderate agreement between self-report and urinalysis. However, the kappa for the appeal condition (0.60) was much improved over the condition with no appeal (0.49).

Tables 5.5 and 5.6 are a pair of two-by-two tables further examining the relationship between self-reported 30-day use and urinalysis results in the core and repeat questions. Table 5.5 presents data based on reports of marijuana use in the past 30 days in *both* the core and repeat questions (i.e., those who reported use on both core and repeat are included on the X axis as "+"). Table 5.6 presents data based on reports of use in the past 30 days in *either* the core or repeat questions. The proportion who would be considered to have underreported use was greater under the requirement in Table 5.5 that use in the past 30 days had to be self-reported in both the core and repeat questions (4.8 percent), compared with the percentage under the requirement in Table 5.6 that self-reported use in either set of questions was sufficient (3.7 percent). Similarly, there was greater sensitivity when an affirmative report of marijuana use in the past 30 days was

Table 5.4 Marijuana Use: Comparison of Responses to 30-Day Self-Report *Repeat* Questions and Urinalysis, by Receipt of Appeal

Respondent Received Appeal

(*n* = 1,833)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Repeat Questions)	(-)	84.0%	3.1%	(1,600)
	(+)	5.2%	7.7%	(233)
	C	(1,633)	(200)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
107.69	1	<0.0001	0.714	0.942	0.604

Respondent Did Not Receive Appeal

(*n* = 1,915)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Repeat Questions)	(-)	82.4%	5.0%	(1,680)
	(+)	5.9%	6.7%	(235)
	C	(1,701)	(214)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
86.02	1	<0.0001	0.573	0.934	0.491

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use of marijuana. Respondents then listen to one of two possible introductions to the next series of questions in which respondents are asked again to report their recency of use. One scenario very broadly introduces the next series of questions. This is defined as *repeat questions without appeal.* The second scenario gives a broad overview of the study and emphasizes the importance of the respondent's responses. An appeal is made to the respondent to answer the questions as honestly as he or she can. This is defined as *repeat questions with appeal*.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of marijuana in the repeat questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: A screening cutoff concentration of 30 nanograms per milliliter (ng/mL) for cannabinoids and a confirmatory cutoff concentration of 2 ng/mL for carboxy-THC were used to identify a specimen as positive for marijuana.

Table 5.5 Marijuana Use: Comparison of Responses to 30-Day Self-Report *Core and Repeat* Questions and Urinalysis

(*n* = 3,748)	A	Urinalysis (-)	Urinalysis (+)	B
30-Day Self-Report (Core and Repeat Questions)	(-)	83.9%	4.8%	(3,328)
	(+)	4.9%	6.5%	(420)
	C	(3,334)	(414)	D

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
164.06	1	<0.0001	0.574	0.945	0.518

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of marijuana in the core and *repeat* questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: A screening cutoff concentration of 30 nanograms per milliliter (ng/mL) for cannabinoids and a confirmatory cutoff concentration of 2 ng/mL for carboxy-THC were used to identify a specimen as positive for marijuana.

sufficient in either the core or repeat questions (0.67) than it was when a self-report of use in the past 30 days was required in both places (0.57). The specificity was similar for both instances. The CMH chi squares for both tables were significant, and the kappa values both indicated moderate agreement. Results of these tables show that a better fit was obtained by combining the responses to the core or repeat questions, although a higher percentage were classified as overreporters (cell C) when this strategy was adopted.

5.4.1 Passive Exposure

Because marijuana typically is a smoked substance, the survey included questions about passive exposure to environmental marijuana smoke to determine any correlation between passive exposure and testing positive. However, scientific studies have shown that, unless there is significant, long-term exposure to extensive amounts of marijuana smoke in a closed, small environment, a detectable concentration of the drug metabolite in the urine cannot be achieved. It is, therefore, highly unlikely that passive exposure is a realistic explanation for positive urine marijuana tests. Among those respondents testing positive for marijuana and reporting no use in the past 30 days, the average carboxy-THC concentrations were much higher than the 2 ng/mL cutoff concentration used to confirm that the specimens were positive. The correlation of claims

Table 5.6 Marijuana Use: Comparison of Responses to 30-Day Self-Report *Core or Repeat* Questions and Urinalysis

(*n* = 3,752)	A	Urinalysis (-)	Urinalysis (+)	B
30-Day Self-Report (Core or Repeat Questions)	(-)	82.2%	3.7%	(3,229)
	(+)	6.5%	7.7%	(523)
	C	(3,335)	(417)	D

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
191.69	1	<0.0001	0.673	0.927	0.543

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of marijuana in the core or *repeat* questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: A screening cutoff concentration of 30 nanograms per milliliter (ng/mL) for cannabinoids and a confirmatory cutoff concentration of 2 ng/mL for carboxy-THC were used to identify a specimen as positive for marijuana.

of passive exposure and testing positive is considered in multivariate logistic regression models later in this chapter.

5.4.2 Comparisons of Self-Report and Urinalysis Results, by Age

Because there were differences in the validity of self-reported tobacco use by age group, it is important to determine whether these differences also occurred with marijuana use. Unlike tobacco, however, purchasing and using marijuana are illegal regardless of age.

Table 5.7 presents two sets of two-by-two tables comparing self-reported use in the core or repeat questions with the urinalysis results, by age group (i.e., 12 to 17 and 18 to 25). The overall congruence rates were very similar, although fewer youths aged 12 to 17 were positive, as might be expected. The sensitivity level showed that the proportion of those testing positive who self-reported use was slightly higher among the young adults aged 18 to 25, and the specificity was slightly higher among the youths, but the differences were small. The kappa value was higher for the older than the younger age group. This suggests that the older age group was slightly more likely to report marijuana use. This same conclusion was drawn for tobacco, but as noted above, marijuana is typically illegal for purchase regardless of age.

Table 5.7 Marijuana Use: Comparison of Responses to 30-Day Self-Report *Core or Repeat* Questions and Urinalysis, by Age Group

12 to 17 Years

(*n* = 1,950)

		Urinalysis		
		A (-)	(+)	**B**
30-Day Self-Report (Core or Repeat Questions)	(-)	87.8%	2.4%	(1,751)
	(+)	5.9%	3.9%	(199)
		C (1,825)	(125)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
64.00	1	<0.0001	0.621	0.937	0.440

18 to 25 Years

(*n* = 1,802)

		Urinalysis		
		A (-)	(+)	**B**
30-Day Self-Report (Core or Repeat Questions)	(-)	77.6%	4.8%	(1,478)
	(+)	6.9%	10.7%	(324)
		C (1,510)	(292)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
151.20	1	<0.0001	0.691	0.918	0.577

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 30 day use of marijuana in the core or *repeat* questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: A screening cutoff concentration of 30 nanograms per milliliter (ng/mL) for cannabinoids and a confirmatory cutoff concentration of 2 ng/mL for carboxy-THC were used to identify a specimen as positive for marijuana.

5.5. Comparison of Self-Report and Urinalysis Results for 7-Day and 3-Day Windows

Tables 5.8 and 5.9 compare self-reported marijuana use in the past 7 days and past 3 days from the repeat questions with the urinalysis results. Although heavy users can test positive for 30 days and there are some reports of people testing positive even longer, the research literature suggests that the 3-day and 7-day reporting timeframes are more comparable with the carboxy-THC window of detection. This is reflected in the two-by-two tables for the repeat questions (Tables 5.3, 5.8, and 5.9). The percentage classified as overreporters (cell C) was lowest for the 3-day reference period in Table 5.9 (1.7 percent), was 2.7 percent for the 7-day reference period in Table 5.8, and was 5.5 percent for the 30-day reference period in Table 5.3. This also is reflected in the overall congruence rates from cells A and D combined (i.e., 93.0 percent for the 3-day reference period, 92.5 percent for the 7-day reference period, and 90.4 percent for the 30-day reference period). As with other drugs, these high congruence rates were dominated largely by those reporting no use and testing negative.

Table 5.8 Marijuana Use: Comparison of Responses to 7-Day Self-Report *Repeat* Questions and Urinalysis

(*n* = 3,748)

		Urinalysis		
	A	(-)	(+)	**B**
7-Day Self-Report (Repeat Questions)	(-)	86.0%	4.8%	(3,412)
	(+)	2.7%	6.5%	(336)
	C	(3,334)	(414)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
167.89	1	<0.0001	0.579	0.969	0.594

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 7 day use of marijuana in the repeat questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: A screening cutoff concentration of 30 nanograms per milliliter (ng/mL) for cannabinoids and a confirmatory cutoff concentration of 2 ng/mL for carboxy-THC were used to identify a specimen as positive for marijuana.

Table 5.9 Marijuana Use: Comparison of Responses to 3-Day Self-Report *Follow-Up or Repeat* Questions and Urinalysis

(*n* = 3,749)

		Urinalysis		
	A	(-)	(+)	**B**
3-Day Self-Report (Follow-Up or Repeat Questions)	(-)	86.9%	5.2%	(3,469)
	(+)	1.7%	6.1%	(280)
	C	(3,334)	(415)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
163.05	1	<0.0001	0.540	0.980	0.601

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those with a valid urine specimen and reporting past 3 day use of cocaine on the *follow-up* or *repeat* questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: A screening cutoff concentration of 30 nanograms per milliliter (ng/mL) for cannabinoids and a confirmatory cutoff concentration of 2 ng/mL for carboxy-THC were used to identify a specimen as positive for marijuana.

The percentage classified as underreporters (cell B) was 4.8 percent for the 7-day reference period and 5.2 percent for the 3-day reference period. Sensitivity levels for the 3-day and 7-day periods were 0.54 and 0.58, respectively, compared with the level of 0.64 for the 30-day period from Table 5.3. This finding suggests that underreporting increased as the reference period of interest moved closer to the interview date. Specificity was somewhat higher for the 3-day and 7-day reference periods than the 30-day period but did not differ appreciably between the 7-day and 3-day timeframes (0.98 for the 3-day period, 0.97 for the 7-day period, and 0.94 for the 30-day period). The kappa statistic also was somewhat higher for the 3- and 7-day reference periods compared with the statistic for the 30-day period, but did not differ between the 7-day and 3-day periods (0.60 for the 3-day period, 0.59 for the 7-day period, and 0.55 for the 30-day period).

5.6. Varying Screening Cutoff Concentrations and Self-Report Timeframes

The Validity Study obtained information on concentrations of carboxy-THC, so it is possible to examine the efficiency of various cutoff concentrations to designate specimens as positive by comparing the distribution of concentrations in relationship with the self-reported use. Table 5.10 examines the recency of marijuana use in the repeat questions and various screening concentrations. Most of those with less than 30 ng/mL of carboxy-THC (i.e., the

Table 5.10 Marijuana Use: Comparison of Responses to Self-Report *Repeat* Questions on Recency of Use at Varying Urinalysis Screening Cutoff Concentrations: Percentages

Recency of Use	Screening Concentrations (*n* = 3,772)				
	< 30 ng/mL (*n* = 3,324)	30 - 49 ng/mL (*n* = 61)	50 - 74 ng/mL (*n* = 63)	75 - 99 ng/mL (*n* = 112)	≥ 100 ng/mL (*n* = 212)
No Past Month Use	93.6	30.7	49.5	32.7	39.5
Within 3 Days	1.5	49.8	38.7	52.2	49.1
More Than 3 Days, but within 7 Days	1.6	9.5	5.0	6.8	6.1
More Than 7 Days, but within 30 Days	3.1	9.1	6.7	5.7	4.9
Don't Know/Refusal	0.2	0.8	0.0	2.6	0.4

ng/mL = nanograms per milliliter.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages have been weighted to be representative of persons aged 12 to 25 in the civilian, noninstitutionalized population in the coterminous United States. Percentages are based on those reporting use on *repeat* questions and submitting a valid urine specimen.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

screening cutoff concentration used in this study) in their urine (93.6 percent) reported no marijuana use in the past month. Using a screening cutoff concentration of 100 ng/mL, almost half (49.1 percent) reported use within the past 3 days.

5.7. Overreporters

The vast majority of respondents (88.7 percent) tested negative for marijuana in their urine. About 1 in 17, or 5.8 percent of this group, self-reported marijuana use in the past 30 days. They may have used outside the 1- to 7-day timeframe for which carboxy-THC is generally thought to be detectable. As the time increased since use, the chances of identifying marijuana use within the window of detection decreased.

The 3-day and 7-day timeframes were examined to further analyze the issue of overreporting. The Validity Study screening test cutoff concentration of 30 ng/mL and confirmatory test cutoff concentration of 2 ng/mL resulted in 66 overreporters for the 3-day timeframe and 99 overreporters for the 7-day timeframe. Although these respondents reported marijuana use, their urine tests were negative. More than 60 percent of the overreporters also reported being daily or frequently exposed to environmental marijuana smoke.

Among persons who reported use between 7 and 30 days in the past month, 16.4 percent tested positive. An explanation for these positive results outside the expected detection window could be the frequency of use. However, there were still unexplained cases (2.8 percent) of respondents who reported marijuana use in the past 3 days and had no traces of carboxy-THC in their urine.

Reasons other than those noted above can account for negative drug tests for individuals who reported drug use. These include high screening and confirmation cutoff concentrations, incorrect urine test results, specimen misidentification, or random errors. In addition, physiological characteristics (high metabolism, low body fat) could have contributed to unexpected negative urine test results. In these situations, respondents' reports that they had used the drug within a given reference period might be correct. Other reasons could include incorrect self-reports due to unclear memory, difficulties in understanding drug-related questions, or untruthfulness. Lowering the screening cutoff concentration may have identified additional positive specimens, thereby reducing the number of overreporters. With regard to the second possible explanation of physiological issues, however, it is not possible for any study of this nature to examine physiological differences among the respondents. Variables controlling for difficulties in answering drug-related questions and concern about the privacy of the interview were included in the multivariate statistical models described below, but provided only limited insight about the nature of overreporting.

As in Chapter 4 on tobacco use, logistic regression models were used to determine whether there were any demographic, behavioral, environmental, or other variables that might distinguish the overreporters from true nonusers and true users. Two groups of models were developed for each age group. Independent variables were largely based on the same variables as in the tobacco logistic regressions with minimal changes. The complete results of these models are provided in Tables 5.11 and 5.12. Models for 3-day and 7-day overreporters were compared with true nonusers and true users. Separate models were run for youths aged 12 to 17 and young adults aged 18 to 25.

Marijuana overreporters were defined as respondents who reported use in the past 3 days or past 7 days but tested negative. These time periods were used because they matched the window of detection for carboxy-THC in urine more closely than the 30-day reporting timeframe. The categorical independent variables included in the models were race/ethnicity, gender, region, receipt of appeal, and truthfulness. Ordinal independent variables included in the models were privacy of the interview, exposure to marijuana smoke in the past 6 months, and friends' use of marijuana. Two composite indices also were created from multiple variables and included in the models: difficulty in answering the drug-related questions and religiosity.

5.7.1 12 to 17 Year Olds

The first two columns of Table 5.11 examine 3-day and 7-day overreporters among youths aged 12 to 17 compared with true nonusers. A total of 30 youths were 3-day overreporters, and 36 were 7-day overreporters. The first column presents the 3-day overreporters compared with the true nonusers, and the second column presents the 7-day overreporters compared with the true nonusers. Two significant independent variables are in both models: passive exposure to environmental marijuana smoke and having friends who use marijuana. Compared with true nonusers, the likelihood of overreporting increased with claims of increased passive exposure to environmental marijuana smoke and increased numbers of friends who used marijuana. Odds ratios for these two environmental measures ranged from 2.3 to 4.4; these effects were particularly significant because there is collinearity between these two predictors.

Table 5.11 Marijuana Use: Logistic Regression Models Predicting 3-Day and 7-Day Overreporting among Youths Aged 12 to 17

Model Covariate	3-Day OR vs. TNU		7-Day OR vs. TNU		3-Day OR vs. TU		7-Day OR vs. TU	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Race								
White	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Black	0.23[a]	(0.05-0.97)	0.30	(0.06-1.43)	1.39	(0.25-7.65)	2.24	(0.49-10.35)
Other	1.72	(0.33-8.90)	1.03	(0.22-4.76)	1.13	(0.19-6.52)	0.23	(0.03-1.57)
Gender								
Male	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Female	1.78	(0.70-4.54)	1.18	(0.47-2.93)	8.32[b]	(2.38-29.09)	4.21[a]	(1.35-13.17)
Region								
West	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Midwest	0.48	(0.12-1.86)	0.78	(0.20-2.99)	0.50	(0.10-2.46)	0.85	(0.20-3.60)
South	0.86	(0.20-3.78)	0.49	(0.11-2.16)	0.32	(0.05-2.19)	0.28	(0.06-1.23)
Northeast	1.10	(0.25-4.83)	0.98	(0.25-3.86)	0.93	(0.19-4.47)	1.06	(0.29-3.89)
Receipt of Appeal								
Received Appeal	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Did Not Receive Appeal	1.00	(0.42-2.41)	0.87	(0.40-1.91)	0.51	(0.14-1.86)	0.54	(0.19-1.53)
Truthfulness								
Completely Truthful	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Not at All/Somewhat/ Mostly Truthful	0.18[a]	(0.04-0.90)	0.90	(0.25-3.25)	0.14[a]	(0.02-0.75)	0.39	(0.08-1.97)
Privacy of Interview[1]	0.84	(0.62-1.13)	0.98	(0.80-1.20)	1.02	(0.74-1.40)	1.12	(0.85-1.48)
Difficulty in Answering Drug-Related Questions[2]	1.27[b]	(1.09-1.47)	1.10	(0.94-1.29)	1.10	(0.88-1.38)	1.01	(0.82-1.23)
Passive Marijuana Smoke Exposure in Past 6 Months[3]	3.90[b]	(2.07-7.35)	3.72[b]	(2.09-6.63)	0.31[b]	(0.14-0.69)	0.38[a]	(0.17-0.87)
Friends UsingMarijuana[4]	4.39[b]	(2.20-8.75)	2.29[b]	(1.38-3.79)	2.77[a]	(1.11-6.94)	1.14	(0.37-3.48)
Religiosity[5]	0.85[a]	(0.74-0.98)	0.88	(0.78-1.00)	0.99	(0.81-1.20)	0.97	(0.82-1.16)
	n = 30/1,704 χ^2 = 136.05 df = 13		n = 36/1,698 χ^2 = 116.28 df = 13		n = 30/56 χ^2 = 29.30 df = 13		n = 36/65 χ^2 = 25.46 df = 13	

CI = confidence interval.

Note: Overreporters (OR) self-report positive and have a negative urine test, true nonusers (TNU) self-report negative and have a negative urine test, and true users (TU) self-report positive and have a positive urine test. The 3-day self-reporting is based on responses to repeat or follow-up questions. The 7-day self-reporting is based on responses to repeat questions. Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

[a] Statistically significant at the 0.05 level.

[b] Statistically significant at the 0.01 level.

[1] Higher score indicates less privacy during interview.

[2] Respondents were asked four questions on their difficulties in (1) understanding the drug-related questions and (2) remembering the drug-related information, as well their (3) clarity of memories regarding drug-related information and (4) how often they made a best guess when answering the drug-related questions. Higher scores indicate more difficulty.

[3] Higher score indicates more frequent exposure to persons smoking marijuana in the past 6 months.

[4] Higher score indicates more friends who use marijuana.

[5] Respondents were asked four questions regarding their religious beliefs and practices: (1) number of times attended religious services in the past year, (2) importance of religious beliefs, (3) influence of religious beliefs on decisionmaking, and (4) importance of friends sharing religious beliefs. Higher scores indicate more religiosity.

Table 5.12 Marijuana Use: Logistic Regression Models Predicting 3-Day and 7-Day Overreporting among Young Adults Aged 18 to 25

Model Covariate	3-Day OR vs. TNU		7-Day OR vs. TNU		3-Day OR vs. TU		7-Day OR vs. TU	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Race								
White	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Black	0.44	(0.08-2.53)	0.35	(0.09-1.40)	0.29	(0.06-1.39)	0.20[a]	(0.05-0.77)
Other	1.13	(0.34-3.82)	1.10	(0.45-2.74)	2.49	(0.45-13.68)	1.70	(0.36-8.00)
Gender								
Male	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Female	0.87	(0.39-1.93)	0.73	(0.36-1.51)	1.60	(0.68-3.81)	1.71	(0.82-3.56)
Region								
West	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Midwest	1.58	(0.50-5.02)	2.01	(0.68-5.95)	1.60	(0.41-6.22)	2.32	(0.79-6.82)
South	2.44	(0.92-6.51)	2.03	(0.80-5.19)	2.06	(0.65-6.50)	2.02	(0.79-5.17)
Northeast	0.72	(0.26-2.02)	0.94	(0.37-2.37)	0.48	(0.15-1.49)	0.63	(0.24-1.68)
Receipt of Appeal								
Received Appeal	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Did Not Receive Appeal	0.51	(0.20-1.27)	1.26	(0.59-2.69)	0.89	(0.34-2.30)	1.61	(0.76-3.40)
Truthfulness								
Completely Truthful	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Not at All/Somewhat/ Mostly Truthful	0.34	(0.05-2.16)	0.49	(0.14-1.72)	0.39	(0.05-3.20)	0.62	(0.15-2.49)
Privacy of Interview[1]	1.02	(0.75-1.38)	0.96	(0.73-1.25)	1.06	(0.81-1.39)	1.03	(0.83-1.29)
Difficulty in Answering Drug-Related Questions[2]	1.11	(0.96-1.29)	1.07	(0.94-1.21)	1.01	(0.80-1.26)	0.94	(0.79-1.12)
Passive Marijuana Smoke Exposure in Past 6 Months[3]	4.67[b]	(2.72-8.02)	3.77[b]	(2.46-5.78)	0.63	(0.35-1.14)	0.50[b]	(0.31-0.80)
Friends Using Marijuana[4]	3.74[b]	(2.33-5.98)	3.76[b]	(2.39-5.94)	0.98	(0.45-2.18)	0.99	(0.51-1.93)
Religiosity[5]	1.00	(0.89-1.13)	0.95	(0.86-1.06)	1.12	(0.97-1.30)	1.07	(0.93-1.23)
	n = 35/1,442 χ^2 = 117.60 df = 13		n = 60/1,417 χ^2 = 172.99 df = 13		n = 35/157 χ^2 = 21.84 df = 13		n = 60/171 χ^2 = 34.11 df = 13	

CI = confidence interval.

Note: Overreporters (OR) self-report positive and have a negative urine test, true nonusers (TNU) self-report negative and have a negative urine test, and true users (TU) self-report positive and have a positive urine test. The 3-day self-reporting is based on responses to repeat or follow-up questions. The 7-day self-reporting is based on responses to repeat questions. Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

[a] Statistically significant at the 0.05 level.

[b] Statistically significant at the 0.01 level.

[1] Higher score indicates less privacy during interview.

[2] Respondents were asked four questions on their difficulties in (1) understanding the drug-related questions and (2) remembering the drug-related information, as well their (3) clarity of memories regarding drug-related information and (4) how often they made a best guess when answering the drug-related questions. Higher scores indicate more difficulty.

[3] Higher score indicates more frequent exposure to persons smoking marijuana in the past 6 months.

[4] Higher score indicates more friends who use marijuana.

[5] Respondents were asked four questions regarding their religious beliefs and practices: (1) number of times attended religious services in the past year, (2) importance of religious beliefs, (3) influence of religious beliefs on decisionmaking, and (4) importance of friends sharing religious beliefs. Higher scores indicate more religiosity.

Three more variables were significant for 3-day overreporters, but not for 7-day overreporters: race/ethnicity, truthfulness, and difficulty answering drug-related questions. These overreporters were less likely to be black compared with the reference group of white. The likelihood of overreporting also was lower among respondents who indicated that they had been less than completely truthful in answering the drug questions. The likelihood of overreporting also increased somewhat with respondents' perceived difficulty in answering the drug-related questions.

The next two models (columns 3 and 4) in Table 5.11 compare the 3-day and 7-day overreporters with the true users, or those who reported use and tested positive (cell D) in the relevant reference period. There were 56 true users aged 12 to 17 for the 3-day comparison and 65 for the 7-day comparison.

Because there were very few overreporters or true users among youths aged 12 to 17, the significant effects in the models were particularly noteworthy. Two predictors were significant in both the 3-day and 7-day comparisons with true users: gender and passive exposure to marijuana smoke. Females aged 12 to 17 were 4.2 to 8.3 times more likely than males to be overreporters. For the passive exposure variable, the direction of the relationship between passive exposure and overreporting was reversed compared with the models with true nonusers. In the models with overreporters and true users, the likelihood of overreporting decreased as claims of passive exposure to environmental marijuana smoke increased. That is, the overreporters were less likely to report passive exposure to marijuana smoke than the true users.

Two additional variables were significant predictors of overreporting in the past 3 days but not for the past 7 days: truthfulness and friends using marijuana. The direction of the truthfulness variable was the same as in the model that included overreporters and true nonusers. That is, the likelihood of overreporting was lower among respondents who indicated that they had been less than completely truthful in answering the drug questions. The likelihood of overreporting for the 3-day reference period increased as the number of friends who smoked marijuana also increased. On the other hand, friends' marijuana use was not a significant predictor of overreporting in the past 7 days.

5.7.2 18 to 25 Year Olds

The results for the older age group of overreporters aged 18 to 25 are presented in Table 5.12. The models for young adults aged 18 to 25 were similar to the models for youths aged 12 to 17 in terms of absolute numbers of overreporters for the 3-day reference period. A total of 35 young adults aged 18 to 25 were 3-day overreporters and 60 were 7-day overreporters, compared with 30 youths who were 3-day overreporters and only 36 who were 7-day overreporters.

As was the case in the models comparing overreporters with true nonusers aged 12 to 17, claims of passive exposure to environmental marijuana smoke and friends using marijuana were predictors of overreporting in the models comparing overreporters with true nonusers aged 18 to 25. Reported passive exposure also was a significant predictor of 7-day overreporting in the models with true users aged 18 to 25, and the relationship was in the same direction as the models comparing overreporters aged 12 to 17 with true users. That is, the overreporters were less likely to report passive exposure to marijuana smoke than the true users. Unlike the situation

with youths, however, reported passive exposure was not a significant predictor of overreporting in the model comparing 3-day overreporters with true users aged 18 to 25.

Gender was a significant predictor of overreporting among youths in the models with overreporters and true users, but not in the corresponding models for young adults. In the model comparing 3-day overreporters with true nonusers aged 12 to 17, blacks were less likely than whites to be overreporters. The only significant odds ratio for race in the young adult models was in the comparison of 7-day overreporters with true users. In that model, blacks were less likely than whites to be overreporters.

5.8. Underreporters

Underreporters were defined as respondents who reported no marijuana use in the past 7 or 30 days, but who nevertheless tested positive for THC metabolite. Possible reasons for underreporting include fear of being identified as a marijuana user, unclear recall regarding the most recent use of marijuana, untruthfulness, or difficulties in understanding the drug-related questions. Other possible reasons for the apparent underreporting include incorrect urine test results, specimen misidentification, lack of sensitivity at the Validity Study cutoffs, or random errors. In these latter situations, the respondents' self-reports that they had not used marijuana within a given reference period might be correct. Among those testing positive for marijuana, 137 reported no use in the past 30 days on either the core or repeat questions. There were 177 who reported no use in the past 7 days in the repeat questions. The percentage of underreporters increased with the 7-day timeframe. The 30-day timeframe had the least number of underreporters, as would be expected, because carboxy-THC should generally not be detectable in the urine for more than 7 days.

As with the previous logistic regression models, underreporters were compared with the group of true nonusers and the group of true users. The models for underreporters used the same independent variables as those used in the models for overreporters and also were run separately for youths aged 12 to 17 and young adults aged 18 to 25.

The elimination period for marijuana metabolites in some cases (particularly for regular users) can be as much as 30 days. Considering this window of detection, the 7-day and 30-day groups should consist predominantly of true underreporters and can therefore provide interesting insights on the nature of underreporting. Consequently, both 7-day and 30-day models are presented for underreporters.

5.8.1 12 to 17 Year Olds

For youths aged 12 to 17, a total of 57 were 7-day underreporters, and 41 were 30-day underreporters. The ratio of males to females in the group of underreporters was about 3 to 1.

The first two columns of Table 5.13 present the results for the 7-day and 30-day underreporter comparisons with the true nonusers. Three variables were significant in both time periods (reported passive exposure in the past 6 months, privacy of interview, and religiosity), and they each had similar odds ratios for the 7-day and 30-day models. The likelihood of underreporting increased as claims of passive exposure to marijuana increased. As noted previously, it is extremely unlikely that passive exposure could cause a positive urine drug test

Table 5.13 Marijuana Use: Logistic Regression Models Predicting 7-Day and 30-Day Underreporting among Youths Aged 12 to 17

Model Covariate	7-Day UR vs. TNU		30-Day UR vs. TNU		7-Day UR vs. TU		30-Day UR vs. TU	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Race								
White	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Black	1.02	(0.39-2.62)	0.96	(0.35-2.62)	2.07	(0.24-17.56)	1.60	(0.23-11.00)
Other	0.89	(0.28-2.84)	0.16[a]	(0.04-0.70)	0.26	(0.05-1.47)	0.06[b]	(0.01-0.46)
Gender								
Male	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Female	0.41[a]	(0.20-0.83)	0.46	(0.20-1.04)	0.73	(0.21-2.52)	1.10	(0.25-4.85)
Region								
West	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Midwest	0.45	(0.15-1.30)	0.71	(0.21-2.38)	0.23	(0.03-1.87)	1.63	(0.24-11.17)
South	0.72	(0.28-1.88)	0.88	(0.28-2.74)	0.74	(0.15-3.73)	1.13	(0.18-7.32)
Northeast	0.56	(0.27-1.20)	0.65	(0.22-1.90)	0.68	(0.24-1.93)	0.65	(0.13-3.21)
Receipt of Appeal								
Received Appeal	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Did Not Receive Appeal	0.81	(0.42-1.56)	0.60	(0.28-1.25)	1.74	(0.47-6.37)	0.56	(0.18-1.69)
Truthfulness								
Completely Truthful	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Not at All/Somewhat/ Mostly Truthful	2.33[a]	(1.21-4.46)	1.80	(0.82-3.94)	0.59	(0.14-2.48)	0.43	(0.08-2.20)
Privacy of Interview[1]	1.18[a]	(1.02-1.36)	1.18[a]	(1.00-1.38)	1.28	(0.92-1.78)	1.03	(0.81-1.31)
Difficulty in Answering Drug-Related Questions[2]	1.05	(0.95-1.17)	1.09	(0.97-1.23)	1.07	(0.86-1.32)	1.29[a]	(1.06-1.56)
Passive Marijuana Smoke Exposure in Past 6 Months[3]	2.27[b]	(1.46-3.52)	2.26[b]	(1.53-3.34)	0.15[b]	(0.07-0.34)	0.22[b]	(0.10-0.48)
Friends Using Marijuana[4]	1.24	(0.80-1.92)	1.03	(0.56-1.90)	0.55	(0.22-1.36)	0.42	(0.16-1.07)
Religiosity[5]	0.86[b]	(0.78-0.95)	0.84[b]	(0.75-0.94)	0.91	(0.72-1.16)	0.82	(0.67-1.01)
	n = 57/1,698 χ^2 = 93.72 df = 13		n = 41/1,619 χ^2 = 68.10 df = 13		n = 57/65 χ^2 = 69.89 df = 13		n = 41/81 χ^2 = 69.36 df = 13	

CI = confidence interval.

Note: Underreporters (UR) self-report negative and have a positive urine test, true nonusers (TNU) self-report negative and have a negative urine test, and true users (TU) self-report positive and have a positive urine test. The 7-day self-reporting is based on responses to repeat questions. The 30-day self-reporting is based on responses to core or repeat questions. Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

[a] Statistically significant at the 0.05 level.

[b] Statistically significant at the 0.01 level.

[1] Higher score indicates less privacy during interview.

[2] Respondents were asked four questions on their difficulties in (1) understanding the drug-related questions and (2) remembering the drug-related information, as well their (3) clarity of memories regarding drug-related information and (4) how often they made a best guess when answering the drug-related questions. Higher scores indicate more difficulty.

[3] Higher score indicates more frequent exposure to persons smoking marijuana in the past 6 months.

[4] Higher score indicates more friends who use marijuana.

[5] Respondents were asked four questions regarding their religious beliefs and practices: (1) number of times attended religious services in the past year, (2) importance of religious beliefs, (3) influence of religious beliefs on decisionmaking, and (4) importance of friends sharing religious beliefs. Higher scores indicate more religiosity.

for marijuana. These findings suggest that when underreporters are compared with true nonusers, the positive urine test results may reflect drug users who deny use but report passive exposure to marijuana smoke. The likelihood of underreporting also increased as religiosity decreased.

In the 7-day model, females were about 40 percent as likely as males (odds ratio = 0.40) to underreport marijuana use in the past 7 days. Also in the 7-day model with underreporters and true nonusers, truthfulness was a significant predictor of underreporting: Youths who reported being less than completely truthful in answering the drug questions were 2.3 times more likely to underreport than those who indicated that they were completely truthful. In the 30-day model, persons in the "other" race/ethnicity category were less likely than whites to underreport marijuana use.

Privacy level was a significant predictor of underreporting for youths in the models comparing either 7-day or 30-day underreporters with true nonusers. The likelihood of underreporting increased with decreasing levels of privacy based on the presence of an adult or family member. These results support prior research that suggests that the presence of an adult or another family member played an important role in inhibiting youths from reporting their marijuana use when the models controlled for potential confounding effects of other variables.

Results for the comparisons of 12- to 17-year-old underreporters with true users are presented in columns 3 and 4 of Table 5.13; again, separate models are shown for marijuana use in the past 7 days or past 30 days. For the 7-day model, there were slightly more true users (n = 65) than underreporters (n = 57). In contrast, for the 30-day model, there were almost twice as many true users (n = 81) as underreporters (n = 41). These numbers of underreporters for the 7-day model may suggest a hesitancy for youths to self-report recent illegal activity.

One variable was a significant predictor of underreporting in both models: passive exposure to environmental marijuana smoke in the past 6 months. When compared with the youths who were true users, the likelihood of youths underreporting in both the 7-day and 30-day periods *decreased* as claims of passive exposure increased. In contrast, the models comparing underreporters with the true nonusers showed a positive relationship between underreporting and claims of passive exposure to marijuana smoke. In the 30-day model, race/ethnicity and difficulty in answering the drug-related questions also were significant predictors of underreporting. Persons in the "other" race/ethnicity category were less likely than whites to underreport. The likelihood of underreporting also increased as youths reported more difficulties in answering the drug-related questions.

5.8.2 18 to 25 Year Olds

For young adults aged 18 to 25, a total of 114 were 7-day underreporters, and 90 were 30-day underreporters. As with the youths, the number of underreporters relative to true users was greater for the 7-day reference period (114 underreporters vs. 171 true users) than for the 30-day period (90 vs. 196, respectively).

The logistic regression results for the 18- to 25-year-old underreporters are presented in Table 5.14. In the 7-day and 30-day models for underreporters and true nonusers, five variables were significant predictors of underreporting: race, receipt of the appeal, truthfulness, reported

passive exposure to marijuana smoke in the past 6 months, and region. Specifically, blacks were 2.5 to 2.8 times more likely than whites to underreport. Young adults who did not receive the appeal were up to 2.2 times more likely to underreport compared with young adults who received the appeal. Those who reported being less than completely truthful when answering the drug-related questions were up to 2.4 times more likely to underreport than those who reported being completely truthful. Compared with true nonusers, the likelihood of underreporting increased as claims of passive exposure to environmental marijuana smoke increased. In both models, young adults in the Midwest were less likely to underreport compared with those in the West. Also, in the 7-day model, young adults in the South were less likely to underreport compared with those in the West. Unlike the models of underreporting compared with true nonusers among youths, religiosity was not a significant predictor of underreporting in the corresponding models for young adults.

When the underreporters were compared with the true users (columns 3 and 4), truthfulness in answering the drug-related questions was no longer a significant predictor of underreporting. Race and receipt of the appeal were significant predictors of underreporting in one of the models but not both. Blacks were more likely than whites to underreport use in the past 7 days, but not in the model for use in the past 30 days. Respondents who did not receive the appeal were more likely than those who received the appeal to underreport use in the past 30 days but not in the model for use in the past 7 days.

Although reported passive exposure to environmental marijuana smoke in the past 6 months was a significant predictor of underreporting in the models comparing underreporters with true users, the direction of the relationship was reversed compared with the models for underreporters and true nonusers. Specifically, the likelihood of underreporting decreased in the models comparing underreporters with true users as claims of passive exposure increased. As noted above, these findings also were observed in the models for youths.

In addition, region and friends' use of marijuana were significant predictors of underreporting in these models comparing young adults who were underreporters with those who were true marijuana users. Specifically, young adults in the Midwest were less likely than those in the West to underreport. Relative to true users, the likelihood of underreporting among young adults decreased as the reported number of friends who used marijuana increased.

5.8.3 Summary of Marijuana Regression Model Results for Underreporters

Some general similarities and effects are to be noted.

- Findings in the models were consistent for the passive exposure variable for both youths aged 12 to 17 and young adults aged 18 to 25, with a positive relationship being observed between underreporting and reported passive exposure in the models comparing underreporters with true nonusers, but a negative relationship in models comparing underreporters with true users.

Table 5.14 Marijuana Use: Logistic Regression Models Predicting 7-Day and 30-Day Underreporting among Young Adults Aged 18 to 25

Model Covariate	7-Day UR vs. TNU		30-Day UR vs. TNU		7-Day UR vs. TU		30-Day UR vs. TU	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Race								
White	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Black	2.76[b]	(1.66-4.59)	2.51[b]	(1.37-4.60)	2.20[a]	(1.06-4.59)	1.29	(0.49-3.44)
Other	0.30	(0.06-1.45)	0.37	(0.08-1.70)	0.66	(0.17-2.51)	0.98	(0.24-4.06)
Gender								
Male	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Female	0.68	(0.42-1.11)	0.83	(0.48-1.43)	1.30	(0.66-2.56)	1.91	(0.85-4.29)
Region								
West	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Midwest	0.29[b]	(0.12-0.72)	0.34[a]	(0.12-1.00)	0.34[a]	(0.14-0.81)	0.29[a]	(0.10-0.87)
South	0.38[a]	(0.18-0.79)	0.44	(0.19-1.03)	0.58	(0.26-1.28)	0.75	(0.30-1.90)
Northeast	0.64	(0.32-1.26)	0.65	(0.29-1.46)	1.30	(0.49-3.44)	1.02	(0.30-3.42)
Receipt of Appeal								
Received Appeal	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Did Not Receive Appeal	1.89[a]	(1.17-3.07)	2.20[b]	(1.35-3.60)	1.76	(0.82-3.77)	2.41[a]	(1.01-5.77)
Truthfulness								
Completely Truthful	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)	1.00	(1.00-1.00)
Not at All/Somewhat/ Mostly Truthful	2.15[a]	(1.01-4.57)	2.37[a]	(1.04-5.39)	1.47	(0.49-4.37)	1.46	(0.38-5.52)
Privacy of Interview[1]	0.96	(0.83-1.10)	0.92	(0.79-1.08)	1.10	(0.81-1.49)	0.96	(0.71-1.30)
Difficulty in Answering Drug-Related Questions[2]	1.05	(0.96-1.14)	1.04	(0.95-1.15)	1.02	(0.86-1.20)	1.01	(0.85-1.19)
Passive Marijuana Smoke Exposure in Past 6 Months[3]	1.88[b]	(1.47-2.41)	1.45[a]	(1.09-1.93)	0.43[b]	(0.29-0.63)	0.24[b]	(0.14-0.41)
Friends Using Marijuana[4]	1.32	(0.93-1.87)	1.30	(0.91-1.85)	0.30[b]	(0.18-0.49)	0.32[b]	(0.18-0.55)
Religiosity[5]	0.94	(0.88-1.01)	0.94	(0.87-1.00)	1.02	(0.90-1.16)	1.01	(0.88-1.15)
	n = 114/1,417 χ^2 = 116.24 df = 13		n = 90/1,354 χ^2 = 66.65 df = 13		n = 114/171 χ^2 = 117.56 df = 13		n = 90/196 χ^2 = 145.96 df = 13	

CI = confidence interval.

Note: Underreporters (UR) self-report negative and have a positive urine test, true nonusers (TNU) self-report negative and have a negative urine test, and true users (TU) self-report positive and have a positive urine test. The 7-day self-reporting is based on responses to repeat questions. The 30-day self-reporting is based on responses to core or repeat questions. Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

[a] Statistically significant at the 0.05 level.

[b] Statistically significant at the 0.01 level.

[1] Higher score indicates less privacy during interview.

[2] Respondents were asked four questions on their difficulties in (1) understanding the drug-related questions and (2) remembering the drug-related information, as well their (3) clarity of memories regarding drug-related information and (4) how often they made a best guess when answering the drug-related questions. Higher scores indicate more difficulty.

[3] Higher score indicates more frequent exposure to persons smoking marijuana in the past 6 months.

[4] Higher score indicates more friends who use marijuana.

[5] Respondents were asked four questions regarding their religious beliefs and practices: (1) number of times attended religious services in the past year, (2) importance of religious beliefs, (3) influence of religious beliefs on decisionmaking, and (4) importance of friends sharing religious beliefs. Higher scores indicate more religiosity.

- Truthfulness in answering the drug-related questions differentiated underreporters from true nonusers for young adults in both the 7-day and 30-day models and for youths with regard to underreporting of use in the past 7 days.

- The appeal did not significantly affect underreporting for youths, but in most models, it reduced underreporting among young adults. The exception was the 7-day model for young adults comparing overreporters with true users, where this variable was not significant.

6. Cocaine

This chapter examines the validity of self-reported cocaine use. Both cocaine hydrochloride and crack are considered because crack is a derivative of powdered cocaine. The urine drug test cannot distinguish between these two forms of cocaine. The target analyte for the screening tests was benzoylecgonine (BZE), the primary urinary metabolite of cocaine. The screening cutoff used was 50 nanograms per milliliter (ng/mL). The gas chromatography/mass spectrometry (GC/MS) confirmation test analyzed for BZE using a cutoff of 5 ng/mL to confirm the screening test results. BZE has an excretion half-life of approximately 7.5 hours (Ambre, 1985; Cone, Tsadik, Oyler, & Darwin, 1998). In occasional users, the metabolite concentrations drop below 150 ng/mL in 2 to 4 days after cessation of use.

As with other drugs, there is no exact window of detection. BZE and other cocaine metabolites may be excreted in the urine for up to several weeks. It has been reported that chronic users of cocaine may produce cocaine-positive urine specimens at a 300 ng/mL threshold for several weeks after cessation of use and that their urine concentrations may move back and forth across the 300 ng/mL cutoff threshold (Burke, Ravi, Dhopesh, Vandegrift, & Maany, 1990; Cone & Weddington, 1989; Jufer, Walsh, & Cone, 1998; Jufer, Wstadik, Walsh, Levine, & Cone, 2000; Weiss & Gawin, 1988). Thus, it is possible for someone who has not used cocaine in the past 3 days to be defined as a recent user of cocaine when he or she may have ceased use well before the conventionally accepted "72 hour window of detectability" (Mieczkowski & Newel, 1997).

The stigma associated with some drugs, such as heroin and cocaine (specifically, crack) may affect the validity of self-reported use. Several studies have found that cocaine is less likely to be correctly self-reported than marijuana (Fendrich & Xu, 1994; Harrison, 1995; Mieczkowski, Barzelay, Gropper, & Wish, 1991a). Magura, Goldsmith, Casriel, Goldstein, and Lipton (1987) found that self-reporting by methadone clients was least valid for opiates, while benzodiazepine and cocaine reporting were moderately valid and highly valid, respectively.

6.1. Self-Reported Cocaine Use: Comparison of Responses to Core Questions and Urinalysis

Table 6.1 presents a comparison of the results of self-reported cocaine use in the past 30 days on the core questions with the results of the urine test. The two-by-two table includes those with a valid urine specimen. Table 6.1 shows that 97.9 percent (cell A) of the respondents said they had not used cocaine in the past 30 days and tested negative. This is such a high percentage that it is obvious that none of the other cells in the table can reach much more than 1 percent. However, the next largest cell was underreporters (cell B): 1.1 percent tested positive and did not report use. A scant 0.3 percent (cell D) reported cocaine use and tested positive. The sensitivity level of 0.21 was very poor, as would be expected looking at the table. The specificity was much higher at 0.99. It might be anticipated that this number would not be as high because there were nearly equivalent proportions of under- and overreporters, but this number is influenced by the large proportion of respondents who reported no cocaine use in the past 30 days.

Table 6.1 Cocaine Use: Comparison of Responses to 30-Day Self-Report *Core* Questions and Urinalysis

(*n* = 3,761)		Urinalysis (-)	Urinalysis (+)	
	A			**B**
30-Day Self-Report (Core Questions)	(-)	97.9%	1.1%	(3,718)
	(+)	0.6%	0.3%	(43)
	C	(3,709)	(52)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
12.11	1	0.0007	0.206	0.994	0.243

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 30 day use of cocaine on the *core* questions.

Note 3: A screening cutoff concentration of 50 nanograms per milliliter (ng/mL) for cocaine and a confirmatory cutoff concentration of 5 ng/mL for benzoylecgonine were used to identify a specimen as positive for cocaine.

This table suggests that there has been some underreporting of cocaine use in the National Household Survey on Drug Abuse (NHSDA) and other general population surveys of youths and young adults aged 12 to 25. However, 0.9 percent reported use in the past 30 days on the core, and only 1.4 percent tested positive. Overreporters were about as large a proportion as underreporters, but overreporters possibly used the drug outside the window of detection. Combining self-reported use and positive urinalysis responses resulted in an overall prevalence rate of 2.0 percent for past 30 day cocaine use, which although small, was about twice as large as the past 30 day prevalence rate based on self-reports.

6.2. Self-Reported Cocaine Use: Comparison of Responses to Core and Repeat Questions

Table 6.2 presents the results of self-reported cocaine use in the past 30 days on both the core and the repeat questions. Only respondents who reported lifetime use were prompted to respond to questions about use in the past 30 days. The two-by-two table shows that the largest proportion reported "no" cocaine use on both questions at 98.9 percent (cell A). A small percentage, 0.7 percent, said "yes" in both questions (cell D). Equivalently small proportions, 0.2 percent, changed their answers on the core and repeat questions (cell B and cell C), which is noteworthy, when only 0.3 percent had congruent positive urine results and core self-reports

(Table 6.1). This includes about 20 respondents who changed their answers between the repeat and core questions. The pattern of changing answers was found with marijuana and tobacco as well. If all those who reported use on either the core or repeat questions were combined, the overall self-reported prevalence rate would be 1.1 percent.

Table 6.2 Cocaine Use: Comparison of Responses to 30-Day Self-Report *Core* and *Repeat* Questions

(*n* = 4,444)		30-Day Self-Report (Core Questions) (-)	30-Day Self-Report (Core Questions) (+)	
	A			B
30-Day Self-Report (Repeat Questions)	(-)	98.9%	0.2%	(4,398)
	(+)	0.2%	0.7%	(46)
	C	(4,397)	(47)	D

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
23.16	1	<0.0001	0.770	0.998	0.758

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those reporting past 30 day use of cocaine on the *core* and/or *repeat* questions.

6.3. Self-Reported Cocaine Use: Comparison of Responses to Repeat Questions and Urinalysis

Table 6.3 presents the results of self-reported cocaine use in the past 30 days on the repeat questions with the results of the urine test. It shows that 97.9 percent (cell A) of respondents said they had not used cocaine in the past 30 days and tested negative—the same as in the core questions. There were slightly more congruent positive urine tests and self-reports among the repeat than in the core questions. There were similar proportions of under- and overreporters. Between the core and repeat questions, the sensitivity level remained poor at 0.22, and neither it nor any of the "goodness-of-fit" measures showed any improvement over the core questions.

6.4. Self-Reported Cocaine Use: Comparison of Responses to 3-Day Self-Report and Urinalysis

Most research suggests that cocaine's detection is much shorter than 30 days (more like 2 to 4 days). All chronic cocaine users should have reported use in the past 7 days, so Table 6.4

compares self-reported cocaine use in the past 7 days with the urinalysis results. The results were not much changed over the 30-day comparisons because the majority of respondents neither self-reported nor tested positive. However, the proportion of overreporters was lower (cell C).

Table 6.3 Cocaine Use: Comparison of Responses to 30-Day Self-Report *Repeat* Questions and Urinalysis

		Urinalysis		
(*n* = 3,753)	**A**	(-)	(+)	**B**
30-Day Self-Report (Repeat Questions)	(-)	97.9%	1.1%	(3,710)
	(+)	0.7%	0.3%	(43)
	C	(3,701)	(52)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
12.03	1	0.0008	0.218	0.993	0.252

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 30 day use of cocaine on the *repeat* questions.

Note 3: A screening cutoff concentration of 50 nanograms per milliliter (ng/mL) for cocaine and a confirmatory cutoff concentration of 5 ng/mL for benzoylecgonine were used to identify a specimen as positive for cocaine.

For the 3-day self-reports, the follow-up and repeat questions were combined, with any discrepancy recoded to use. Table 6.5 shows that the proportion of overreporters was lower in the past 7 days than in the 30-day core and repeat and was even lower when the time period was reduced to the past 3 days. However, overreporters were still present in both the 3- and 7-day timeframes. The proportion of underreporters (cell B) was unchanged in these timeframes, although in these cases, the urinalysis results produced larger prevalence rates than did the self-reports. Of those who tested positive, 81.5 percent in the 7-day timeframe and 83.7 percent in the 3-day timeframe denied use of this highly stigmatized drug.

Table 6.4 Cocaine Use: Comparison of Responses to 7-Day Self-Report and Urinalysis

(*n* = 3,753)		Urinalysis (-)	Urinalysis (+)	
		A		**B**
7-Day Self-Report (Repeat Questions)	No	98.2%	1.2%	(3,727)
	Yes	0.3%	0.3%	(26)
		C (3,701)	(52)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
9.75	1	0.0023	0.185	0.997	0.257

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 7 day use of cocaine on the *repeat* questions.

Note 3: A screening cutoff concentration of 50 nanograms per milliliter (ng/mL) for cocaine and a confirmatory cutoff concentration of 5 ng/mL for benzoylecgonine were used to identify a specimen as positive for cocaine.

Table 6.5 Cocaine Use: Comparison of Responses to 3-Day Self-Report and Urinalysis

(*n* = 3,753)		Urinalysis (-)	Urinalysis (+)	
		A		**B**
3-Day Self-Report (Follow-Up or Repeat Questions)	No	98.4%	1.2%	(3,735)
	Yes	0.1%	0.2%	(18)
		C (3,701)	(52)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
7.71	1	0.0066	0.163	0.999	0.255

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 3 day use of cocaine on the *follow-up* or *repeat* questions.

Note 3: A screening cutoff concentration of 50 nanograms per milliliter (ng/mL) for cocaine and a confirmatory cutoff concentration of 5 ng/mL for benzoylecgonine were used to identify a specimen as positive for cocaine.

7. Opiates

Opiates are primarily central nervous system depressants and narcotic analgesics. Opium is the milky latex fluid contained in the unripened seedpod of the opium poppy. Opium contains a number of different alkaloid families, but only one of them can be converted to narcotic substances. The natural opiates in this family are morphine, codeine, and thebaine. Heroin, hydromorphone, hydrocodone, and oxycodone (OxyContin®) are semisynthetic opiates derived from morphine, codeine, or thebaine. There is also a group of fully synthetic narcotics, including meperidine (Demerol®), methadone (Dolophine®), propoxyphene (Darvon®), and many other drugs. They are designed and produced mostly as prescription pain relievers (PPRs), but can be used illegally. Heroin may be the most popular street drug among the opiates. It typically is found in white to brown powdered form and can be injected, sniffed, or smoked. In recent years, street supplies have become much purer. The availability of higher purity heroin allows for more users to sniff or smoke the drug, instead of injecting it, and still achieve the desired effect.

Morphine and codeine are the two most common opiates tested in the urine. Another opiate analyte encountered in urine drug testing is 6-monoacetylmorphine (6-MAM), a metabolite found in the urine only after heroin use. Because it has a very short half-life and can be detected for 8 to 10 hours and as much as 34 hours (Reiter et al., 2001), its presence in urine generally indicates very recent use of heroin (Jenny, 1989; Makkai, 2000; White & Irvine, 1999). Additionally, heroin metabolizes to morphine; however, the presence of morphine alone may suggest either the use of morphine or heroin. Morphine takes longer to be excreted than 6-MAM and may be detected in urine for 2 to 4 days after the last use. Codeine, present in some prescription cough suppressants and pain relievers, also metabolizes to morphine. Codeine stays in the urine for 1 to 2 days. A codeine concentration higher than that of morphine suggests that codeine has been ingested (Makkai, 2000; Reiter et al., 2001).

The Validity Study used a screening cutoff of 50 nanograms per milliliter (ng/mL). The screening tests were specific to morphine and codeine, but also cross-reacted with semisynthetic opiates. Specimens screened positive (i.e., at or above the screening cutoff concentration) were tested only for codeine and morphine by gas chromatography/mass spectrometry (GC/MS) using a cutoff concentration of 5 ng/mL.

Survey questions related to opiate use asked about the use of heroin and the use of PPRs. Because the GC/MS test only analyzed for morphine and codeine, this precluded any analysis of the data for other opiate use. In addition, although the core questions asked about heroin and specific PPRs (i.e., by drug or by trade name), the follow-up and repeat sections of the interview only asked about heroin. There were no follow-up questions for PPRs. Therefore, the core and repeat questions could be used to compare self-reports with the urine drug test results for heroin and PPRs.

However, there are also problems with using only the core questions. All of the core questions (i.e., the separate questions about specific drugs and the general question referring to PPRs listed on an interview card) included drugs that do not produce the analytes used to detect opiate use (i.e., codeine and morphine). Respondents using these drugs would not be expected to test positive. Therefore, the number of overreporters may be inflated. The information from the

core questions does not provide an accurate picture of recent use because the drug-specific questions asked only about the lifetime use of the drug and not about its use in the past month. Questions about past month use asked about the nonmedical use of any PPR. Respondents who reported nonmedical use of multiple PPRs were not able to report the recency of use of a specific drug. The number of overreporters also may be inflated because the 30-day timeframe is much longer than the detection time window for codeine and morphine in urine.

The repeat questions asked about legitimate use of any PPRs (i.e., the use of prescribed PPRs). If the repeat questions had asked about specific drugs, this information could have been used to exclude those reporting legitimate use from the sample. Because the questions were not drug-specific and the PPRs in the core questions included drugs that would not test positive, no respondents were excluded based on these questions.

Table 7.1 compares data on self-reported opiate use, including the use of PPRs, in the past 30 days on the core questions with the urinalysis results. Fifty tested positive, and 32 reported use in the past 30 days. Table 7.2 compares the data on self-reported opiate use in the past 30 days on the core and repeat questions. One percent reported "yes" on the core, but "no" on the repeat. More than 2 times as many, however, reported use on the repeat, but not on the core. It would appear that the appeal encouraged more reporting of opiates. The higher

Table 7.1 Opiate and Prescription Pain Reliever (PPR) Use: Comparison of Responses to 30-Day Self-Report *Core* Questions and Urinalysis

(*n* = 3,771)		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Core Questions)	(-)	97.7	0.8	(3,739)
	(+)	1.5	0.1	(32)
	C	(3,721)	(50)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
1.33	1	0.2510	0.119	0.985	0.073

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: *Core* questions ask about drug use and replicate the NHSDA format.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 30 day use of any opiate product on the *core* questions.

Note 3: Percentages include those reporting past 7 day use of prescribed prescription pain relievers.

Note 4: A screening cutoff concentration of 50 nanograms per milliliter (ng/mL) for opiates and, for codeine and morphine confirmatory testing, a cutoff concentration of 5 ng/mL were used to identify a specimen as positive for opiates.

Table 7.2 Opiate Use: Comparison of Responses to 30-Day Self-Report *Core* and *Repeat* Questions

(*n* = 4,434)	A	30-Day Self-Report (Core Questions) (-)	30-Day Self-Report (Core Questions) (+)	B
30-Day Self-Report (Repeat Questions)	(-)	97.4	0.4	(4,349)
	(+)	1.0	1.1	(85)
	C	(4,379)	(55)	D

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
22.49	1	<.0001	0.524	0.996	0.602

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those reporting past 30 day use of opiates on the *core* and/or *repeat* questions.

prevalence rate on the repeat was somewhat unexpected because the core asked individual questions about a variety of opiate drugs, while the repeat question was restricted to the overall class of drugs without questions about specific opiate drugs.

Tables 7.3 and 7.4 present findings for comparisons of urinalysis results with 1-day and 3-day self-reports based on repeat questions. A total of 32 tested positive, but 40 reported use based on the 7-day questions and 26 reported use based on the 3-day questions.

Some poppy seeds and poppy seed pastes used in the preparation of foodstuffs (e.g., bagels, muffins, poppy seed cakes) contain codeine and morphine. Individuals who have eaten such products may test positive for codeine and/or morphine using the low cutoffs employed in this study. Survey questions did not address this issue. Therefore, the number of underreporters may be inflated.

Due to the small sample sizes for opiates (in terms of both self-reported use and positive urine drug test results), the issues discussed above concerning the survey, and the lack of urine test data for semisynthetic opiates, no meaningful conclusions about the validity of self-reported opiate use can be drawn.

Table 7.3 Opiate and Prescription Pain Reliever (PPR) Use: Comparison of Responses to 7-Day Self-Report *Repeat* Questions and Urinalysis

(*n* = 3,754)

		Urinalysis		
	A	**(-)**	**(+)**	**B**
7-Day Self-Report (Repeat Questions)	(-)	97.9	0.8	(3,714)
	(+)	1.2	0.0	(40)
	C	(3,722)	(32)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
0.42	1	0.5207	0.034	0.988	0.018

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 7 day use of any opiate product on the *repeat* questions..

Note 3: Percentages include those reporting past 7 day use of prescribed prescription pain relievers.

Note 4: A screening cutoff concentration of 50 nanograms per milliliter (ng/mL) for opiates and, for codeine and morphine confirmatory testing, a cutoff concentration of 5 ng/mL were used to identify a specimen as positive for opiates.

Table 7.4 Opiate and Prescription Pain Reliever (PPR) Use: Comparison of Responses to 3-Day Self-Report *Repeat* Questions and Urinalysis

(*n* = 3,754)

		Urinalysis		
	A	(-)	(+)	**B**
3-Day Self-Report (Repeat Questions)	(-)	98.4	0.8	(3,728)
	(+)	0.8	0.0	(26)
	C	(3,722)	(32)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
0.60	1	0.4417	0.0345	0.992	0.027

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 3 day use of any opiate product on the *repeat* questions.

Note 3: Percentages include those reporting past 7 day use of prescribed prescription pain relievers.

Note 4: A screening cutoff concentration of 50 nanograms per milliliter (ng/mL) for opiates and, for codeine and morphine confirmatory testing, a cutoff concentration of 5 ng/mL were used to identify a specimen as positive for opiates.

8. Amphetamines

Amphetamines (i.e., amphetamine and methamphetamine) are classified as "stimulants," a name given to several groups of drugs that tend to increase alertness and physical activity. Stimulants are abused to produce a sense of exhilaration, enhance self-esteem, improve mental and physical performance, increase activity, reduce appetite, produce prolonged wakefulness, and enable users to "get high" (U.S. Department of Justice, Drug Enforcement Administration, 2005).

Amphetamine and methamphetamine have legitimate medical uses, including the treatment of obesity, narcolepsy, and attention deficit/hyperactivity disorder (ADHD), but these uses have been limited in recent years. Illicit amphetamines can be pharmaceuticals diverted from legitimate sources for illicit use or drugs manufactured in clandestine laboratories. Amphetamines can be ingested in a number of ways, including sniffing, oral ingestion, smoking, and injection. Stimulants including amphetamine and methamphetamine are commonly referred to as "uppers" or "speed." Street terms for amphetamine include "black beauties," "white cross," and "dexies" (Office of National Drug Control Policy [ONDCP], 2006b). Methamphetamine comes in a powder form that resembles granulated crystals, commonly called "crystal meth," and in a rock form known as "ice," which is the smokeable version (ONDCP, 2006a).

Amphetamines are generally detectable in the urine 24 to 48 hours after use (Cone, 1997; Hawks & Chiang, 1986). Methamphetamine metabolizes to amphetamine, so both drugs may be present in the urine after methamphetamine use.

The design of the Validity Study had several limitations that should be considered when interpreting the urine drug test results. The survey questions (i.e., core, follow-up, repeat) asked about the use of stimulants, only some of which contain or metabolize to amphetamine or methamphetamine. Because only amphetamine and methamphetamine were tested by the laboratory in the Validity Study, the use of other stimulants would not be identified by urinalysis. At least one prescription drug containing amphetamine commonly used by youths and young adults aged 12 to 25 in the general population (i.e., Adderall®) was not addressed in the survey questions. In addition, respondents were asked about the use of designer drugs, such as methylene dioxymethamphetamine (MDMA) or methylene dioxyamphetamine (MDA), commonly used by the young adult population at the time of the study, in the core but not the follow-up and repeat questions. Although these drugs would not by themselves cause a positive amphetamine test, it is known that illicit formulations of these drugs can contain amphetamine or methamphetamine and that sometimes amphetamine or methamphetamine pills are sold as these drugs. Finally, legitimate use was not differentiated from illicit use in this study for these drugs.

Table 8.1 compares self-reported stimulant use on the core questions in the past 30 days with the urinalysis results for amphetamine. Congruency between self-reported stimulant use and urine tests for amphetamines was relatively high, 98.2 percent, primarily because 98.1 percent of all respondents reported no use of stimulants in the past 30 days and tested negative. The number of cases in the two incongruent categories—under- and overreporters—was much smaller. If only self-report data were available, the 30-day prevalence rate would be 1.0 percent. If only

urinalysis data were available, the prevalence rate would be the same (1.0 percent). Combining any positive self-report and urinalysis data would result in a prevalence rate of 1.9 percent.

Of the 42 people who tested positive for amphetamines in their urine, only 4 reported stimulant use in the past 30 days. This created a high percentage of underreporters among the positive participants. As noted above, underreporting could be explained by the fact that respondents were not asked about all drugs that could cause a positive amphetamine test, or they were not asked about a particular drug and did not recognize that the drug would fall in the stimulants category (e.g., Adderall®, the amphetamine used in the treatment of ADHD and narcolepsy). Also, for most stimulants, the survey only asked about nonmedical use. Respondents using valid prescription drugs that could cause a positive amphetamine test result were unable to report that use.

There was also a high percentage of overreporters; in fact, the numbers of overreporters and underreporters were the same. Of the respondents who reported use of any stimulants in the past 30 days (n = 42), 90 percent (n = 38) did not test positive. There are two plausible explanations for the high percentage of overreporters. First, the timeframe in the core questions was 30 days, which was much longer than the timeframe for amphetamine detection (up to 3 days). The 3-day timeframe for the self-report of stimulant use is more consistent with the detection time for these drugs. Second, the survey questions about stimulants included drugs that would not cause a positive amphetamine test result. Respondents may have reported honestly about their use of a stimulant that did not contain or metabolize to the drug test analytes (i.e., amphetamine and methamphetamine).

Table 8.1 Stimulant Use: Comparison of Responses to 30-Day Self-Report *Core* Questions and Urinalysis for Amphetamines

(n = 3,762)

		Urinalysis		
	A	(-)	(+)	**B**
30-Day Self-Report (Core Questions)	(-)	98.1%	0.9%	(3,720)
	(+)	0.9%	0.1%	(42)
	C	(3,720)	(42)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
2.74	1	0.1013	0.075	0.991	0.065

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: *Core* questions ask about drug use and replicate the NHSDA format.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 30 day use of any stimulant product on the *core* questions.

Note 3: A screening cutoff concentration of 500 nanograms per milliliter (ng/mL) for amphetamines and, for amphetamine and methamphetamine confirmatory testing, a cutoff concentration of 25 ng/mL were used to identify a specimen as positive for amphetamines.

Compared with the 30-day self-report from the core questions, Table 8.2 shows that the 30-day repeat questions produced significantly fewer overreporters (n = 19 vs. n = 38), while preserving the number of underreporters. The percentage of congruent cases reached 98.7 percent, a 0.5 percent increase compared with the core questions. It appears that the appeal, while having no effect on the underreporting, reduced overreporting (there were only 4 overreporters among those who received the appeal compared with 15 who did not receive the appeal). However, the 30-day prevalence rate based on the repeat questions was 0.6 percent compared with 1.0 percent if only urinalysis data were available.

Table 8.2 Stimulant Use: Comparison of Responses to 30-Day Self-Report *Repeat* Questions and Urinalysis for Amphetamines

(*n* = 3,748)	**A**	**Urinalysis (-)**	**Urinalysis (+)**	**B**
30-Day Self-Report (Repeat Questions)	(-)	98.6%	0.9%	(3,726)
	(+)	0.5%	0.1%	(22)
	C	(3,707)	(41)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
2.20	1	0.1407	0.061	0.995	0.071

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 30 day use of stimulants on the *repeat* questions.

Note 3: A screening cutoff concentration of 500 nanograms per milliliter (ng/mL) for amphetamines and, for amphetamine and methamphetamine confirmatory testing, a cutoff concentration of 25 ng/mL were used to identify a specimen as positive for amphetamines.

Tables 8.3 and 8.4 show the congruence between self-reported stimulant use in the repeat questions and the amphetamine urine test results for the past 7 day and 3 day timeframes. The dynamic of the two congruent groups, congruent nonusers and congruent users, was rather stable. More interesting, however, was the dynamic of the two incongruent groups—the under- and overreporters (cells B and C, respectively). The group of underreporters was consistent across the different time periods mainly because the number of positive cases did not change over time and also because those who did not report any use in a more distant time were less likely to report more recent use. The group of overreporters, however, varied with time, with 19 respondents overreporting use in the past 30 days in the repeat questions, 9 in the past 7 days, and 5 in the past 3 days.

Table 8.3 Stimulant Use: Comparison of Responses to 7-Day Self-Report *Repeat* Questions and Urinalysis for Amphetamines

(*n* = 3,748)		Urinalysis (-)	Urinalysis (+)	
	(-)	98.8%	0.9%	(3,737)
7-Day Self-Report (Repeat Questions)	(+)	0.2%	0.0%	(11)
		C (3,707)	(41)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
1.81	1	0.1810	0.051	0.998	0.075

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 7 day use of stimulants on the *repeat* questions.

Note 3: A screening cutoff concentration of 500 nanograms per milliliter (ng/mL) for amphetamines and, for amphetamine and methamphetamine confirmatory testing, a cutoff concentration of 25 ng/mL were used to identify a specimen as positive for amphetamines.

Table 8.4 Stimulant Use: Comparison of Responses to 3-Day Self-Report *Repeat* Questions and Urinalysis for Amphetamines

(*n* = 3,748)		Urinalysis (-)	Urinalysis (+)	
	(-)	98.9%	0.9%	(3,741)
3-Day Self-Report (Repeat Questions)	(+)	0.1%	0.0%	(7)
		C (3,707)	(41)	**D**

CMH	*df*	*p*	Sensitivity	Specificity	Kappa
1.89	1	0.1726	0.051	0.999	0.083

CMH = Cochran-Mantel-Haenszel chi-square test.

df = degrees of freedom.

Note 1: The *core* and *follow-up* portions of the study interview ask respondents to report their recency of use. Respondents then listen to one of two possible introductions to the next series of questions in which respondents again report their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on those with a valid urine specimen and reporting past 3 day use of stimulants on the *repeat* questions.

Note 3: A screening cutoff concentration of 500 nanograms per milliliter (ng/mL) for amphetamines and, for amphetamine and methamphetamine confirmatory testing, a cutoff concentration of 25 ng/mL were used to identify a specimen as positive for amphetamines.

Due to the small sample sizes for stimulants (in terms of both self-reported use and positive amphetamine urine test results), the Validity Study question format, and the limited stimulant test panel (i.e., only amphetamine and methamphetamine), it is difficult to draw meaningful conclusions about the validity of self-reported stimulant use.

9. Summary and Implications

9.1. Fieldwork

Because drug use is a relatively rare behavior in the general population, it is necessary to have large sample sizes to examine the validity of self-report. The Validity Study focused on youths and young adults aged 12 to 25 in order to maximize the efficiency of the sample by including those in the general population most likely to be currently using drugs. Actually, the decision to include youths between the ages of 12 and 17 was not based on higher use patterns because this age group is at the point in their developmental process when experimentation with drugs begins. The decision to include youths was based on a body of research that suggests that reporting patterns may differ among youths compared with adults, and during a time when drug use had been increasing slightly among youths for several years. There were no comparable increases among adults. However, the young adult population, which the National Household Survey on Drug Abuse (NHSDA)[6] defines as those between the ages of 18 and 25, has always shown the highest rates of active drug use (past year and past month).

9.2. Bivariate Results

Debriefing questions were unique to the Validity Study and were not included in the 2000 and 2001 NHSDAs. In response to a question about their truthfulness in responding to the drug-related questions, a robust 90 percent of respondents said they were completely truthful. Another 7.9 percent said they were mostly truthful. However, respondents thought that most people would not be as truthful as they were. Respondents reported little difficulty in understanding the drug-related questions or in remembering the information requested. The overall pattern in the debriefing questions was for respondents to rate themselves as having less difficulty than most people and reporting more accurately and honestly than most people. Three quarters of respondents said they were "not at all" embarrassed by answering the drug-related questions. Despite all the attention in the Validity Study to assuring respondents about the confidentiality of the answers they provided, over a quarter (27.6 percent) suggested they were somewhat to very concerned about others having access to their answers. Two thirds reported that they thought most people would be somewhat to very concerned that others would have access to their answers.

The debriefing questions were followed by a "repeat" of questions about the use of the five drugs that were tested for in urine. Comparing the answers on the repeat questions with the answers on the core questions (the first time they were asked about drug use) revealed no significant differences in prevalence rates, although a surprising number of respondents had inconsistent answers on the core and repeat questions. The major lesson from comparing the core and repeat questions on the two-by-two tables comes from an examination of the tobacco questions. Greater sensitivity was achieved in the core questions that asked about each type of tobacco product individually. Most tobacco users smoked cigarettes, and, even among those who used more than one tobacco product, they were most likely to smoke cigarettes and cigars. Still, excluding smokeless tobacco use and pipe smoking from the repeat questions resulted in less

[6] Since 2002, the survey has been called the National Survey on Drug Use and Health (NSDUH).

congruence between self-report and urinalysis among those who tested positive. This suggests the need for surveys to ask more detailed questions about drug types to generate more accurate prevalence rates. For tobacco specifically, it also indicates the need to include questions about kreteks (clove cigarettes) and bidis (which are used by a small but noticeable proportion of youths) because the use of these tobacco products leads to a positive urine test for cotinine. Questions were included about the use of bidis and kreteks in the second year of the Validity Study (as well as in the 2001 NHSDA).

The repeat questions were delivered after the persuasion experiment or appeal, and it was hypothesized that the appeal would increase self-reporting rates. It appears that the appeal did in fact decrease underreporting, and this was especially robust in the two-by-two tables for marijuana. In the multivariate models, the appeal was only significant for marijuana and not tobacco and among young adults and not youths.

It is important to point out that higher prevalence rates were self-reported for past 30 day use than would have been generated from the urine test results alone. More respondents self-reported use than tested positive. Some researchers have suggested that combining positive self-reports and positive urinalysis results would create the best prevalence estimates of drug use. This strategy would increase the prevalence rates for drug use generated by the Validity Study.

9.3. Multivariate Results

There were sufficient cases to examine the predictors of over- and underreporters of tobacco and marijuana use by age group using logistic regression. Separate models were run for youths aged 12 to 17 and young adults aged 18 to 25 because tabular analyses suggested some differences between these age groups.

The results from the tobacco and marijuana sections of the Validity Study underscore that the underreporting of drug use, even under conditions of research confidentiality, continues to be an issue for research studies. There were clearly small proportions who had used these drugs recently, but did not report use. Some of the tobacco underreporters may have tested positive due to passive exposure to environmental tobacco smoke. This is suggested by the results of the logistic regression models in which passive exposure was significantly related to underreporting of tobacco use. However, as shown in Table F.12 in Appendix F, the cotinine concentrations of respondents who denied use and tested positive were not consistent with their reported frequency of passive exposure (i.e., the lowest average concentration was for those respondents reporting daily passive exposure). This was also true for the marijuana underreporters (see Table F.22 in Appendix F). Another possible explanation for some of the tobacco underreporting may be that the Validity Study only employed a screening immunoassay test for cotinine without a confirmatory test.

9.4. Implications and Lessons Learned

The Validity Study has provided valuable information concerning the validity of self-reported drug use in general population surveys and the use of biological drug test results to verify self-reports. Although many of the study objectives were accomplished, problems

encountered during the study and issues raised during evaluation of the results have provided information that will be extremely useful in designing future studies of this type.

Most prior research has been confined to criminal justice or treatment populations and, therefore, cannot be generalized to the large population not involved with the criminal justice system or in treatment. This is the first large-scale study that compares self-reported drug use with the results of urine and hair drug tests. It utilized the state-of-the-art methodology employed in the NHSDA, which is the only nationally representative survey of drug use in the general population in the United States and has been ongoing since 1971. The NHSDA monitors patterns and changes in drug use for the Nation as a whole and among population subgroups, such as youths and minorities. The NHSDA methodological design features using computer-assisted interviewing (CAI) methods, with the sensitive questions on drug use administered by audio computer-assisted self-interviewing (ACASI) methods, have been researched and adjusted accordingly to continually improve the survey and enhance the quality of the survey data. Efforts must be continued to refine the methods to assuage respondents' concerns about the confidentiality of the answers they provide.

Recognizing that it takes a large sample of respondents from the general population to detect something as unusual as recent use of illicit drugs, this study has shown that additional validity studies of greater magnitude are sorely needed.

There were some problems with the questions included in the survey to characterize the use of certain classes of drugs (e.g., opiates and amphetamines). These are discussed in Chapters 7 and 8. This study presented an excellent opportunity to learn more about the abuse of prescription analgesics and the ability to resolve illicit use of amphetamines[7] from legitimate use of methylphenidate and amphetamine. To gather this information, it would have been necessary to include the appropriate questions in the survey and to measure additional analytes in urine and hair specimens.

The Validity Study demonstrated that it is possible to collect urine and hair specimens in a household survey environment with a high response rate. Almost 90 percent of the youths aged 12 to 17 and young adults aged 18 to 25 who were interviewed agreed to provide either a hair or urine specimen, and about 81 percent provided both. Slightly more provided a hair specimen than provided a urine specimen. Unfortunately, many of the hair specimens had an insufficient quantity for testing.

In general, a 3 to 4 centimeter (cm) segment of hair closest to the scalp is collected and analyzed for drugs and is thought to reflect drug use over approximately 3 months. The Validity Study was designed to collect a 1.3 cm segment of hair to reflect drug use over the previous 30 days, a period of time covered in the core questions of the NHSDA. However, the same amount of hair (the size of a pencil eraser) was collected for the 1.3 cm specimen as routinely utilized for the 3 to 4 cm or 3-month test. The amount collected proved to be insufficient for the testing laboratory to conduct both the screening and confirmation drug tests for multiple drugs.

[7] For example, the illicit use of methamphetamine, amphetamine, methylene dioxymethamphetamine (MDMA) or Ecstasy, 3,4-methylenedioxy-*N*-ethylamphetamine (MDEA), and methylene dioxyamphetamine (MDA).

It was somewhat more difficult to obtain hair specimens from men than women, 82.9 percent as compared with 86.6 percent (see Appendix F, Table F.3). The data also show that a lower percentage of black respondents' hair specimens were obtained when compared with white respondents' specimens (i.e., 76.8 and 86.8 percent, respectively). The explanation for these demographic differences may be cultural (i.e., most men have shorter lengths of hair than women and many African-American men and women select shorter hairstyles). Fendrich et al. (1999b) noted a similar problem of insufficient hair specimens in a household study of the general population. Those authors raised concern over sample bias, noting key demographic differences (e.g., race, gender) between participants and nonparticipants. This was not evaluated in the Validity Study report.

The cutoff concentrations used by the testing laboratory were too high for hair tests, especially for opiates and cannabinoids. The cutoffs utilized did not meet industry standards at the time of the testing. Thus, for marijuana, the primary drug of abuse throughout the world, there were no positive hair tests in year 1 and only 18 in year 2, dramatically underrepresenting marijuana use.

The testing laboratory did not use the most sensitive instrumentation or methodology available. The laboratory attempted to use gas chromatography/mass spectrometry (GC/MS) for all drug analytes in hair. This method does not provide sufficient sensitivity for some analytes. In addition, the selected cutoff concentrations and monitored analytes were not always appropriate. This was especially evident for the opiate assay when use of codeine, morphine, heroin, poppy seeds, or prescription analgesics (e.g., oxycodone, hydromorphone) could not be distinguished, severely limiting the usefulness of these data. Similarly, use of methylphenidate for attention deficit-hyperactivity disorder is particularly prevalent in this age group; therefore, the evaluation of the validity of self-report could be hampered for the amphetamines class as well.

Basic drugs (e.g., cocaine, methamphetamine) are deposited into hair in greater relative amounts than neutral or acidic drugs (e.g., cannabinoids). This remains a controversial topic, but much research has documented higher concentrations of basic drugs in dark, coarse hair than in fine, blond hair. Thus, detection of drugs such as cocaine and methamphetamine in hair is facilitated, while detection of cannabinoid analytes in hair after marijuana use is problematic. Although hair testing offers the advantage of a wider window of drug detection than urine for basic drugs, it has reduced sensitivity for detecting marijuana use.

There are two additional issues concerning the use of hair testing to validate the truthfulness of self-reports of drug use: the potential for external contamination of the hair with drug from the environment, and the requirement for sophisticated and expensive technology to detect low analyte concentrations (picograms of drug per milligram of hair). Contamination of hair with drug from the environment can potentially be misinterpreted as drug use. For example, the only true metabolite detected in hair testing is carboxy-THC, a metabolite of delta-9-tetrahydrocannabinol (THC), the major psychoactive compound in marijuana. Measurement of the extremely small concentrations of drug analytes found in hair requires sensitivity in the range of 0.05 to 200 pg/mg of hair, depending upon the drug of interest. Although low concentration cutoffs were utilized for urine testing in the Validity Study, relatively high cutoff concentrations were employed when testing hair for marijuana use, limiting the significance of the hair test results.

Windows of drug detection vary based on the matrix tested, drug analytes monitored, analytical method employed, cutoffs applied, and the route, amount, and frequency of drug consumed. For the drugs tested in the study, a single use occasion may be detected in the urine for up to 2 to 7 days. The amount of drug ingested, the route and frequency of ingestion, the time since last ingested, the individual's state of hydration, and urinary frequency are all factors that influence the amount of drug found in a urine specimen. However, those respondents reporting drug use more than 7 days previously generally would not be expected to test positive. It could be argued that some of the inconsistency in the self-reports and urine test results were due to the fact that expected detection times for drug analytes in urine are not exact, but are estimated time ranges based on research and case studies.

The testing laboratory used two different immunoassay methods for the urine drug screening test during the study: fluorescence polarization immunoassay (FPIA) was used from January 2000 to May 2001, and enzyme-multiplied immunoassay technique (EMIT) was used from June 2001 to December 2001. Because these two test methods have different performance characteristics (i.e., different specificity for the drugs of interest, different cross-reactivity to other compounds), the change may have introduced variability in drug test results.

Future studies could include analysis of a proportion of the urine and hair specimens with negative screening test results by the more sensitive and specific confirmation methodology to gain some perspective on false-negative drug tests. This would allow researchers to critically evaluate the appropriateness of the immunoassays and cutoffs that were employed.

The Validity Study only employed a screening immunoassay to test for cotinine in urine without a confirmatory test, so the results should be interpreted with some caution in regard to tobacco use. Although it would have been preferable to use a confirmatory test to quantify cotinine, there were no good confirmatory test methods readily available at the time of the study. Tobacco has been one of the least researched drugs in terms of developing guidelines for urine testing.

A number of respondents who reported high concentrations of tobacco or marijuana use unexpectedly tested negative in urine, even though the Validity Study cutoff concentrations were set to capture light or infrequent users. More concern generally is focused on those who test positive and deny use, but overreporters need to be understood as well. Most will find it easier to interpret the underreporters as being untruthful than to interpret the overreporters as such. Many of the overreporters probably were telling the truth about their use of the drug, but the drug tests were not designed to identify the drugs they were using (e.g., many of prescription pain relievers and stimulants listed in the survey questions). More comprehensive urine drug testing involving immunoassays with increased cross-reactivity and expanded confirmation test panels, with additional drug analytes, may resolve many of the discordant cases (i.e., under- and overreporting).

Drug test results are only pieces of information. Ideally, this information should be integrated into a meaningful whole by a human evaluator who, equipped with additional knowledge, will be less likely to make an error in judgment than if he or she were deprived of that information (Mieczkowski & Newel, 1997). Researchers employing drug tests in epidemiological studies must be knowledgeable of the performance characteristics of analytical

procedures used for the tests (e.g., capabilities of the test methods, validation of procedures used by the testing laboratory), as well as the pharmacology of the tested drugs, to enable an acceptable study design and correct interpretation of the drug test results in the different biological specimen matrices.

References

Akinci, I. H., Tarter, R. E., & Kirisci, L. (2001). Concordance between verbal report and urine screen of recent marijuana use in adolescents. *Addictive Behaviors, 26*, 613-619.

Ambre, J. (1985). The urinary excretion of cocaine and metabolites in humans: A kinetic analysis of published data. *Journal of Analytical Toxicology, 9*, 241-245.

Babor, T. F., Brown, J., & DelBoca, F. K. (1990). Validity of self-reports in applied research on addictive behaviors: Fact or fiction? *Behavioral Assessment, 12*, 5-31.

Benuck, I., Gidding, S. S., & Binns, H. J. (2001). Identification of adolescent tobacco users in a pediatric practice. *Archives of Pediatrics & Adolescent Medicine, 155*, 32-35.

Beumer, J. H., Bosman, I. J., & Maes, R. A. (2001). Hair as a biological specimen for therapeutic drug monitoring. *International Journal of Clinical Practice, 55*, 353-357.

Bono, R., Arossa, W., Russo, R., Scursatone, E., Castello, D., & Gilli, G. (1994). Environmental tobacco smoke and urinary cotinine in a group of adolescents. *Journal of Environmental Science and Health, A29*, 1439-1449.

Borges, C. R., Wilkins, D. G., & Rollins, D. E. (2001). Amphetamine and *N*-acetylamphetamine incorporation into hair: An investigation of the potential role of drug basicity in hair color bias. *Journal of Analytical Toxicology, 25*, 221-227.

Bowman, K. R., Penne, M. A., Chromy, J. R., & Odom, D. M. (2001, April). 1999 National Household Survey on Drug Abuse: Sample design report. In *1999 National Household Survey on Drug Abuse: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008, Deliverable No. 10, RTI/7190.130.400). Research Triangle Park, NC: RTI. [Available as a PDF at http://www.oas.samhsa.gov/NHSDA/1999/SampleDesign/index.htm]

Bronson, M., Latour, C., & Nahas, G. G. (1984). Distribution and disposition of delta-9-Tetrahydrocannabinol (THC) in different tissues of the rat. In S. Agurell, W. L. Dewey, & R. E. Willette (Eds.), *The cannabinoids: Chemical, pharmacologic and therapeutic aspects* (pp. 309-317). New York: Academic Press.

Burke, W. M., Ravi, N. V., Dhopesh, V., Vandegrift, B., & Maany, I. (1990). Prolonged presence of metabolite in urine after compulsive cocaine use. *Journal of Clinical Psychiatry, 51*, 145-148.

Cairns, T., Hill, V., Schaffer, M., & Thistle, W. (2004). Removing and identifying drug contamination in the analysis of human hair. *Forensic Science International, 145*(2-3), 97-108.

Charles, B. K., Day, J. E., Rollins, D. E., Andrenyak, D., Ling, W., & Wilkins, D. G. (2003). Opiate recidivism in a drug-treatment program: Comparison of hair and urine data. *Journal of Analytical Toxicology, 27*, 412-428.

Chen, P., Emrich, S. L., Gordek, H., & Singh, A. C. (2006, January). *2000–2001 National Household Surveys on Drug Abuse: Sampling weight calibration for the 2000 and 2001 Validity of Self-Reported Drug Use in Population Surveys (VSRDU)* (Report No. RTI/07190, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/NSDUH/Methods.cfm]

Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the 1979 American Statistical Association, Survey Research Methods Section, Washington, DC* (pp. 401-406). Washington, DC: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/SRMS/proceedings/]

Chromy, J. R. (1981). Variance estimators for a sequential sample selection procedure. In D. Krewski, R. Platek, & J. N. K. Rao (Eds.), *Current topics in survey sampling: Proceedings of the symposium held May 7-9, 1980, sponsored by the Ottawa Chapter and the Survey Research Methods Section of the American Statistical Association* (pp. 329-347). New York: Academic Press.

Clark, J. P., & Tifft, L. L. (1966). Polygraph and interview validation of self-reported deviant behavior. *American Sociological Review, 31*, 516-523.

Clauwaert, K. M., Van Bocxlaer, J. F., Lambert, W. E., & De Leenheer, A. P. (2000). Segmental analysis for cocaine and metabolites by HPLC in hair of suspected drug overdose cases. *Forensic Science International, 110*(3), 157-166.

Colon, H. M., Robles, R. R., & Sahai, H. (2002). The validity of drug use self-reports among hard core drug users in a household survey in Puerto Rico: Comparison of survey responses of cocaine and heroin use with hair tests. *Drug and Alcohol Dependence, 67*, 269-279.

Cone, E. (1994a, October 29-30). *Drug testing in hair*. Paper presented at the SOFT Conference on Drug Testing in Hair, Tampa, FL.

Cone, E. (1994b, October 29-30). *Potential for ethnic bias in hair testing*. Paper presented at the SOFT Conference on Drug Testing in Hair, Tampa, FL.

Cone, E. J. (1997). New developments in biological measures of drug prevalence. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 108-130). Rockville, MD: National Institute on Drug Abuse. [Available at http://www.drugabuse.gov/pdf/monographs/monograph167/download167.html]

Cone, E. J., & Johnson, R. E. (1986). Contact highs and urinary cannabinoid excretion after passive exposure to marijuana smoke. *Clinical Pharmacology and Therapeutics, 40*, 247-256.

Cone, E. J., Johnson, R. E., Darwin, W. D., Yousefnejad, D., Mell, L. D., Paul, B. D., & Mitchell, J. (1987). Passive inhalation of marijuana smoke: Urinalysis and room air levels of delta-9-tetrahydrocannabinol. *Journal of Analytical Toxicology, 11*, 89-96.

Cone, E. J., Tsadik, A., Oyler, J., & Darwin, W. D. (1998). Cocaine metabolism and urinary excretion after different routes of administration. *Therapeutic Drug Monitoring, 20*, 556-560.

Cone, E. J., & Weddington, W. W. Jr. (1989). Prolonged occurrence of cocaine in human saliva and urine after chronic use. *Journal of Analytical Toxicology, 13*, 65-68.

Cook, R. F., & Bernstein, A. (1994). Assessing drug use prevalence in the workplace: A comparison of self-report methods and urinalysis. *International Journal of the Addictions, 29*, 1057-1068.

Cook, R. F., Bernstein, A. D., & Andrews, C. M. (1997). Assessing drug use in the workplace: A comparison of self-report, urinalysis, and hair analysis. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 247-272). Rockville, MD: National Institute on Drug Abuse. [Available at http://www.drugabuse.gov/pdf/monographs/monograph167/download167.html]

Council on Scientific Affairs. (1987). Scientific issues in drug testing. *Journal of the American Medical Association , 257*, 3110-3114.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology, 24*, 349-354.

Dackis, C. A., Pottash, A. L., Annitto, W., & Gold, M. S. (1982). Persistence of urinary marijuana levels after supervised abstinence. *American Journal of Psychiatry, 139*, 1196-1198.

Davis, K. H., McDaniel, I. A., Cadwell, L. W., & Moody, P. L. (1984). Some smoking characteristics of marijuana cigarettes. In S. Agurell, W. L. Dewey, & R. E. Willette (Eds.), *The cannabinoids: Chemical, pharmacologic and therapeutic aspects* (pp. 245-261). New York: Academic Press.

DuPont, R. L., & Baumgartner, W. A. (1995). Drug testing by urine and hair analysis: Complementary features and scientific issues. *Forensic Science International, 70*(1-3), 63-76.

Ellis, G. M. Jr., Mann, M. A., Judson, B. A., Schramm, N. T., & Tashchian, A. (1985). Excretion patterns of cannabinoid metabolites after last use in a group of chronic users. *Clinical Pharmacology and Therapeutics, 38*, 572-578.

Fendrich, M., Johnson, T. P., Sudman, S., Wislar, J. S., & Spiehler, V. (1999a). Validity of drug use reporting in a high-risk community sample: A comparison of cocaine and heroin survey reports with hair tests. *American Journal of Epidemiology, 149*, 955-962.

Fendrich, M., Johnson, T., Wislar, J. S., & Sudman, S. (1999b). The feasibility of hair testing in a household survey on drug abuse. In T. Mieczkowski (Ed.), *Drug testing methods: Assessment of field applications* (pp. 235-253). Boca Raton, FL: CRC Press.

Fendrich, M., Johnson, T. P., Wislar, J. S., & Hubbell, A. (2004). Drug test feasibility in a general population household survey. *Drug and Alcohol Dependence, 73*, 237-250.

Fendrich, M., & Xu, Y. (1994). The validity of drug use reports from juvenile arrestees. *International Journal of the Addictions, 29*, 971-985.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 598-603). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/SRMS/proceedings/]

Gygi, S. P., Joseph, R. E. Jr., Cone, E. J., Wilkins, D. G., & Rollins, D. E. (1996). Incorporation of codeine and metabolites into hair: Role of pigmentation. *Drug Metabolism and Disposition: The Biological Fate of Chemicals, 24*, 495-501.

Harkey, M. R., & Henderson, G. L. (1989). Hair analysis for drugs of abuse. In R. C. Baselt (Ed.), *Advances in analytical toxicology* (pp. 298-329). Chicago: Year Book Medical Publishers.

Harrell, A. V. (1997). The validity of self-reported drug use data: The accuracy of responses on confidential self-administered answer sheets. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 37-58). Rockville, MD: National Institute on Drug Abuse. [Available as a PDF at http://www.drugabuse.gov/pdf/monographs/monograph167/download167.html]

Harrison, L. (1997). The validity of self-reported drug use in survey research: An overview and critique of research methods. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 17-36). Rockville, MD: National Institute on Drug Abuse. [Available as a PDF at http://www.drugabuse.gov/pdf/monographs/monograph167/download167.html]

Harrison, L. D. (1992). The validity of self-reported drug use. In *Proceedings of the Third Annual Conference on Drug Issues and Drug Policy, September 28-30, 1992* (pp. 158-177). Loughborough, Leicestershire, England: European Social Science Research Group on Drug Issues and the Loughborough University of Technology.

Harrison, L. D. (1995). The validity of self-reported data on drug use. *Journal of Drug Issues, 25*, 91-111.

Harrison, L. D. (2001). Understanding the differences in youth drug prevalence rates produced by the MTF, NHSDA, and YRBS studies. *Journal of Drug Issues, 31*, 665-694. [Special issue, available to paid subscribers at http://www2.criminology.fsu.edu/~jdi/31n3.htm]

Harrison, L. D., & O'Neil, J. A. (1989, January 12-13). *Concealment of recent drug use among arrestees*. Paper presented at the Drug Use Forecasting Project Directors Meeting, Phoenix, AZ.

Haufroid, V., & Lison, D. (1998). Urinary cotinine as a tobacco-smoke exposure index: A minireview. *International Archives of Occupational and Environmental Health, 71*(3), 162-168.

Hawks, R. L., & Chiang, C. N. (1986; reprinted 1987). Examples of specific drug assays. In R. L. Hawks & C. N. Chiang (Eds.), *Urine testing for drugs of abuse* (DHHS Publication No. ADM 87-1481, NIDA Research Monograph 73, pp. 84-114). Rockville, MD: National Institute on Drug Abuse. [Available at http://www.drugabuse.gov/pdf/monographs/download73.html]

Hecht, S. S., Carmella, S. G., Murphy, S. E., Akerkar, S., Brunnemann, K. D., & Hoffmann, D. (1993). A tobacco-specific lung carcinogen in the urine of men exposed to cigarette smoke. *New England Journal of Medicine, 329*, 1543-1546.

Henderson, G. L., Harkey, M. R., & Jones, R. (1993). *Hair analysis for drugs of abuse: Final report* (Grant No. NIJ90-NIJ-CX-0012). Washington, DC: National Institute of Justice, National Institute on Drug Abuse.

Henderson, G. L., Harkey, M. R., & Jones, R. T. (1995). Analysis of hair for cocaine. In E. J. Cone, M. J. Welch, & M. B. Grigson Babecki (Eds.), *Hair testing for drugs of abuse: International research on standards and technology* (NIH Publication No. 95-3727, pp. 91-120). Rockville, MD: National Institute on Drug Abuse.

Henderson, G. L., Harkey, M. R., Zhou, C., Jones, R. T., & Jacob, P. 3rd. (1996). Incorporation of isotopically labeled cocaine and metabolites into human hair: 1. Dose-response relationships. *Journal of Analytical Toxicology, 20*, 1-12.

Herzig, K. E., Callaway, E., Halliday, R., Naylor, H., & Benowitz, N. L. (1998). Effects of cotinine on information processing in nonsmokers. *Psychopharmacology, 135*, 127-132.

Hindin, R., McCusker, J., Vickers-Lahti, M., Bigelow, C., Garfield, F., & Lewis, B. (1994). Radioimmunoassay of hair for determination of cocaine, heroin, and marijuana exposure: Comparison with self-report. *International Journal of the Addictions, 29*, 771-789.

Hoffman, B. H. (1999). Analysis of race effects on drug-test results. *Journal of Occupational and Environmental Medicine/American College of Occupational and Environmental Medicine, 41*, 612-614.

Holl, R. W., Grabert, M., Heinze, E., & Debatin, K. M. (1998). Objective assessment of smoking habits by urinary cotinine measurement in adolescents and young adults with type 1 diabetes: Reliability of reported cigarette consumption and relationship to urinary albumin excretion. *Diabetes Care, 21*, 787-791.

Huestis, M. A., Henningfield, J. E., & Cone, E. J. (1992). Blood cannabinoids. I. Absorption of THC and formation of 11-OH-THC and THCCOOH during and after smoking marijuana. *Journal of Analytical Toxicology, 16*, 276-282.

Hunt, C. A., & Jones, R. T. (1980). Tolerance and disposition of tetrahydrocannabinol in man. *Journal of Pharmacology and Experimental Therapeutics, 215*, 35-44.

Jacob, P. 3rd, Yu, L., Shulgin, A. T., & Benowitz, N. L. (1999). Minor tobacco alkaloids as biomarkers for tobacco use: Comparison of users of cigarettes, smokeless tobacco, cigars, and pipes. *American Journal of Public Health, 89*, 731-736.

Jarvis, M. J., Tunstall-Pedoe, H., Feyerabend, C., Vesey, C., & Saloojee, Y. (1987). Comparison of tests used to distinguish smokers from nonsmokers. *American Journal of Public Health, 77*, 1435-1438.

Jenny, R. (1989). Quality assurance. In D. G. Deutsch (Ed.), *Analytic aspects of drug testing*. New York: John Wiley & Sons.

Johansson, E., & Halldin, M. M. (1989). Urinary excretion half-life of delta 1-tetrahydrocannabinol-7-oic acid in heavy marijuana users after smoking. *Journal of Analytical Toxicology, 13*, 218-223.

Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (1988). *Illicit drug use, smoking, and drinking by America's high school students, college students, and young adults: 1975-1987* (DHHS Publication No. ADM 89-1602). Rockville, MD: National Institute on Drug Abuse.

Jones, R. T. (1980). Human effects: An overview. In R. C. Petersen (Ed.), *Marijuana research findings: 1980* (DHHS Publication No. ADM 80-1001, NIDA Research Monograph 31, pp. 54-80). Rockville, MD: National Institute on Drug Abuse. [Available as a PDF at http://www.nida.nih.gov/pdf/monographs/download31.html]

Joseph, R. E. Jr., Hold, K. M., Wilkins, D. G., Rollins, D. E., & Cone, E. J. (1999). Drug testing with alternative matrices II. Mechanisms of cocaine and codeine deposition in hair. *Journal of Analytical Toxicology, 23*, 396-408.

Jufer, R. A., Walsh, S. L., & Cone, E. J. (1998, October 8). *Urinary detection times following chronic oral cocaine administration*. Presented at Society of Forensic Toxicologists and the International Association of Forensic Toxicologists Joint Annual Meeting; Albuquerque, NM.

Jufer, R. A., Wstadik, A., Walsh, S. L., Levine, B. S., & Cone, E. J. (2000). Elimination of cocaine and metabolites in plasma, saliva, and urine following repeated oral administration to human volunteers. *Journal of Analytical Toxicology, 24*, 467-477.

Jurado, C., Kintz, P., Menendez, M., & Repetto, M. (1997). Influence of the cosmetic treatment of hair on drug testing. *International Journal of Legal Medicine, 110*, 159-163.

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*(1), 31-41.

Kelly, P., & Jones, R. T. (1992). Metabolism of tetrahydrocannabinol in frequent and infrequent marijuana users. *Journal of Analytical Toxicology, 16*, 228-235.

Kelly, R. C., Mieczkowski, T., Sweeney, S. A., & Bourland, J. A. (2000). Hair analysis for drugs of abuse: Hair color and race differentials or systematic differences in drug preferences? *Forensic Science International, 107*(1-3), 63-86.

Kidwell, D. A., & Blank, D. L. (1994, October 29-30). *Incorporation of drugs of abuse into and their removal from hair*. Paper presented at the SOFT Conference on Drug Testing in Hair, Tampa, FL.

Kidwell, D. A., & Blank, D. L. (1995). Mechanisms of incorporation of drugs into hair and the interpretation of hair analysis data. In E. J. Cone, M. J. Welch, & M. B. Grigson Babecki (Eds.), *Hair testing for drugs of abuse: International research on standards and technology* (NIH Publication No. 95-3727, pp. 19-90). Rockville, MD: National Institute on Drug Abuse.

Knight, J. M., Eliopoulos, C., Klein, J., Greenwald, M., & Koren, G. (1996). Passive smoking in children: Racial differences in systemic exposure to cotinine by hair and urine analysis. *Chest, 109*, 446-450.

Kronstrand, R., Forstberg-Peterson, S., Kagedal, B., Ahlner, J., & Larson, G. (1999). Codeine concentration in hair after oral administration is dependent on melanin content. *Clinical Chemistry, 45*, 1485-1494.

Kronstrand, R., Nystrom, I., Josefsson, M., & Hodgins, S. (2002). Segmental ion spray LC-MS-MS analysis of benzodiazepines in hair of psychiatric patients. *Journal of Analytical Toxicology, 26*, 479-484.

Law, B., Mason, P. A., Moffat, A. C., & King, L. J. (1984). Confirmation of cannabis use by the analysis of delta 9-tetrahydrocannabinol metabolites in blood and urine by combined HPLC and RIA. *Journal of Analytical Toxicology, 8*, 19-22.

Lemberger, L., & Rubin, A. (1976). *Physiological disposition of drugs of abuse*. New York: Spectrum Publications, Inc.

Magura, S., Freeman, R. C., Siddiqi, Q., & Lipton, D. S. (1992). The validity of hair analysis for detecting cocaine and heroin use among addicts. *International Journal of the Addictions, 27*, 51-69.

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 the Addictions, 22*, 727-749.

Magura, S., & Kang, S.-Y. (1997). The validity of self-reported cocaine use in two high-risk populations. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 227-246). Rockville, MD: National Institute on Drug Abuse. [Available as a PDF at http://www.drugabuse.gov/pdf/monographs/monograph167/download167.html]

Makkai, T. (2000). *Drug Use Monitoring in Australia (DUMA): Drug detection testing* (Report No. 25, Australian Institute of Criminology Research and Public Policy Series). Canberra, Australian Capital Territory (ACT): Australian Institute of Criminology. [Available at http://www.aic.gov.au/publications/rpp/25/]

Mieczkowski, T. (1990). The accuracy of self-reported drug use: An evaluation and analysis of new data. In R. Weisheit (Ed.), *Drugs and crime and the criminal justice system* (pp. 275-302). Cincinnati, OH: Anderson Publishing Co.

Mieczkowski, T., Barzelay, D., Gropper, B., & Wish, E. (1991a). Concordance of three measures of cocaine use in an arrestee population: Hair, urine, and self-report. *Journal of Psychoactive Drugs, 23*, 241-249.

Mieczkowski, T., Landress, H. J., Newel, R., & Coletti, S. D. (1991b, August). *The concordance of drug use indicators: Urine, hair and self-report in an arrestee population* (Grant #90-IJ-CX-0023). Washington, DC: National Institute of Justice.

Mieczkowski, T., Mumm, R., & Connick, H. F. (1995). The use of hair analysis in a pretrial diversion program in New Orleans. *International Journal of Offender Therapy and Comparative Criminology, 39*, 222-241.

Mieczkowski, T., & Newel, R. (1997). Patterns of concordance between hair assays and urinalysis for cocaine: Longitudinal analysis of probationers in Pinellas County, Florida. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 161-199). Rockville, MD: National Institute on Drug Abuse. [Available as a PDF at http://www.drugabuse.gov/pdf/monographs/monograph167/download167.html]

Mieczkowski, T., & Newel, R. (2000). Statistical examination of hair color as a potential biasing factor in hair analysis. *Forensic Science International, 107*(1-3), 13-38.

Mumola, C. J. (1999, January). *Substance abuse and treatment: State and federal prisoners, 1997* (NCJ 172871). Washington, DC: U.S. Department of Justice, National Institute of Justice, Bureau of Justice Statistics. [Available at http://www.ojp.usdoj.gov/bjs/abstract/satsfp97.htm]

Murphy, D. A., Durako, S., Muenz, L. R., & Wilson, C. M. (2000). Marijuana use among HIV-positive and high-risk adolescents: A comparison of self-report through audio computer-assisted self-administered interviewing and urinalysis. *American Journal of Epidemiology, 152*, 805-813.

Murray, R. P., Connett, J. E., Lauger, G. G., & Voelker, H. T. (1993). Error in smoking measures: Effects of intervention on relations of cotinine and carbon monoxide to self-reported smoking: The Lung Health Study Research Group. *American Journal of Public Health, 83*, 1251-1257.

Musshoff, F., & Madea, B. (2006). New trends in hair analysis and scientific demands on validation and technical notes [Corrected proof available online]. *Forensic Science International*.

Nakahara, Y., Takahashi, K., & Kikura, R. (1995). Hair analysis for drugs of abuse. X. Effect of physicochemical properties of drugs on the incorporation rates into hair. *Biological & Pharmaceutical Bulletin, 18*, 1223-1227.

National Institute of Justice. (1990, March). *DUF: 1988 Drug Use Forecasting annual report*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.

Office of Applied Studies. (1996, April). *Development and implementation of a new data collection instrument for the 1994 National Household Survey on Drug Abuse* (DHHS Publication No. SMA 96-3084). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of National Drug Control Policy. (2006a, August 7). *Methamphetamine drug fact sheet*. Retrieved September 12, 2006, from http://www.whitehousedrugpolicy.gov/drugfact/index.html

Office of National Drug Control Policy. (2006b, August 7). *Street terms: Drugs and the drug trade*. Retrieved September 12, 2006, from http://www.whitehousedrugpolicy.gov/streetterms/

Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: A review and meta-analysis. *American Journal of Public Health, 84*, 1086-1093.

Pichini, S., Poudevida, S., Pujadas, M., Menoyo, E., Pacifici, R., Farre, M., & de la Torre, R. (2006). Assessment of chronic exposure to MDMA in a group of consumers by segmental hair analysis. *Therapeutic Drug Monitoring, 28*, 106-109.

Pokorski, T. L., Chen, W. W., & Bertholf, R. L. (1994). Use of urine cotinine to validate smoking self-reports in U.S. Navy recruits. *Addictive Behaviors, 19*, 451-454.

Potsch, L., Skopp, G., & Moeller, M. R. (1997). Influence of pigmentation on the codeine content of hair fibers in guinea pigs. *Journal of Forensic Sciences, 42*, 1095-1098.

Randall, T. (1992). Infants, children test positive for cocaine after exposure to second-hand crack smoke. *Journal of the American Medical Association, 267*, 1044-1045.

Reiter, A., Hake, J., Meissner, C., Rohwer, J., Friedrich, H. J., & Oehmichen, M. (2001). Time of drug elimination in chronic drug abusers: Case study of 52 patients in a "low-step" detoxification ward. *Forensic Science International, 119*(2), 248-253.

Riley, K. J., Lu, N. T., & Taylor, B. G. (2000). Drug screening: A comparison of urinalysis results from two independent laboratories. *Journal of Drug Issues, 30*, 171-186.

Rollins, D. E., Wilkins, D. G., Krueger, G. G., Augsburger, M. P., Mizuno, A., O'Neal, C., Borges, C. R., & Slawson, M. H. (2003). The effect of hair color on the incorporation of codeine into human hair. *Journal of Analytical Toxicology, 27*, 545-551.

Romano, G., Barbera, N., & Lombardo, I. (2001). Hair testing for drugs of abuse: Evaluation of external cocaine contamination and risk of false positives. *Forensic Science International, 123*(2-3), 119-129.

Romano, G., Barbera, N., Spadaro, G., & Valenti, V. (2003). Determination of drugs of abuse in hair: Evaluation of external heroin contamination and risk of false positives. *Forensic Science International, 131*(2-3), 98-102.

Rosenberg, N. M., Marino, D., Meert, K. L., & Kauffman, R. F. (1995). Comparison of cocaine and opiate exposures between young urban and suburban children. *Archives of Pediatrics & Adolescent Medicine, 149*, 1362-1364.

Rouse, B. A., Kozel, N. J., & Richards, L. G. (Eds.). (1985). *Self-report methods of estimating drug use: Meeting current challenges to validity* (DHHS Publication No. ADM 85-1402, NIDA Research Monograph 57). Rockville, MD: National Institute on Drug Abuse. [Available as a PDF at http://www.drugabuse.gov/pdf/monographs/download57.html]

Schaffer, M., Hill, V., & Cairns, T. (2005). Hair analysis for cocaine: The requirement for effective wash procedures and effects of drug concentration and hair porosity in contamination and decontamination. *Journal of Analytical Toxicology, 29*, 319-326.

Schaffer, M. I., Wang, W. L., & Irving, J. (2002). An evaluation of two wash procedures for the differentiation of external contamination versus ingestion in the analysis of human hair samples for cocaine. *Journal of Analytical Toxicology, 26*, 485-488.

Scheidweiler, K. B., Cone, E. J., Moolchan, E. T., & Huestis, M. A. (2005). Dose-related distribution of codeine, cocaine, and metabolites into human hair following controlled oral codeine and subcutaneous cocaine administration. *Journal of Pharmacology and Experimental Therapeutics, 313*, 909-915.

Schuman, H., & Presser, S. (1981). The attitude-action connection and the issue of gun control. *Annals of American Academy of Political and Social Science*, (455), 40-47.

Shah, B. V., Barnwell, B. G., & Bieler, G. S. (1997). *SUDAAN user's manual*. Research Triangle Park, NC: Research Triangle Institute.

Skopp, G., Potsch, L., & Moeller, M. R. (1997). On cosmetically treated hair—Aspects and pitfalls of interpretation. *Forensic Science International, 84*(1-3), 43-52.

Smith, F. P., Kidwell, D. A., & Cook, L. F. (1994, October 29-30). *Cocaine in children's hair when they live with drug dependent adults*. Paper presented at the SOFT Conference on Drug Testing in Hair, Tampa, FL.

Stephens, R. (1972). The truthfulness of addict respondents in research projects. *International Journal of the Addictions, 7*, 549-558.

Stout, P. R., Ropero-Miller, J. D., Baylor, M. R., & Mitchell, J. M. (2006). External contamination of hair with cocaine: Evaluation of external cocaine contamination and development of performance-testing materials. *Journal of Analytical Toxicology, 30*, 490-500.

Substance Abuse and Mental Health Services Administration, Division of Workplace Programs. (2000-2001). *Drafts of proposed revisions to the 1994 guidelines*. Rockville, MD: Author. [Available at http://dwp.samhsa.gov/DrugTesting/Level_1_Pages/Drug%20Testing%20Archives.aspx]

Substance Abuse and Mental Health Services Administration, Division of Workplace Programs. (2004, April 13). *Mandatory guidelines for federal workplace drug-testing programs*. *Federal Regist*er, 69, 19644-19673. [Available at http://dwp.samhsa.gov/FedPgms/Fed_DFWP.aspx]

Substance Abuse and Mental Health Services Administration, Division of Workplace Programs. (2004, April 13). *Notice of proposed revisions to the mandatory guidelines for federal workplace drug testing programs*. *Federal Register, 69*, 19673-19732. [Available at http://dwp.samhsa.gov/FedPgms/Fed_DFWP.aspx]

Substance Abuse and Mental Health Services Administration, Division of Workplace Programs. (2006). *Alternative specimens pilot PT program results*. Retrieved December 1, 2006, from http://dwp.samhsa.gov/DrugTesting/Level_1_Pages/Alternative%20Specimens.aspx

Turner, C. E. (1980). Chemistry and metabolism. In R. C. Petersen (Ed.), *Marijuana research findings: 1980* (DHHS Publication No. ADM 80-1001, NIDA Research Monograph 31, pp. 81-97). Rockville, MD: National Institute on Drug Abuse. [Available as a PDF at http://www.nida.nih.gov/pdf/monographs/download31.html]

Turner, C. F., Lessler, J. T., & Devore, J. W. (1992). Effects of mode of administration and wording on reporting of drug use. In C. F. Turner, J. T. Lessler, & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (DHHS Publication No. ADM 92-1929, pp. 177-219). Rockville, MD: National Institute on Drug Abuse.

U.S. Department of Health and Human Services. (2006, June 27). *The health consequences of involuntary exposure to tobacco smoke: A report of the Surgeon General*. Rockville, MD, & Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. [Available at http://www.surgeongeneral.gov/library/secondhandsmoke/report/]

U.S. Department of Justice, Drug Enforcement Administration. (2005). *Drugs of abuse*. Retrieved September 12, 2006, from http://www.usdoj.gov/dea/pubs/abuse/index.htm

Vine, M. F., Hulka, B. S., Margolin, B. H., Truong, Y. K., Hu, P. C., Schramm, M. M., Griffith, J. D., McCann, M., & Everson, R. B. (1993). Cotinine concentrations in semen, urine, and blood of smokers and nonsmokers. *American Journal of Public Health, 83*, 1335-1338.

Wall, M. A., Johnson, J., Jacob, P., & Benowitz, N. L. (1988). Cotinine in the serum, saliva, and urine of nonsmokers, passive smokers, and active smokers. *American Journal of Public Health, 78*, 699-701.

Wall, M. E., Brine, D. R., Pitt, C. G., & Perez-Reyes, M. (1972). Identification of delta-9-tetrahydrocannabinol and metabolites in man. *Journal of the American Chemical Society, 94*, 8579-8581.

Wang, W. L., & Cone, E. J. (1995). Testing human hair for drugs of abuse. IV. Environmental cocaine contamination and washing effects. *Forensic Science International, 70*(1-3), 39-51.

Weatherby, N. L., Needle, R., Cesari, H., Booth, R., McCoy, C. B., Watters, J. K., Williams, M., & Chitwood, D. D. (1994). Validity of self-reported drug use among injection drug users and crack cocaine users recruited through street outreach. *Evaluation and Program Planning, 17*, 347-355.

Weiss, R. D., & Gawin, F. H. (1988). Protracted elimination of cocaine metabolites in long-term high-dose cocaine abusers. *American Journal of Medicine, 85*, 879-880.

Welch, M. J., & Sniegoski, L. T. (1995). Interlaboratory studies on the analysis of hair for drugs of abuse: Results from three exercises. In E. J. Cone, M. J. Welch, & M. B. Grigson Babecki (Eds.), *Hair testing for drugs of abuse: International research on standards and technology* (NIH Publication No. 95-3727, pp. 133-147). Rockville, MD: National Institute on Drug Abuse.

Welch, M. J., Sniegoski, L. T., Allgood, C. C., & Habram, M. (1993). Hair analysis for drugs of abuse: Evaluation of analytical methods, environmental issues, and development of reference materials. *Journal of Analytical Toxicology, 17*, 389-398.

White, J. M., & Irvine, R. J. (1999). Mechanisms of fatal opioid overdose. *Addiction , 94*, 961-972.

Wilkins, D. G., Haughey, H. M., Krueger, G. G., & Rollins, D. E. (1995). Disposition of codeine in female human hair after multiple-dose administration. *Journal of Analytical Toxicology, 19*, 492-498.

Wilkins, D. G., Rollins, D. E., Valdez, A. S., Mizuno, A., Krueger, G. G., & Cone, E. J. (1999). A retrospective study of buprenorphine and norbuprenorphine in human hair after multiple doses. *Journal of Analytical Toxicology, 23*, 409-415.

Wish, E. D. (1990-1991). U.S. drug policy in the 1990s: Insights from new data from arrestees. *International Journal of the Addictions, 25*(3A), 377-409.

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* (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 200-226). Rockville, MD: National Institute on Drug Abuse. [Available as a PDF at http://www.drugabuse.gov/pdf/monographs/monograph167/download167.html]

Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Reports, 62*, 1432-1449.

Zanis, D. A., McLellan, A. T., & Randall, M. (1994). Can you trust patient self-reports of drug use during treatment? *Drug and Alcohol Dependence, 35*, 127-132.

Appendix A:
Prior Research on the Validity of Self-Report

The possibility that respondents "edit" their survey answers to what they perceive is the socially desirable response has long been a concern in survey research (Crowne & Marlowe, 1960; Schuman & Presser, 1981), and the concern is particularly salient for drug use (Babor et al., 1990; Rouse et al., 1985). There is some small amount of overreporting, particularly among youths who may boast about drugs they have not used. However, practically and theoretically, it is the possibility of underreporting that may attenuate the validity of self-report of drug use. Due to advances in drug testing, it is possible to determine the accuracy of self-reported drug use in recent timeframes. However, there is no "gold standard" against which questions on attitudes and behaviors other than recent drug use can be compared.

Perhaps what is most surprising about survey research on drugs is not that some people are not truthful, but rather that, with some safeguards, most drug users will admit to illicit drug use on questionnaires. There is *prima facie* evidence of this truth telling based on data from the Monitoring the Future (MTF) study, which in the late 1980s showed over 80 percent of young adults reporting lifetime experience with some illicit drug by the time they were 27 years old (Johnston, O'Malley, & Bachman, 1988). We expect these rates to be reasonably accurate as there is little perceived reason for respondents to exaggerate their experimentation with illicit drugs that occurred at an earlier time in their lives. Among State prison inmates in 1997, 83 percent reported using an illicit drug in their lifetime (Mumola, 1999). Although this is not much higher than evident among young adults a decade earlier, representative surveys of prison inmates show much higher rates of regular drug use, past month drug use, and use of "harder" drugs, such as cocaine and heroin, than the general population. Similarly, studies of individuals seeking substance abuse treatment regularly reveal high rates of drug use (Weatherby et al., 1994; Zanis, McLellan, & Randall, 1994). Therefore, there is evidence showing that people are willing to self-report their illicit drug use on the confidential questionnaires utilized in survey research.

Because illicit drug use is by definition illegal, the task is even more complicated than, for example, measuring political party affiliation or other behaviors that are not illegal or socially proscribed. Respondents must feel assured that they can answer honestly without fear of recrimination and that their anonymity and confidentiality are protected. This is critical, but other features inherent in the interview process also can cause distortions in the data, such as respondent recall and question clarity. Respondents may misreport due to memory lapse, poor comprehension, or desire to deceive—either to conceal undesirable attributes or to exaggerate desirable ones (Schuman & Presser, 1981). So, despite the obvious advantages of survey methods over indirect measures, such as arrests, seizures, and treatment entries, they are frequently criticized because they rely on self-reporting of sensitive and highly stigmatized behavior.

A.1 Overview of Prior Validity Studies

The history of validity studies on self-report of drug use is relatively brief because validation requires comparison with some method that we know is accurate. Efforts to examine validity prior to the advent of urine-testing technology were especially infrequent. One study designed as a criterion validity test of the National Household Survey on Drug Abuse (NHSDA) was conducted in 1986 using a known sample of former substance abuse treatment clients who were interviewed within a year of receiving drug treatment, using methods to appear as if they were randomly selected to participate in the survey. These respondents reliably reported their lifetime use, but they underreported their more recent (i.e., past year) use. Those in treatment for heroin, the most highly stigmatized drug at the time of the study, were less likely to report their heroin use than those in treatment for cocaine (Harrell, 1997). Official record checks, such as checks of criminal justice or treatment records, have been used to validate self-reports of drug use. Reports by family, close friends, or counselors also have been used (Stephens, 1972). Even polygraph tests have been used to validate self-reports of drug use (Clark & Tifft, 1966). These types of validation procedures, which rely on checking validity against an external criterion, are examining external or empirical validity.

In the past two decades, urinalysis has been the major validation technique used to check self-report obtained in questionnaires. Validation studies conducted prior to the mid-1980s using urinalysis techniques primarily with criminal justice populations suggested that drug use was reported fairly accurately in self-report surveys (Mieczkowski, 1990). However, in the mid-1980s, the introduction of new instrumentation and higher standards for forensic urine drug testing brought about increased specificity and sensitivity of urine screening and confirmation tests. Urine drug testing became considerably more effective in detecting drug use. These new procedures revealed that considerably more drug use was underreported among criminal justice samples than was previously thought.

A.1.1 Criminal Justice Populations

The first large-scale study to employ urine testing for illicit drugs was the Drug Use Forecasting (DUF) study of arrestees. Begun in several major U.S. cities in 1987, the study grew to include sites in 23 cities by 1989. The DUF study administered a short, 10-minute questionnaire to ascertain recent drug use after promising confidentiality of results. Urine specimens were collected and analyzed with an immunoassay test, enzyme-multiplied immunoassay technique (EMIT). Interviews were conducted with as much privacy as possible, but in a jail setting privacy is often not possible. Promises of confidentiality also may be viewed with more skepticism in a jail's group holding cell. It was determined that only about half of those testing positive reported to the interviewer that they had used drugs in the past 2 to 3 days (Harrison, 1992; Harrison & O'Neil, 1989; Magura, Freeman, Siddiqi, & Lipton, 1992; Mieczkowski et al., 1991a; National Institute of Justice [NIJ], 1990). The results of the DUF study continued to cast doubt on the validity of all survey research on drug use in the 1980s and 1990s (Harrison, 1995; Magura et al., 1987; Wish, 1990-1991).

In 1997, DUF was reborn as ADAM (Arrestee Drug Abuse Monitoring program). ADAM sought to improve quality control in data collection and sample selection. The DUF study essentially used a convenience sample of arrestees, conducting interviews at one central

booking facility in each of the identified cities or sites for several days in a row. Every city included in the DUF study in the late 1980s and early 1990s had more than one central booking facility. Still, the sample was not representative of all arrestees. The majority of arrestees in many sites were not taken to a central booking facility, but received a field citation or were processed at a local precinct—so they never were eligible for the DUF study. The catchment areas in the various sites were not only unstandardized, but largely unknown. Local factors often had an impact on arrest rates or changes in arrest practices (even targeted Office of Justice initiatives temporarily could change the numbers and characteristics of arrestees), so trend data had to be interpreted cautiously—especially in the late 1980s with police resources increasingly targeted to drug arrests. Intersite comparisons also were not valid due to differences in arrest practices and/or methodological procedures in the various cities. Although the DUF study was an important new drug indicator, its data rarely were presented within the context of the strengths and limitations of the study design. The DUF data were misused in making national estimates or projections.

The ADAM program made many improvements. Each site included a defined catchment area, usually a county, and results were generalized to arrestees in the catchment area. There had been an improvement in methods and a more conservative approach to reporting data, even site-specific data. Nevertheless, the DUF/ADAM studies fairly consistently found that only about half of those who tested positive for a drug reported use in the past 2 to 3 days. Studies also have shown that the discrepancy between urinalysis and self-report varies by drug type (Harrison, 1995). Still, it must be emphasized that both sampling issues and interview circumstances are such that the DUF/ADAM studies of self-report are not generalizable to other populations. Consequently, results of validation studies employing urinalysis techniques with criminal justice groups cannot be generalized to other populations.

A.1.2 Treatment Populations

Contrary to the DUF/ADAM findings, there are studies in the literature that suggest relatively high rates of concordance between self-reported drug use and urinalysis results. For example, Zanis et al. (1994) found that, for a sample of patients in methadone treatment for at least 6 months, only 13 and 19 percent of those with positive EMIT urine specimens for opiates and cocaine, respectively, failed to self-report use in the prior month. And, on the other hand, 58 and 28 percent of those with negative urine specimens for opiates and cocaine, respectively, reported use of the drug during the prior month.

Substance abuse treatment studies suggest that the validity of self-report varies by treatment status. Perhaps not surprisingly, some studies have found that self-report is more accurate at intake—with clients more likely to provide self-reports that are congruent with biological test results in the beginning stages of treatment—than they are at follow-up posttreatment (Hindin et al., 1994; Wish, Hoffman, & Nemes, 1997). For example, a study by Hindin et al. (1994) compared hair test results with self-reported drug use from 109 entrants to two New England treatment facilities: 89 percent of the 87 with positive cocaine hair tests and 96 percent of the 45 with positive heroin hair tests were confirmed by self-report. However, among the 86 followed up, only 51 percent of the 43 with positive cocaine hair tests and 67 percent of the 18 with positive heroin hair tests were confirmed by self-report. This suggests the social

desirability hypothesis in which respondents may wish the interviewers, and perhaps themselves, to believe that the treatment was useful.

The results of urinalyses from 154 subjects in four cities in a study of human immunodeficiency virus (HIV) risk behaviors showed that 71.2 percent tested positive and 73.2 percent reported using cocaine in the past 48 hours (Weatherby et al., 1994). This was a highly drug-involved sample, with 76 percent reporting injecting drugs in the past 30 days. Self-reports and urinalysis results agreed for 85 percent of those reporting heroin use in the past 48 hours. Self-reported drug use in the past 48 hours was not confirmed by urinalysis among 9.7 percent of those reporting heroin use and 7.8 percent of those reporting cocaine use. Positive urinalysis results were found for 5.2 and 5.8 percent, respectively, of respondents who did not self-report heroin or cocaine use in the past 48 hours. This suggests that heavily drug-involved individuals can self-report recent drug use with some degree of validity and that jail setting and arrest status may affect the sensitivity of drug questions. It also demonstrates that urine testing does not identify drug use in all cases.

It seems clear that although drug use may vary substantially among different populations, such as arrestees, treatment clients, and the general population, the accuracy of their self-report also may vary substantially (Harrison, 1997). The research literature suggests that self-report may be the least reliable among criminal justice clients. Magura and Kang (1997) reported the results of two validity studies conducted using similar methods in New York City. One study included a random sample of patients in two methadone treatment programs, with the sampling "stratified to overrepresent patients with recent cocaine-positive clinic urinalysis" (p. 233). The patients were interviewed about drug use and provided both hair and urine specimens to the researchers. The other study included a sample of criminally involved young male adults. The young adults were recruited while they were in jail, but were followed up in the community about 5 months after release. Self-report information and hair specimens were obtained at the follow-up interviews. For the methadone sample, 60 percent self-reported recent cocaine use and 80 percent had positive hair specimens for cocaine. For the criminally involved young adult sample, 23 percent self-reported recent cocaine use, but almost 3 times as many, 67 percent, were positive for cocaine in their hair. The relatively more accurate self-reporting by the treatment clients led the researchers to conclude that the accuracy of self-report depends on the characteristics of the populations studied. An interesting aside is that, whereas only 23 percent of the criminal justice–involved young adults self-reported recent cocaine use, 75 percent reported recent marijuana use, and fully 41 percent reported drug dealing in the past month. Of particular note, 36 percent of those denying cocaine use reported dealing. This would indicate that the young adults were not afraid of self-reporting all sensitive information. Magura and Kang (1997) suggested that, for the young adults, cocaine or, more specifically, crack had become stigmatized, though dealing of these drugs was not. So, although the young adults reported drug dealing, they were more reluctant to be identified as cocaine or crack users. There also may be an important explanation overlooked by the authors, which is that hair analysis might detect cocaine that had entered the young adults' hair through environmental contamination via the young adults' handling of cocaine or crack when selling the drug or being present where it was smoked.

A.1.3 General Population

In a validity study conducted among a workplace population (which more closely resembles a general population than either treatment or criminal justice populations), Cook and Bernstein (1994) and Cook, Bernstein, and Andrews (1997) found that self-reports produced higher prevalence rates than either urinalysis or hair analysis. For the entire sample (n = 928) of employees from a large steel plant who were interviewed about drug use and had urine tests, only 7.8 percent tested positive for any drug by urinalysis, while 9.4 percent reported recent drug use. For the subsample who also had hair tests (n = 307), 6.5 percent were positive for an illicit drug, and 7.5 percent reported recent use. The most frequently detected drug was marijuana. Subjects who reported legitimate prescription drug use and tested positive for that drug were classified as negative for both self-report and drug test. However, it was noted that some of the 20 subjects who tested positive for codeine, morphine, and/or sedatives but reported no recent drug use may have failed to report legitimate prescription drug use. Because of the small number with positive drug tests, the validity analyses were combined across all drug types. Although self-report methods produced higher prevalence rates than testing, Cook and colleagues found that less than half of those positive for any drug by either a urine test or a hair test self-reported recent use. They suggested that the best strategy for prevalence assessment is to use drug tests in addition to self-report. Cook and colleagues also varied the method of data collection setting between telephone interview, personal interview in the workplace, questionnaire administration to small groups in the workplace, and personal interview offsite. They found that rates of self-reported drug use were similar for the telephone interviews and individual interviews (whether in the workplace or offsite), but were much lower for the group questionnaire method.

A validity study more closely approximating a household or general population survey was conducted in a high-risk community sample of adults in Chicago (Fendrich et al., 1999a, 1999b). The study compared self-reported cocaine and heroin use with hair tests. The questionnaire was based on the NHSDA, and the neighborhoods selected were those with above-average admission rates to State-supported drug and alcohol treatment programs. A third of the neighborhoods were predominantly African-American, a third Hispanic, and a third were predominantly white. Upon conclusion of the interview, respondents were asked to provide a hair specimen for a $10 incentive. About 56 percent of the sample provided hair specimens. Of the 322 specimens, 111 (34.5 percent) were positive for cocaine, and 13 (4.0 percent) were positive for heroin. Past month or past year cocaine use was self-reported by fewer than one in five subjects who tested positive. Of the 13 with heroin positive hair tests, only 4 self-reported use in the past month or past year. The authors found that those with higher levels of cocaine metabolites in their hair were more likely to self-report cocaine use. Other correlates of self-reported lifetime use were that those with less than a high school education and whites had higher self-report rates. Subjects with positive drug tests reported lifetime use more frequently than use in the past year or past month, leading the researchers to conclude that prevalence rates based on lifetime self-reports are more accurate than rates based on self-reports of more recent use.

In a representative household survey conducted in Puerto Rico, Colon, Robles, and Sahai (2002) compared self-reported cocaine and heroin use in a confidential questionnaire with the results of hair tests. They found that few of those testing positive for cocaine self-reported using the drug in the past 3 months, and a third of those who tested positive for heroin self-reported

use. The household sample was supplemented with drug users who had been in a prior treatment study and interviewed to appear as if they had been randomly selected to participate in the household survey. Among those with a positive hair test, posttreatment respondents were more likely to self-report use than the general population. About 70 percent of those who tested positive self-reported using cocaine, and 79 percent of those who tested positive for heroin self-reported use in the past 3 months. The results of this study suggest that treated users of cocaine and heroin will report their recent drug use accurately.

A.2 Strengths and Limitations of Drug Tests

The history of validity studies is relatively brief because validation requires comparison with an accurate test method. Over the past several decades, sophisticated technology-based methods have been developed to analyze drug metabolites in body fluids or tissues. Urine is most often used for the analysis of drugs of abuse, but drugs also can be detected in hair, blood, saliva, nails, perspiration, semen, and meconium. Each biological specimen is unique and offers a somewhat different pattern of information regarding drug use over time. Also, each specimen has particular strengths and weaknesses regarding the type of information that may be obtained from drug testing. Following sample preparation steps designed for the specific matrix, the same testing methods generally can be used for the various specimen matrices.

The usefulness of a drug test resides in its ability to detect accurately the presence of the parent drug or its metabolites in a biological fluid or tissue following drug administration, using a realistic amount of specimen. Analytical factors that influence test outcomes include sensitivity (the least amount of detectable drug), specificity (how selective the test is for the drug), and accuracy. These test performance characteristics depend on the individual laboratory's procedures for sample preparation and for analysis (i.e., screening and confirmatory), including the amount of specimen required to conduct the test and the cutoff concentration used to distinguish positive and negative specimens. Specimen handling, specimen storage time, and specimen storage conditions may affect drug stability in urine. Physiological factors affect drug pharmacokinetics and the excretion of the drug and metabolites, thereby influencing the window of detection. Pharmacological considerations include the time of drug administration, drug dose, ingestion method, drug chemical structure, and urine pH, as well as individual metabolism, state of hydration, fluid balance, diet, disease states, and the individual's frequency of use (Cone, 1997; Council on Scientific Affairs, 1987). Drug test results must be interpreted with caution in light of the numerous factors that may affect test results.

A.2.1 Urinalysis

After self-report, urinalysis is the most frequently employed technique to measure drug use. Urine drug testing is widely used (e.g., in workplace programs, sports programs, judicial settings, drug treatment monitoring) and is accepted as a valid check of recent drug use.

For the majority of illicit drugs, a single use occasion may be detected in the urine for 2 to 7 days. Frequent, multiple dosing over an extended period may extend the detection period. Normally, specimens collected within 6 hours of drug use contain the highest concentrations of the parent drug and their metabolites. Drug excretion in urine occurs at an exponential rate, with the majority of the drug dose eliminated within 2 days (Cone, 1997). Typically, cocaine is

detectable for 1 to 3 days in the urine, and opiates for 1 to 3 days, depending on the opiate. Amphetamines are generally detectable for 2 to 4 days. Marijuana metabolites may be detected for up to a month in chronic users (Cone, 1997).

For tobacco use, testing for the presence of cotinine is considered the most accurate biochemical measure of exposure to cigarette smoke (Bono et al., 1994; Haufroid & Lison, 1998; Jacob et al., 1999; Jarvis et al., 1987; Murray, Connett, Lauger, & Voelker, 1993; Patrick et al., 1994). Cotinine has a relatively long half-life (average ≈ 20 hours) (Pokorski et al., 1994). The detection of cotinine in urine has been estimated to have a minimum window of 2 to 3 days (Vine et al., 1993), with a maximum window of 7 days (Jarvis et al., 1987). Cotinine also can be used to test for the extent of smoke exposure. Studies have found that cotinine concentrations in urine increased as the number of cigarettes smoked increased (Knight et al., 1996; Vine et al., 1993).

Testing methods may be classified as either screening or confirmation tests. Screening tests are used to distinguish specimens that may contain drugs or drug metabolites from those that do not. Enzyme immunoassay (EIA), specifically EMIT, is probably the most frequently used screening test for drugs of abuse in urine. Specimens with positive screening test results must be analyzed using a more specific method as the confirmation test to identify definitively the presence of a drug analyte. Gas chromatography/mass spectrometry (GC/MS) is highly sensitive and specific in identifying drug analytes and is the most commonly used confirmatory test for drugs of abuse in urine.

One innovative study split 931 urine specimens gathered in the 1998 ADAM program, and each was analyzed by a different laboratory using EIA (Riley et al., 2000). There was 99 percent agreement between the two laboratories using a 10-drug screen. The largest numbers of discordant cases were for marijuana (20 cases) and cocaine (11 cases). This demonstrated good reliability across two laboratories that utilize EIA.

A.2.2 Hair Analysis

Hair testing for drugs of abuse is currently used in various settings, including workplace testing by some private-sector employers and criminal justice settings. The application of hair testing as a check on the accuracy of self-reported drug use is not well-documented in the literature. One of the early studies comparing hair, urine, and self-report was conducted by Mieczkowski and colleagues in Florida using a prototype of the DUF study (Mieczkowski et al., 1991b). Hair was analyzed by radioimmunoassay (RIA). Both EMIT and fluorescence polarization immunoassay (FPIA) were used to test urine. The authors concluded that about 4 times as many arrestees had a positive hair test as self-reported cocaine use within the previous 30 days. There was a ninefold increase in the number with opiate-positive hair tests as compared with self-reported opiate use in the previous 30 days. This study concluded that hair analysis discloses more past 30 day drug users than can be found through either urinalysis or self-reports. The study found that individuals were less likely to accurately report use in the immediate past (48 hours) and more likely to report use over longer time periods (30 and 60 days). It also was concluded that self-report was least reliable for cocaine.

Another early study involved a pretrial diversionary program in New Orleans and assessed hair testing, in addition to urinalysis and self-report, to monitor compliance with program rules of abstinence from illicit drugs (Mieczkowski, Mumm, & Connick, 1995).

After hair has undergone sample preparation steps, the same analytical methods commonly used for initial and confirmatory urine tests (i.e., immunoassay and GC/MS, respectively) can be used for most analytes for drugs of abuse in hair. Urine and hair tests often are considered complementary for identifying drug use. Although urine testing usually is limited to detecting drug use that has occurred within the past few days, hair has the potential to measure drug use over a longer timeframe—generally, months, depending on hair length. Hair specimens commonly are taken from the posterior vertex region of the scalp, where hair grows at a rate of about 1.3 centimeters (cm) per month. If hair can provide a chronological record of drug use, as some research suggests, a 1.3 cm segment measured from the proximal end should provide a record of drug use over the past month. Hair that is 7.8 cm long could provide information about an individual's drug use over the past 6 months.

However, that window of detection may correspond with a time period that is not covered by the hair testing. DuPont and Baumgartner (1995) proposed that, because hair does not grow out of the follicle sufficiently to be clipped for hair testing for about 6 to 7 days, urine and hair are complementary. However, drugs have been found in the hair within hours of administration (Henderson et al., 1996).

Some research suggests that hair test results can be quantified into heavy, moderate, and light use of particular drugs over the detection period (DuPont & Baumgartner, 1995). However, this issue is still debated (Musshoff & Madea, 2006). It has been established that the methods used for hair analysis in the Validity Study (a preliminary immunoassay screen followed by GC/MS confirmation) can measure drug analytes in hair.

The National Institute of Standards and Technology (NIST) conducted an international study in which 12 laboratories were sent eight samples of hair that were either drug-free or contained cocaine or morphine (Welch & Sniegoski, 1995). They used specimens from drug users, drug-free hair, and samples that had been soaked in drug solutions. Most laboratories reliably detected the presence of drugs or their metabolites in hair. Most of the positives missed had drug concentrations of less than 1 nanogram per milligram (ng/mg) of hair. The incidence of false positives was low. Another round of 14 laboratories found similar results, although 5 of the 14 laboratories reported a drug metabolite present when it was not (false positive).

Since April 2000, the Substance Abuse and Mental Health Services Administration (SAMHSA) has assessed the hair-testing capabilities in the United States through a voluntary pilot performance testing (PT) program (Division of Workplace Programs [DWP], SAMHSA, 2006). Laboratories have demonstrated their ability to analyze specimens from drug users, drug-free samples, and hair treated to contain drugs or drug metabolites.

Both the NIST study and the SAMHSA pilot PT program found that limitations of hair testing include the sensitivity of the individual laboratories' procedures and the ability to determine accurately drug analyte concentrations in the sub-nanogram per milligram range.

Within the SAMHSA pilot PT program, no laboratories have reported the presence of a drug when it was not present (false positive).

Although the technology of hair testing has progressed since the time of the Validity Study, several controversial aspects remain unresolved, as discussed below.

Currently, it is unclear how drugs enter the hair. Some drugs are able to bind readily to hair, but binding depends on several variables, such as pH of the hair at the time of exposure, ionic strength, hair type and pigmentation, and melanin affinity and lipophilicity of the drug (Gygi et al., 1996; Joseph, Hold, Wilkins, Rollins, & Cone, 1999; Kidwell & Blank, 1994; Kronstrand et al., 1999; Nakahara et al, 1995; Potsch et al., 1997; Rollins et al., 2003; Scheidweiler et al., 2005; Stout et al., 2006; Wilkins et al., 1995).

This uncertainty creates concern about the source of drug detected by a hair test. The question that has yet to be answered definitively is whether the drug was present as a result of drug use or external contamination from dust, smoke, or vapor containing drug. Several studies have found cocaine in the hair of children, suggesting that external contamination is an important consideration (Randall, 1992; Rosenberg et al., 1995; Smith et al., 1994). There is conflicting information from studies on the efficacy of decontamination procedures in removing externally deposited drug (Romano et al., 2003; Welch et al., 1993). Although some contend that extensive washing procedures can remove drug present from external contamination (Cairns et al., 2004; Schaffer et al., 2002, 2005), others have demonstrated that externally deposited drug remains even after extensive washing (Romano et al., 2001; Stout et al., 2006; Wang & Cone, 1995).

Further investigation is needed to determine whether criteria (e.g., analyte cutoff, metabolite-to-parent drug ratio) could be established to effectively distinguish a drug user from an individual who has been exposed to drug residue in the environment. Differentiating an externally deposited drug from a consumed drug is not definitive in hair tests for drugs in which the analyte is the parent drug (e.g., cocaine, amphetamine, methamphetamine, morphine, codeine) or a compound for which there are nonmetabolic pathways for their formation (e.g., cocaine metabolites: benzoylecgonine, cocaethylene, and norcocaine). This is not an issue in marijuana testing because the analyte detected in the confirmatory test (carboxy-THC) is a true metabolite and will only be present in the hair following marijuana ingestion. However, carboxy-THC has a low rate of incorporation into hair (Nakahara et al., 1995). Therefore, it is necessary to use a test method more sensitive than GC/MS (e.g., GC/MS/MS, GC/GC/MS, LC/MS/MS) to detect this analyte at the concentrations found in hair.

Another controversial issue is the interpretation of dose and time relationships. Some research has suggested that the amount of drugs in the hair is proportional to the amount of use (DuPont & Baumgartner, 1995). Studies with labeled cocaine have found only a limited dose and time relationship (Cone, 1994a; Henderson et al., 1993, 1996; Kidwell & Blank, 1995), although heavier drug users tend to have higher concentrations in their hair (Henderson et al., 1996; Kidwell & Blank, 1995; Mieczkowski et al., 1991b; Welch & Sniegoski, 1995).

Some studies support that segmental hair analysis can provide a chronological record of drug use (Beumer et al., 2001; Kronstrand et al., 2002; Pichini et al., 2006). However, others found high variability in segmental analysis results (Charles et al., 2003; Clauwaert et al., 2000;

Wilkins et al., 1999). Researchers should use caution when interpreting drug test results from hair segments.

One study used deuterium-labeled cocaine that could be detected separately from street cocaine use (Henderson et al., 1996). Different doses and dosing regimens of cocaine were given to 25 moderate cocaine users for up to 10 months. The study found, after a single dose of cocaine, that the drug was not always confined to a discrete area adjacent to the root. In some subjects, the drug was distributed over multiple segments extending far from the root. There was also considerable variability in the amount of drug incorporated into hair and the time until the drug first appeared in hair. This study also found that the primary analyte for cocaine in hair was the parent drug, cocaine. Only 20 percent of the subjects had BZE (benzoylecgonine) and then in only a few of the hair specimens from these individuals. It usually was found in those who received higher doses of cocaine. No drug could be detected in the hair of subjects who received the lower doses of cocaine, even when two small doses were given a week apart. There were no differences in detection based on route of administration. Typically, the maximum amount of drug was found in specimens obtained 1 to 2 months after cocaine administration, although this time varied considerably between subjects. Larger doses could be detected for up to 8 months. There was considerable variability in the amount of drug incorporated into hair, the time until the drug first appeared in hair, and the distribution of the drug along the hair shaft over time.

An important issue is that there is also evidence of bias in hair testing, with coarse dark hair retaining more of some drugs than other hair types (Cone, 1994b; Henderson et al., 1993, 1995; Holl et al., 1998; Kidwell & Blank, 1994, 1995). Kidwell and Blank (1995) found that black Asian or African hair incorporated higher concentrations of drugs than Caucasian hair—whether it was black, brown, or blonde. Researchers have shown that basic drugs may have higher concentrations in dark hair compared with lighter hair (Borges et al., 2001; Gygi et al., 1996; Kronstrand et al., 1999; Potsch et al., 1997; Rollins et al., 2003; Scheidweiler et al., 2005; Wilkins et al., 1995). However, the limited number of population studies in the literature do not indicate a significant association between hair color and drug analyte (Hoffman, 1999; Kelly et al., 2000; Mieczkowski & Newel, 2000).

Hair treatments can alter drug concentrations in hair. Bleaching, dyeing, and permanent waves have been shown to decrease the concentration of drugs in hair (Harkey & Henderson, 1989; Henderson et al., 1993; Jurado et al., 1997; Skopp et al., 1997), although the percentage decrease is similar regardless of hair type (Kidwell & Blank, 1995). Skopp et al. (1997) also found that the amount of drugs incorporated into the hair via blood or sweat varied as a function of perming the hair.

Despite these limitations, hair is being used to detect drug use. Its characteristics, including relatively noninvasive collection and ease of storage, make hair an attractive option for many drug testing programs. However, researchers relying on hair drug test results as an indicator of recent drug use in epidemiological studies must be cognizant of the limitations discussed above.

Appendix B:
Methodological Design of the Validity Study

B.1 Study Partners

The Validity Study is the result of a unique collaboration among Federal agencies, a research institution, and a university research center. The fieldwork was conducted by RTI International (a trade name of Research Triangle Institute) under contract to the Office of Applied Studies (OAS) within the Substance Abuse and Mental Health Services Administration (SAMHSA). OAS is responsible for directing the National Survey on Drug Use and Health (NSDUH),[8] and they managed and funded the fieldwork for the Validity Study, which was conducted as a methodological substudy of the 2000 and 2001 NHSDAs. The study was originated by the researchers at the University of Delaware (UD), and the planning, specimen testing, and analysis parts of the project were funded by a grant from the National Institute on Drug Abuse (NIDA). The respective organizations have worked collaboratively to support and administer this study.

To fund the specimen collection and testing, UD became a subcontractor to RTI with respect to their contract with SAMHSA/OAS to conduct the 2000 and 2001 methodological field test(s) on the NHSDA. RTI was responsible for all data collection activities for the NHSDA and the Validity Study, with the exception of those specified below as UD responsibilities. RTI drew the sample and conducted the sampling and listing field activities. They trained the interviewers and provided the computers, programmed all the instruments (screener, survey questionnaire), and subsequently conducted the interviews and collected the urine and hair specimens.

RTI provided raw data files to the researchers at UD at quarterly intervals for merging with biological specimen data. Most important, at the end of the field period, RTI cleaned, edited, and weighted the data in the same manner as was done for the NHSDA. UD researchers added the biological data results to this data tape and returned copies to RTI and SAMHSA/OAS. UD researchers, as a subcontractor to RTI, are covered by existing confidentiality agreements on the NHSDA and are bound to treat all data as confidential.

UD researchers developed the abbreviated version of the NHSDA questionnaire with input from both OAS and RTI. The core section of the NHSDA on drug use was left unchanged. Other sections of the questionnaire were eliminated, and some questions were dropped, so that questions specific to the Validity Study could be added without increasing the overall length of survey administration. UD researchers prepared the protocols for specimen collection and trained those who trained the interviewers in these protocols. RTI interviewers collected the specimens from respondents and mailed them in the postage-paid mailers provided by UD to the U.S. Drug Testing Laboratory (USDTL) for analysis. UD selected the laboratory and paid for the laboratory testing (analyzing urine and hair for the presence of marijuana, amphetamines, cocaine, and opiate metabolites or analogues, and analyzing urine for the presence of cotinine—the principal

[8] Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

metabolite of nicotine). UD was responsible for data analyses and developing reports and manuscripts, with publications support and technical review from RTI.

B.2 Sample Design

The design of the Validity Study limited the sample to a subset of the 2000-2001 NHSDA sample. The NHSDA used an independent, multistage area probability sample design for each of the 50 States and the District of Columbia. The NHSDA multistage sampling process began with the systematic selection of predefined regions and ended with the selection of one or two eligible participants per household. This design provided a sample large enough to support direct State estimates of drug use for the eight States with the largest population. For the remaining States and the District of Columbia, smaller samples were selected to support State estimates using small area estimation (SAE) techniques.

The Validity Study was not designed to report State estimates, but rather estimates among the general population living in the coterminous United States (i.e., excluding Alaska and Hawaii). To obtain the required precision, the Validity Study sample size per year for the 2 years of data collection consisted of approximately 2,000 persons with completed interviews and at least one biological specimen. Respondents between the ages of 12 and 25 years old were divided into two age groups (12 to 17 and 18 to 25). Within each age group, there were about 1,000 cases per year. The respondent universe for the Validity Study mirrored the NHSDA in each of the 2000 and 2001 survey years. It consisted of the civilian, noninstitutionalized population, including noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes) and civilians residing on military bases. Persons excluded from this universe were those with no fixed household address (e.g., homeless transients not in shelters) and residents of institutional group quarters, such as jails and hospitals.

The biggest difference between the sampling design of the NHSDA and the Validity Study is that the Validity Study limited the respondent universe to persons aged 12 to 25. With the redesign of the NHSDA in 1999, the two age groups (12 to 17 and 18 to 25) were oversampled by allocating a third of the NHSDA sample in each of these age groups. The Validity Study focused on the 18- to 25-year-old group because they have the highest prevalence of recent drug use, as identified by repeated administrations of the NHSDA. The 12- to 17-year-old age group also was targeted because prior methodological research on the NHSDA consistently demonstrated that they appear to have the highest rates of inconsistent reporting (Harrison, 2001). These age restrictions allowed us to increase precision in the estimates by examining age groups more likely than older adults to be current drug users.

Optimal allocation procedures were developed to address the issues of precision and data collection costs. For cost-efficiency purposes, the Validity Study data collection was overlaid with the regular 2000 and 2001 NHSDA data collection using a subset of the NHSDA interviewing staff. Four steps were included in this process. First, it was necessary to determine the optimal number of interviews needed to satisfy precision requirements for the Validity Study's selected drug outcome measures. Next, it was necessary to determine the number of field interviewer (FI) regions required for obtaining the optimal number of interviews. The third step involved the determination of the optimal number of selected dwelling units from which potential respondents would be selected. The last step involved respondent selection. Allocation

procedures were developed to ensure that the appropriate numbers for each age group (12 to 17 and 18 to 25) would be achieved.

Certain criteria were employed specifically for selecting the Validity Study sample. Based on analysis from the 1999 NHSDA, segments with an above-average rate of marijuana use were selected with higher probability. Moreover, the Validity Study allowed for the selection of only one person per dwelling unit, whereas the NHSDA allowed for the selection of one or two persons. The within-segment sampling procedures were modeled after the procedures normally used in NHSDA sample selection except that the target population was restricted to persons aged 12 to 25 years old. However, all final estimates from the Validity Study were weighted to account for selection bias. It was necessary to screen approximately 14,000 households per year to obtain the requisite sample size. More details on sample selection are included in Appendix C.

B.3 Questionnaire

The Validity Study was administered via computer-assisted interviewing (CAI) methods. In the NHSDA, sensitive and personal questions were administered via audio computer-assisted self-interviewing (ACASI). This included all the drug-related questions. The Validity Study used the same method of administration. In the NHSDA, less sensitive questionnaire content was administered using computer-assisted personal interviewing (CAPI), and the Validity Study did so as well.

CAI/ACASI technology affords a number of improvements in survey data collection. First, this methodology permits more complex routings in the questionnaire compared with a paper-and-pencil interviewing (PAPI) instrument. The computer is programmed to implement complex skip patterns and fill-in specific wording based on answers previously provided by the respondent. Errors made by interviewers (and respondents) due to faulty implementation of skip instructions are virtually eliminated. The computer is programmed to identify inconsistent responses and resolve them through respondent prompts and queries. This reduces the need for most manual and machine editing and saves both time and money. Additionally, asking the respondent to resolve inconsistencies results in data that are more accurate than when inconsistencies are resolved using machine-editing rules.

CAI/ACASI technology permits greater expediency with respect to data processing and analysis. For example, a number of data-processing steps, including editing, coding, and data entry, become part of the data collection process. Data are transmitted via modem. These efficiencies save time due to the speed of data transmission, as well as receipt in a format suitable for analysis. Tasks formerly completed by clerical staff are accomplished by the CAI/ACASI program. In addition, the costs of printing paper questionnaires and the associated mailing of completed questionnaires are eliminated.

The Validity Study used Apple Newton handheld computers to conduct household screening interviews, just as it was done in the NHSDA. The primary advantage of this methodology was improved accuracy in selecting the correct household member for an interview. The complex paper-and-pencil procedure formerly used was prone to occasional human error. However, the handheld computer automatically selected the correct household

member based on the demographic variables entered, thus substantially reducing the probability for human error.

B.3.1 Questionnaire Content

The 2000-2001 NHSDA questionnaire and interview methods were designed to retain respondent interest, ensure confidentiality, and maximize the validity of responses. The questionnaire was administered in such a way that interviewers did not know respondents' answers to the sensitive questions, including those on drug use. These questions were self-administered, and respondents entered their responses directly into the computer using ACASI procedures. The only topical areas that were interviewer administered (e.g., CAPI) were sections on demographic information, income, and hair characteristics.

A revised version of the 2000 NHSDA questionnaire was used for the Validity Study. The questionnaire was updated in 2001 to correspond with changes made to the NHSDA questionnaire. The questionnaire focused on self-reported drug use to permit the addition of some questions and the collection of the biological specimens. The average amount of time required to administer the base questionnaire for the Validity Study was approximately 45 minutes. The process of explaining the Validity Study component, obtaining additional consent, asking questions, and collecting the biological specimens took about 20 minutes per respondent. Therefore, overall administration time for the Validity Study took about 5 minutes longer, on average, than the 1 hour response time for the NHSDA.

The NHSDA questionnaire was divided into sections based on specific substances or other main topics. The same sequence of questions was presented for each substance or substance class. These questions constitute the core section of the NHSDA and allow estimation of drug use prevalence and patterns. The core section of the questionnaire was identical in the Validity Study and NHSDA. Following the questionnaire's core section, the Validity Study continued to follow the ordering of questions presented in the NHSDA with the exception that portions of the NHSDA questionnaire were deleted. In some cases, whole modules were deleted.[9] Parts of other sections were deleted.[10] Several replacement questions were added to allow personal and family income to be measured in the Validity Study but not with the precision measured in the NHSDA. The deletions were selected carefully to streamline the questionnaire in terms of reducing administration time to allow for the collection of urine and hair specimens. The questions deleted were considered less critical to the purposes of the Validity Study. For example, there was little reason to believe the validity of self-reported drug use varied with health insurance status.

The organization of the Validity Study questionnaire is shown in Table B.1, which lists the modules included in the questionnaire and the mode of administration of each.

[9] Deleted modules included the social environment, health insurance, marijuana market, and youth and parenting experiences sections.

[10] Sections with partial deletions included the noncore demographics, treatment, personal and family income, and special topics sections.

Table B.1 Organization of the 2001 NHSDA Validity Study Questionnaire

Module	Mode of Administration
Core Demographics	CAPI
Calendar (30-Day and 12-Month Reference Dates)	FI reads instructions and identifies reference dates.
Beginning ACASI Section	FI reads description of keyboard and helps respondent adjust headphones.
Computer Tutorial	Respondent completes computer practice session with FI's help.
Tobacco, Alcohol	ACASI
Marijuana, Cocaine, Crack	ACASI
Heroin, Hallucinogens, Inhalants	ACASI
Pain Relievers	ACASI
Tranquilizers	ACASI
Stimulants	ACASI
Sedatives	ACASI
Specialty Cigarettes	ACASI
Validity Study Follow-Up (3-Day and 180-Day Use)	***ACASI***
Tobacco, Marijuana	***ACASI***
Cocaine, Heroin	***ACASI***
Methamphetamine, Prescription Diet Pills	***ACASI***
Specialty Cigarettes	***ACASI***
Friends Use/What Parents Think about Drugs	***ACASI***
Special Drugs	ACASI
Risk/Availability	ACASI
Drug Treatment	ACASI
Health (Pregnancy status), Religion	ACASI
End ACASI	Respondent returns computer to FI.
Back-End Demographics:	CAPI
Education	CAPI
Employment	CAPI
Household Roster	CAPI
Resume ACASI	Turn computer back over to respondent.
Begin Validity Study Persuasion Experiment (½ Get Appeal / ½ Do Not Get Appeal)	***ACASI***
Validity Study Debriefing Questions	***ACASI***
Validity Study Repeat Questions (3, 7, 30, and 180-Day Use)	***ACASI***
Cigarettes, Cigars	***ACASI***
Marijuana, Cocaine, Heroin	***ACASI***
Pain Relievers	***ACASI***
Methamphetamine, Prescription Diet Pills	***ACASI***
Specialty Cigarettes	***ACASI***
Over-the-Counter Drugs (Licit Use)	***ACASI***
Pain Relievers (Licit Use)	***ACASI***
Prescription Diet Pills (Licit), Marinol (Licit)	***ACASI***
Exposure to Smoke from Drugs	***ACASI***
End Validity Study Persuasion Experiment	***Respondent returns computer to FI.***
Personal Income	CAPI (Proxy Allowed)
Employment (Other Family Members)	CAPI (Proxy Allowed)
Family Income	CAPI (Proxy Allowed)
Validity Study – Consent	***CAPI (FI obtains consent to collect hair and urine).***
Questions on Hair Type, Chemicals Used, etc.	***CAPI***
Take Hair and Urine Specimens	***FI obtains hair and urine specimens from respondent.***
Verification	FI and respondent complete form.
FI Observation Questions	FI records own responses.

ACASI = audio computer-assisted self-interviewing; CAPI = computer-assisted personal interviewing; FI = field interviewer.

B.3.2 Core Questions

The core portion of the NHSDA was designed to be administered consistently from year to year. The core of the 2000-2001 NHSDA was comprised of the initial demographic questions and the modules of drug questions from tobacco through sedatives. The introductory demographic questions included age, gender, race, ethnicity, marital status, educational attainment, and overall health status. Personal and household income and health insurance were also core questions in the NHSDA, although health insurance questions were omitted in the Validity Study. The CAI questionnaire began with the demographic questions administered by the interviewer. The interviewer then turned the computer over to the respondent, instructing him or her to complete the next section using the computer and headphones. The interviewer introduced the respondent to the keyboard, but the ACASI program began with a tutorial to the "RTI self-interviewing system."

The very first questions after the tutorial were about tobacco use. The questionnaire began with, *"Have you ever smoked part or all of a cigarette?"* The next question asked for the age of first use, although youths aged 12 to 17 years who had never smoked were asked about their intention to smoke cigarettes. The next question asked: *"Now think about the past 30 days, from [DATEFILL] up to and including today. During the past 30 days, have you smoked part or all of a cigarette?"* DATEFILL was filled in by the computer with the month and day of the reference period—in this case, 30 days ago. If the respondent used the drug in the past 30 days, he or she was asked, *"During the past 30 days, that is since [DATEFILL], on how many days did you smoke part of all of a cigarette?"* The respondent was instructed to fill in a number between 1 and 30. If the respondent reported no use in the past 30 days, the next question was about the recency of use: *"How long has it been since you last smoked part or all of a cigarette?"* Response categories were: *"More than 30 days ago but within the past 12 months; more than 12 months ago but within the past 3 years; and more than 3 years ago."* Those reporting use in the past year also were asked to indicate on how many days they used by filling in a number between 1 and 365.

This pattern of questioning was repeated for all the drugs included in the NHSDA. The general form was to ask first about lifetime use, then age at first use, followed by use in the past 30 days and in the past year. If no lifetime use was reported, the ACASI skipped to the next drug class. If the respondent used the drug in the past year, the number of days used was asked. If the respondent used the drug in the past 30 days, the number of days used also was asked. This general pattern of questioning was followed for alcohol, marijuana/hashish, cocaine, crack cocaine, heroin, hallucinogens, inhalants, pain relievers (analgesics), tranquilizers, stimulants, and sedatives. For stimulants, separate questions were asked about methamphetamine (i.e., illicit methamphetamine and the prescription methamphetamine drugs Desoxyn® and Methedrine®), prescription diet pills (i.e., amphetamine, Benzedrine®, Biphetamine®, Fastin®, or phentermine), methylphenidate (Ritalin®), and a list of some prescription stimulants (i.e., drug names or trade names). These were followed by questions asking about the nonmedical use of any other prescription stimulant. For pain relievers, separate questions were asked about some specific prescription drugs by trade name, then respondents were asked about the nonmedical use of selected prescription pain relievers (i.e., from a list of drug names or trade names). These were followed by questions about the nonmedical use of any other prescription pain reliever. There were many more questions about cigarettes and alcohol than most of the illicit drugs. Brand

preference of tobacco products also was asked about, and alcohol questions asked about more frequent usage patterns. The tobacco section also included questions on the use of smokeless tobacco (snuff and chewing tobacco), cigars, and pipes.

B.3.3 Follow-Up Questions

Follow-up questions were specific to the Validity Study and did not occur in the NHSDA. Recall that the NHSDA asked about drug use in specific time periods. The DATEFILL function was defined based on the day the ACASI questionnaire was completed and was designed to print the date in the question stem that occurred 12 months ago or 30 days ago, depending on the timeframe referent. The most recent time period asked in the NHSDA was the past 30 days. However, the biological tests (urine and hair) had timeframes that varied by drug type. Only the detection times for drugs in hair and, perhaps, marijuana metabolites in the urine of chronic users overlapped with the 30-day time period. To address this, follow-up questions were added that more closely correspond to the time periods matching the window of detection for most drugs in urine, which is about 3 days. Those Validity Study respondents reporting use of tobacco products, marijuana, cocaine, or opiates were queried at the end of the core section on whether that use occurred within the past 3 days.[11] The prototype question was: *"Earlier, you said you smoked [part or all of a cigarette] in the past 30 days. Did you last smoke [all or part of a cigarette] within the past 3 days, that is, since [DATEFILL]?"* For these questions, DATEFILL was computed by ACASI to reflect the date 3 days ago, including the month and day.

The ability to program the Validity Study using CAI methods to bring up these questions at the end of the core section allowed for the core questions to remain in their entirety. Then, invisibly, questions were introduced about recency of use corresponding more closely with the window of detection for the drug tests. For respondents reporting their last drug use occurring in the past year, the CAI was programmed to bring up the question: *"Earlier, you said you had used [cocaine] in the past 12 months. Did you last use cocaine within the past 6 months, that is, since [DATEFILL]?"*

B.3.4 Debriefing Questions

The debriefing questions also were specific to the Validity Study and were not asked in the NHSDA. These questions occurred midway through the supplemental demographics module. The interviewer administered these demographic questions and turned the computer back to the respondent. At this point, one half of the respondents received an appeal designed to increase their willingness to respond accurately to the drug use questions. The strategy was to remind respondents of the history of the survey in measuring drug use in the United States. Respondents were informed that there was no other way to gather credible information on drug use, although scientists and policymakers always had concerns about people's willingness to respond honestly. The overall strategy was to appeal to "science" and to people's willingness to disclose the correct information to inform science and policy. Research suggests that this improves the validity of self-reports (Charles Cannell, personal communication, May 1991). By embedding an experiment in the study, randomizing half to the appeal and half to a short introduction, we

[11] The follow-up questions also included amphetamines in 2001, but the questions mistakenly were omitted from the 2000 questionnaire.

elicited self-reports both under normal survey conditions and under the enhanced debriefing condition. The text of the appeal is shown in the following boxed text:

> *For over 25 years, the study in which you are participating has been the major source of information on the nature and extent of drug use in the United States. You are one of more than 5,000 individuals who have been selected at random to represent the American people. No cities, neighborhoods, or individuals are targeted any more than others. Because we use scientific sampling procedures to select a random sample, your experience and views represent those of 21,000 people.*
>
> *One of the big concerns that scientists, health professionals, and legislators have had about the results is whether people will tell the truth about their drug use. This is a critical issue, since there is no other way to get this information other than surveying a large and representative sample of the population.*
>
> *Scientists already know there are many misconceptions about drugs. Misconceptions will continue and policies addressing drug use will be based on flawed information if this study's results are wrong. Therefore, the scientists who designed this study have worked hard to make you feel comfortable giving us your honest answers. Remember, you were never asked to sign your name on anything. You were assigned a code number, and all your responses are tied to that code number. The computer interview is designed to give you the utmost privacy. The interviewer never sees your answers to drug questions, and no one in your household even knows the questions you are being asked.*
>
> *Perhaps more importantly, this study has a Federal Confidentiality Certificate. This authorizes us to protect the privacy of respondents by withholding their names and other identifying characteristics from all persons not connected with this study. Only the researchers working on this study will ever see your individual responses.*
>
> *Finally, the team of researchers working on this study will not evaluate your individual answers. Rather, they are interested in the overall pattern of answers from everyone. The researchers hope their efforts to make you feel comfortable giving us your truthful answers are effective.*
>
> *Now, there will be a few questions about what you think about our study. These questions will be followed by several more questions about your drug use. Please answer these questions as honestly as you can.*

Those not receiving the appeal received a brief introduction to the next section of questions. Following that, all respondents received the debriefing questions and repeat questions. Those randomized to not get the appeal saw the following text on the ACASI screen:

> *Now, there will be a few questions about what you think about our study. These questions will be followed by a brief repeat of some of the questions about drug use. Please answer these questions as honestly as you can.*

All respondents completed the debriefing questions, which asked how the respondent thought that most people would respond to the study. Questions asked how much difficulty the respondent thought that most people would have understanding the drug-related questions and remembering the types of drug-related information the study requests. Respondents were asked

whether they thought most people would be concerned about others having access to the answers they gave on the questionnaire. They were asked how truthful they thought most people would be and how embarrassed most people would find it to answer these types of questions. Then respondents were asked virtually the same series of questions with respect to their own behavior (i.e., *"How much difficulty did 'you' have understanding the drug-related questions and remembering the types of drug-related information?"*). Respondents were asked to rate the accuracy of their answers and their concern about others having access to their answers. They also were asked to rate their level of embarrassment in responding to the questions and their own level of truthfulness in answering the drug-related questions.

B.3.5 Repeat Questions

The repeat questions also were specific to the Validity Study. Following the debriefing questions, respondents were instructed, *"These next questions ask how recently–if ever–you have used cigarettes and certain other drugs."* All respondents were asked the repeat questions, whether they received the appeal or the short introduction. The repeat questions followed a standard pattern and asked about the drugs that were tested: tobacco (with separate questions for cigarettes and cigars), marijuana/hashish, cocaine (including crack), stimulants (including amphetamines), heroin, and prescription pain relievers (including opiates).

The repeat questions module included questions that were intended to focus on the use of drugs that could cause a positive test. Respondents were asked about use of over-the-counter (OTC) or prescribed nicotine products to help quit smoking (e.g., nicotine gum, patches). These products contain nicotine and may produce a urine test positive for cotinine. Questions were asked about use of prescribed tetrahydrocannabinol (i.e., dronabinol, Marinol®), which may cause a positive marijuana test. Respondents also were asked whether they had taken prescribed pain relievers because the opiate analytes tested (i.e., codeine and morphine) fall within this broad drug category. However, there were some problems with the format of the questions concerning prescription pain relievers and stimulants in light of the drug test analytes selected for the study. The repeat questions concerning prescription pain relievers would not elicit information relevant to the use of many drugs that could cause a positive opiates test and included some drugs that would not cause a positive opiates test. Repeat questions concerning stimulants would not elicit information relevant to the use of many drugs that could cause a positive amphetamines test. Questions were asked about the *nonmedical* use of methamphetamine (i.e., illicit methamphetamine and two prescription drugs, Desoxyn® and Methedrine®), but did not ask about the *legitimate* use of drugs that could result in a positive amphetamines test (i.e., prescription amphetamine or methamphetamine, prescription drugs that metabolize to methamphetamine and/or amphetamine). The only questions concerning prescribed stimulants focused on prescription diet pills, only some of which contain amphetamine or metabolize to amphetamines (e.g., Didrex®).

The first question in the series of repeat questions measured lifetime use with a "yes" or "no" response. If the answer was "yes," ACASI asked how long it had been since the respondent used the drug. The response categories were: "*Within the past 6 months–that is, since [DATEFILL]; more than 6 months ago but within the past 12 months–that is, since [DATEFILL]; and more than 12 months ago."* For those reporting use within the past 6 months, the next question asked: "*Thinking about the past 6 months, when did you last* [use drug]?" The

response categories were: "*Within the past 3 days–that is, since [DATEFILL]; more than 3 days ago but within the past 7 days–that is, since [DATEFILL]; more than 7 days ago but within the past 30 days–that is, since [DATEFILL]; and more than 30 days ago, but within the past 6 months–that is, since [DATEFILL]."* The DATEFILL function wrote the day and month of the reference period in the question. The repeat questions contained timeframes more in line with the window of detection for the urine tests.

A concern about the validity of biological specimen testing is the potential of passive exposure to cause a positive drug test. For example, people who are near someone who is smoking a cigarette could be passively exposed to nicotine through environmental tobacco smoke. The danger appears greatest for the smoked drugs. Questions in the repeat section asked respondents to rate how often (i.e., daily, frequently but not daily, seldom, never) they had been around people smoking cigarettes or any other tobacco product, marijuana/hashish, cocaine or crack, heroin, or methamphetamine.

After completing these question modules, respondents were instructed to return the computer to the interviewer. The interviewer continued with the demographic questions to ascertain personal income. Respondents then were asked to list their age and how they were related to other household members. This preceded questions ascertaining the total household income, and respondents were given the opportunity to invite another household member, who might be better able to answer these questions, to do so. For those under age 18, the expectation was that the parent would be invited to answer these questions. The survey procedures were designed so that a parent would be available to provide parental consent for those under age 18 to provide the biological specimens. After the income questions, respondents were invited to participate in the biological specimen collection. Parents had to give their consent for youths under age 18, then the youths had to give assent as well. Generally, interviewers reported that parents ascertained their child's interest in providing the specimens before giving their consent.

B.4 Interview Administration

B.4.1 Interviewer Assignment and Training

The 2000-2001 NHSDA field interviewers (FIs) resided throughout the country and were supervised by field supervisors (FSs), who in turn were supervised by regional supervisors (RSs). Supervisors maintained regular contact with interviewers and helped to ensure the quality of the work. The NHSDA interviewers were trained in several sites throughout the country, generally the site closest to their home. Training for the NHSDA increased from 3 full days (and an additional day for novice interviewers) in the early 1990s to 7 days in 1999 when the CAI technology was introduced. The training is well designed, and, like good surveys, was delivered consistently to the FIs by the trainers. There were extensive interviewer manuals and instructions to assist NHSDA interviewers and provide guidance on how to handle different situations. An extensive interviewer manual was similarly developed by RTI for the Validity Study. Even the training materials used by the trainers and those who trained the trainers were well developed and written to facilitate standardized training for both the NHSDA and the Validity Study. The high response rates obtained in the NHSDA are a testament to the thoroughness of the training and the level of supervision the interviewers receive.

All the Validity Study interviewers were trained NHSDA interviewers. Interviewers must have successfully completed one quarter of data collection on the main study before being considered for a Validity Study assignment. Initially, volunteers were recruited due to concern that interviewers would not be comfortable collecting the urine and hair specimens. However, the enthusiasm for the Validity Study was underestimated, and many of the NHSDA interviewers volunteered. An incentive for the interviewers was a $25 bonus for each completed Validity Study interview.

At training, supplies necessary to complete the CAI and collection of biological specimens were issued. These included the ACASI instrument, manuals and visual aids, lead letters, consent forms, hair and urine collection kits, incentive payments, and mailing materials. At the onset of year 2, while inexperienced interviewers completed the training, seasoned interviewers were required to complete "veteran home study." The training took 7 days for the NHSDA, with 2 additional days required for the Validity Study. NHSDA interviewers were educated thoroughly in methods for maximizing a respondent's understanding of the government's commitment to confidentiality and in methods of converting refusals. They were taught to attempt to conduct the interview in a setting in the respondent's home that was as private as possible, particularly when the respondent was a youth. They were taught the Validity Study procedures for data and specimen collection through lecture, videos, and instructor demonstrations. An important component of the training was for each interviewer to practice taking a hair specimen from one of the other interviewers. This provided the interviewer with the knowledge of what a respondent experiences during the hair collection.

Many of the interviewer tasks remained the same in both the Validity Study and the NHSDA. Contacting and screening was the first component of the interview process. Interviewers sent lead letters to selected households announcing they would be stopping by in person within a few days. The lead letter was modified only slightly in the Validity Study from the NHSDA. Upon arrival at the address, the interviewer referred the respondent to the letter and answered any questions. If the respondent had no knowledge of the lead letter, the interviewer provided another copy, explained that one was previously sent, and then answered questions. If no one was at home during the initial call at the address, the interviewer left a "Sorry I Missed You" card indicating the interviewer would make another callback at a later date and time. Callbacks were made as soon as possible based on each interviewer's schedule. Interviewers made at least four callbacks (in addition to the initial call) to each address to complete the screening process and possibly obtain an interview. As necessary and appropriate, the interviewer used an appointment card for scheduled return visits with the respondent.

The screening respondent was at least 18 years old. When an in-person contact was made and introductory procedures completed, the interviewer presented a "statement of confidentiality" to assure confidentiality of information. Assuming respondent cooperation, the screening and rostering of the household was initiated through administration of the housing unit screening questions for housing units or the group quarters unit screening questions for group quarters units. The screening questions were administered via a Newton handheld, pen-based computer, which also did the subsequent sample selection routines.

If a potential respondent refused to be screened, the interviewer accepted the refusal in a positive manner, thereby avoiding the possibility of creating an adversarial relationship and

precluding future opportunities for conversion. A conversion usually was attempted by supervisory field staff or specially selected veteran interviewers with established conversion records. A question-and-answer brochure providing answers to commonly asked questions also was given to the respondent at this time. In addition, interviewers carried copies of NHSDA reports and NHSDA newspaper clippings that could be left with the respondent. Following the introductory exchange, the screening interview continued until completion.

If a language barrier was identified, the interviewer had several options. A family member or friend could act as an interpreter if the screening respondent made the suggestion. Also, if the interviewer was an RTI-certified bilingual interviewer, the Spanish version of the Newton program could be used. Only in the screening process was a translator or bilingual interviewer used. There was no Spanish version of the Validity Study CAI interview, although there regularly is a Spanish version of the NHSDA and certified bilingual interviewers trained to administer it.

Interviewers were instructed not to mention the Validity Study incentive so as not to affect the way a respondent answered the CAI questions. No member of the household was to have any prior knowledge of the possibility of being asked to give a hair and urine specimen if selected for an interview. Also, no one was to have any knowledge of the possibility of receiving a cash payment for their participation until the appropriate point in the interview process.

Once the household was rostered and the data verified, the respondent was selected. The interviewer used the Newton to process the special Validity Study selection criteria. Upon selection of the respondent, the interviewer began the process to obtain the informed consent for the CAI interview. If the selected respondent aged 18 or older was currently available, the interviewer immediately began administering the questionnaire in a private setting within the dwelling unit. If the selected individual was 12 to 17 years of age, parental consent for the initial interview was obtained, then the minor was asked to participate. The informed consent procedures explained the purpose of the study, confidentiality procedures, and what was required during the interview. As part of the process for obtaining informed consent, respondents were shown a "statement of confidentiality" describing all the measures taken to assure confidentiality, including the protection offered by the Federal Certificate of Confidentiality under Section 301(d) of the Public Health Service Act. This form was signed and dated by the interviewer and given to the respondent.

These procedures were slightly modified in 2001. The change was the result of the Children's Health Act of 2000 (Public Law 106-310), which included a provision providing for the protection of identifiable data collected by SAMHSA (Section 501(n) of the Public Health Service Act). The confidentiality provisions were modeled after those employed by the National Center for Health Statistics. Therefore, the NIDA certificate of confidentiality was no longer necessary. In its place, a statement of confidentiality was developed, basically spelling out the confidentiality safeguards.

After obtaining consent from the selected respondent, the interviewer administered the interview in the prescribed and uniform manner. At the completion of the CAI interview, the interviewer then explained that the study had a second objective: to determine how accurately respondents in national surveys report their illicit drug use. Interviewers were trained to read, out

loud, the script for the Validity Study informed consent exactly as written to both adults and minors. The respondents were briefed fully on the collection and use of urine and hair specimens for measuring drug use. If the respondent was a minor, the interviewer addressed the parent or guardian first. Permission to obtain biological specimens from a minor had to be obtained first from the parent or guardian before the minor was approached to give his or her consent. The interviewers were instructed to have the correct change in order to pay the respondent $25 for the urine and $25 for the hair sample. Incentives were paid after the specimens had been packaged.

As part of their training, FIs were shown a video developed by UD and RTI showing all the necessary tasks that must be conducted for the biological specimen collection. It emphasized that all procedures must be followed exactly so the specimens would be collected, packaged, and shipped uniformly. Failure to follow any of the procedures could lead to a refusal by Federal Express to ship the collected specimens or the possibility of the specimens being damaged or discarded once they reached the laboratory. Interviewers received their own copy of the video so they could review the biological specimen collection procedures.

B.4.2 Consent Forms

As noted above, the consent forms were read word for word. Interviewers told respondents they were required to read them aloud. Interviewers were trained to answer respondents' questions and concerns, including strategies to convert reluctant respondents. The fact that the procedures worked well is evident by the nearly 90 percent response rate for at least one biological specimen. The "Consent Form for the 2001 NHSDA Validity Study" is shown in the boxed text on the following page.

The same script was read to parents of minors, adapted only slightly to indicate "your child." If the parent agreed, the youth also was read the same consent forms. Youths must have provided their assent, and their parents must have provided their consent, before the youths were interviewed.

B.4.3 Urine Collection Procedures

Interviewers had the respondent collect his or her own urine specimen after the collection of the hair specimen. The respondent was given a paper cup and a plastic urine specimen bottle and was asked to use his or her own bathroom to collect his or her own urine specimen by urinating into the paper cup and pouring enough of the specimen into the plastic bottle to measure to the 30 milliliter (mL) line on the bottle. While the respondent did this, the interviewer completed the packaging of the hair specimen and entered the respondent ID number assigned for the biological specimens into the computer.

When the respondent returned the urine bottle, the interviewer checked the temperature strip on the bottle and recorded whether or not it was within the specified temperature range (90 to 100 °F). Specimens not in range may have been diluted or substituted. The urine specimen bottle then was wrapped in a paper towel and inserted into a small corrugated box. Both this box and the hair specimen envelope were inserted into a Federal Express plastic envelope, along with a requisition sheet. Biological specimens were identified using only unique respondent identification numbers and dates according to specific instructions. The respondent's name or any

Consent Form for 2001 NHSDA Validity Study

Thank you for participating in our study. This study is the major source of information on drug use in the United States. Your answers will be combined with those of others to provide information on drug use for the nation. Policymakers use the information from the study to determine the best way to deal with drug use in our society.

However, if people don't answer the questions truthfully, policies are designed that are based on incorrect information. We are sure you can understand how that may cause more harm than good. The scientists who designed this study worked hard to make you feel comfortable answering the questions we ask. However, we know that despite our efforts, some people will not be comfortable answering our questions truthfully. We are continually trying to determine how we can improve the study to encourage people to answer truthfully. So, the secondary goal of this study is to get a more accurate picture of how truthfully people respond to our questions about their drug use.

In order to do this, we need to compare the questionnaire answers with something we know is accurate. There have been advances in drug testing in recent years that make it possible to determine recent use of specific drugs fairly accurately. You may be aware that urine can be tested for the presence of drugs. However, you may not be aware that hair can also be tested for the presence of specific drugs. Now we would like you to participate in another part of our study. We would like you to provide a urine sample and a small hair sample from two locations on your scalp, each about the area of a pencil eraser. The hair will be clipped close to the scalp to minimize any cosmetic effect, and it will take about 10 minutes to conduct this part of the study. If you agree, you will receive $25 for the urine sample and $25 for the hair sample. We are offering this incentive to encourage you to participate in this critical part of our study.

Be assured that we are not interested in your individual responses. We are not even interested in how your particular responses compare with the results of tests the lab performs on your specimens. What we are interested in is getting an overall picture of how accurately individuals report their drug use. Before you answer, please allow me to read some informed consent information to you. Then I would be happy to answer any questions you have.

Your participation in this study sponsored by the U.S. Public Health Service, Substance Abuse and Mental Health Services Administration, both part of the U.S. Department of Health and Human Services, is completely voluntary. You can choose to withdraw at any time during this part of the study. Section 501 of the Public Health Service Act requires that all information you provide, including the results of lab tests on your hair and urine, must be kept in the strictest of confidence and can not be released to unauthorized persons. This means your results will not be available to you, anyone in your family, or anyone else not authorized by this federal law. I'd be happy to show you the Study Description, where this federal law is cited, if you'd like. (SHOW STUDY DESCRIPTION.) This Act authorizes the researchers conducting this study to protect the privacy of individuals who participate by withholding their names and other identifying characteristics from all persons not connected with the conduct of this study.

Your specimens will only be identified by a research identification number that will be matched to your questionnaire results. You will recall that your name is not recorded anywhere on the questionnaire. Only the researchers working on the study have access to information that links your address to your study results and that information will be destroyed at the end of the study. No names are ever kept. Scientists are not interested in the results for any particular individual, but rather for people overall.

If you think of something after I have left, you may phone _____, the NHSDA Project Representative, toll-free at _____. If you have questions related to your rights as a survey respondent, you may contact _____, the Chairman of RTI's Committee for the Protection of Human Subjects at _____, extension _____. You can also visit the project Website at http://nhsda.rti.org for more information. I will sign this form and give you a copy to show that I have read it to you and that you consent to participate in this part of the study. I will be glad to answer any questions you have about the study now.

Note: The project's URL is now https://nsduhweb.rti.org/.

other easily traceable identifiers were never noted. The interviewer dropped off the package at a Federal Express station that day or the next or arranged for home pickup of the package within the next day or two.

Neither the questionnaire data nor the biological specimens contained any respondent names. On the data file, respondents were identified only by an ID number assigned to the screening files, questionnaires/interviews, and the biological specimens. Interviewers could not access completed interviews to review or to edit questionnaire data. Data from the electronic interviews were transmitted daily via secured data transmission. Although the ID number was associated with a location number and a dwelling unit number, this location information was deleted from the data file. The address information was maintained in a separate file by RTI for use in sampling, fielding, and weighting cases and was archived at the completion of data processing. Questionnaire data were processed immediately upon receipt and added to a raw data file in preparation for internal processing. All links to any personal identifying information were removed following completion of data-processing activities.

B.4.4 Hair Collection Procedures

Interviewers explained and collected the hair specimen before the urine specimen. Once the interviewer was certain the respondent was comfortable with the collection technique, questions relating to hair texture and use of hair treatment applications were asked via CAPI. Before an interviewer began collecting specimens, procedures called for donning latex gloves, which were worn during both the hair and urine collection.

For hair collection, the interviewer asked the respondent to sit in a chair with his or her head tilted forward. Hair was usually collected from just below the crown of the head (i.e., the posterior vertex region of the head). However, if a respondent objected or if his or her hair was not long enough, the specimen could be collected from the nape or another part of the head. The hair specimen needed to be a minimum of 1 centimeter (cm) in length to be tested at the laboratory. Using a new comb, the interviewer parted the hair to divide a section in a wedge and continued to section the hair until holding a specimen about the area of a pencil eraser. The specimen was clipped as close to the scalp as possible using blunt scissors. During the first quarter of the study (January to April 2000), one specimen per respondent was collected. Thereafter, in an attempt to collect sufficient specimen, interviewers were instructed to obtain two specimens per respondent. The second specimen was taken in the same manner, and the two specimens were combined by the interviewer with proximal “root” ends aligned. The two specimens generally came from different sites in order to reduce any cosmetic effect. If the participant preferred, the hair could be sampled once, with the specimen being the area of two pencil erasers side-by-side. The interviewer noted the length of the collected hair. The root ends were wrapped in foil, and the combined specimens were sealed in an envelope for shipment to the laboratory.

The study team was concerned that the change in collection procedures could suppress response rates for hair specimens, so the situation was monitored carefully. The change did not have an appreciable impact on hair specimen response rates. The rates decreased only slightly after the change, and the percentage of respondents who provided urine increased. However, the

percentage of respondents who provided both hair and urine specimens remained nearly the same whether one or two specimens of hair were collected.

B.4.5 Incentive Payments

The amount of the incentive was based on studies conducted at UD, which typically provide respondents with a $50 incentive for both a urine and blood or saliva specimen in their treatment and human immunodeficiency virus (HIV) studies. Long-term experience with 1,720 respondents in their treatment and HIV studies from 1990-1998, using a $50 incentive, yielded a response rate of 86.3 percent for the urine specimens and 84.0 percent for the blood or saliva specimens. The UD experience indicated that respondents would be willing to provide these specimens with a $50 incentive, and this expectation was born out in the high specimen response rates in the Validity Study.

Fendrich et al. (1999b, 2004) provided respondents, in a household survey, a $10 incentive for a hair specimen. Their research suggested that hair testing was not perceived as invasive or burdensome by a majority of subjects, yet 24 percent refused to provide a specimen. Another 19 percent had too little hair to provide a hair specimen. So, hair specimens were obtained on only 57 percent of those interviewed. This led researchers to conclude that an incentive payment of $10 for a hair specimen was too small. They recommended, in general population surveys, a larger incentive would substantially reduce the rate of nonparticipation among subjects eligible to participate in the procedure. Based on the limited research literature on hair testing in a general population survey and the experience at the UD, an incentive of $25 for each specimen seemed well conceived in terms of providing an incentive for respondents to agree to provide the specimens.

B.5 Drug Testing Protocols

Hair and urine specimens were analyzed by U.S. Drug Testing Laboratory (USDTL) for the presence of the following drugs and their metabolites: marijuana, cocaine, amphetamines (amphetamine and methamphetamine), and opiates (codeine and morphine). Urine specimens also were analyzed for cotinine, the principal metabolite of nicotine.

The drugs tested are generally the most prevalent in the NHSDA, with the exception of opiates. Due to alcohol's short half-life via any testing mechanism, it was not practical to test for alcohol. Marijuana has always been the most prevalent of the illicit drugs, and recent use should be detected via urinalysis (and for chronic users up to 30 days). Cocaine is the second most frequently reported illicit drug. We wanted to have a measure of drug use in the past 3 days via urinalysis and for the past month and up to the past 6 months via hair testing. Testing for the tobacco metabolite cotinine was included because tobacco use is more prevalent than any of the illicit drugs. Further, the results from the 1994 NHSDA, which employed both a self-administered answer sheet for the tobacco questions and an interviewer-administered protocol, found significantly higher rates of tobacco use under the self- versus interviewer-administered conditions (OAS, 1996). This methodological study revealed that there is underreporting of tobacco use. Asking about tobacco use can be a sensitive question, particularly for youths for whom the purchase of tobacco products is illegal.

B.5.1 Laboratory Selection

Laboratory selection was based on consultation with Dr. Edward Cone, formerly chief of the chemistry and drug metabolism branch at NIDA. One of the foremost experts, nationally and internationally, in drug testing body fluids and tissues for drugs, he proposed the appropriate uses of hair testing technology that currently are used in epidemiological research and in research to examine the validity of self-reported drug use. He recommended the laboratory that conducted the analysis—USDTL. Two other Validity Study consultants, Dr. Fendrich and Dr. Johnson of the University of Illinois at Chicago, conducted a validity study in Chicago neighborhoods described in Appendix A that used a questionnaire modeled on the NHSDA. They also used USDTL in their validity study, providing a tacit recommendation.

There were developments and improvements in hair testing technology between the time the study was designed and funded. In year 2 of the Validity Study (2001), USDTL revised its hair testing procedures to improve the sensitivity for some analytes. The laboratory lowered the cutoff concentrations for its marijuana and opiates screening tests and for its marijuana confirmatory test.

B.5.2 Urine Testing Protocols

Urine specimens were obtained as unobtrusively as possible. The specimens were shipped overnight by Federal Express the day of collection or the day after, or, in the case of collection over a weekend or in a rural area without manned Federal Express stations, they were mailed within 3 days. Urine maintains stability at room temperature for 1 week, but degrades at higher temperatures. Therefore, receipt by the laboratory within 3 or 4 days better ensured the integrity of the specimen. Upon receipt, after bookkeeping and initial checks, USDTL batched and refrigerated urine specimens for analysis at a later time. USDTL provided a weekly list of specimens received at the laboratory.

The testing laboratory used two different immunoassay methods for the urine drug screening tests during the study: fluorescence polarization immunoassay (FPIA) was used from January 2000 to May 2001, and enzyme-multiplied immunoassay technique (EMIT) was used from June 2001 to December 2001. Abbott FPIA reagents were used for the FPIA tests, and Syva EMIT II reagents were used for the EMIT tests. Screening tests performed were for marijuana (cannabinoids), cocaine, opiates, and amphetamines. Specimens screened positive (i.e., with results at or above the Validity Study cutoff concentrations) were tested using gas chromatography/mass spectrometry (GC/MS) for targeted drug analytes as the confirmatory test. Urine specimens were also screened for cotinine, the principal metabolite of nicotine, using only an immunoassay test (i.e., enzyme-linked immunosorbent assay, ELISA). International Diagnostics Systems ELISA reagents were used for this test. No confirmatory testing was performed for cotinine.

FPIA was chosen as the screening test for drugs because it is a semiquantitative method whose results correlate well with quantitative GC/MS results for the analyzed drugs. Due to FPIA reagent supply problems, the laboratory switched screening test methods in May 2001 and used EMIT for the remainder of the study. The EMIT screening test results were obtained in a semiquantitative mode. Confirmation was performed using GC/MS. Any specimens screened

positive (i.e., results at or above the cutoff shown in Table B.2) were confirmed by GC/MS using the laboratory's established limits of quantitation for its GC/MS tests as the cutoffs (as shown in Table B.3).

Table B.2 Validity Study Screening Cutoff Concentrations for Urine

Urine Screening Test	Cutoff (ng/mL)
Marijuana (Cannabinoids)	30
Cocaine	50
Opiates	50
Amphetamines	500
Cotinine	100

ng/mL = nanograms per milliliter.

Table B.3 Validity Study Confirmatory Cutoff Concentrations for Urine

Urine GC/MS Drug Test	Analyte(s)	Cutoff (ng/mL)
Marijuana	Delta-9-Tetrahydrocannabinol Carboxylic Acid (THCA)	2
Cocaine	Benzoylecgonine	5
Opiates	Codeine	5
	Morphine	5
Amphetamines	Amphetamine	25
	Methamphetamine	25

GC/MS = gas chromatography/mass spectrometry; ng/mL = nanograms per milliliter.

B.5.3 Hair Analysis Protocols

For the Validity Study, the first 1.3 cm segment of hair from the proximal end was analyzed to correspond to questions about drug use in the past month.

ELISA was used as the screening test for hair specimens, and GC/MS was used as the confirmatory test. When the study was launched, USDTL suggested cutoff concentrations for hair screening and confirmatory tests (see Tables B.4 and B.5).

Table B.4 Validity Study Screening Cutoff Concentrations for Hair

Hair Screening Test	Cutoff (ng/mg)
Marijuana (Cannabinoids)	0.05 in Year 1[a] 0.005 in Year 2[a]
Cocaine	0.5
Opiates	0.5 in Year 1[a] 0.2 in Year 2[a]
Amphetamines	0.5

ng/mg = nanograms per milligram.

[a] The cannabinoid and opiate screening test cutoffs were lowered in year 2 of the Validity Study (2001).

Table B.5 Validity Study Confirmatory Cutoff Concentrations for Hair

Hair GC/MS Drug Test	Analyte(s)	Cutoff (ng/mg)
Marijuana	Delta-9-Tetrahydrocannabinol Carboxylic Acid (THCA)	0.005 in Year 1[a] 0.001 in Year 2[a]
Cocaine	Cocaine	1
	Benzoylecgonine	0.1
	Cocaethylene	0.1
Opiates	Codeine	0.2
	Morphine	0.2
	6-Acetylmorphine	0.2
Amphetamines	Amphetamine	0.3
	Methamphetamine	0.3

GC/MS = gas chromatography/mass spectrometry. ng/mg = nanograms per milligram.

[a] The THCA confirmatory test cutoff was lowered in year 2 of the Validity Study (2001).

For the Validity Study, we wanted to use state-of-the-art information on cutoffs for hair analysis. The Federal Government has been working to develop standardized scientific and technical guidelines for workplace drug testing programs for the testing of hair, sweat, and oral fluid specimens in addition to urine specimens. In April 2000, while the study was in the field, SAMHSA issued its draft *Mandatory Guidelines for Federal Workplace Drug Testing Programs* (SAMHSA, 2000-2001) and began its pilot performance testing (PT) program for hair testing laboratories. Subsequent draft versions of the guidelines (i.e., #2, #3, and #4) were published during the study period (SAMHSA, 2000-2001). All drafts proposed the same cutoff concentrations for the detection of drugs in hair, and these cutoffs have been used throughout the pilot PT program for hair. The cutoffs (shown in Table B.6) were based on SAMHSA-organized working groups consisting of representatives from government, research, and the hair testing industry.

Although USDTL worked to reduce their cutoffs for marijuana screening, for much of the study, their screening cutoff was 50 times higher than that specified in the SAMHSA draft guidelines. Likewise, USDTL's confirmatory test for marijuana was not sufficiently sensitive to

use the SAMHSA cutoff of 0.05 pg/mg (0.00005 ng/mg) (Table B.7). The laboratory's cutoff for the marijuana confirmatory test was 0.005 ng/mg, which is 100 times higher than the SAMHSA confirmatory test cutoff in the published drafts. In year 2, USDTL revised its procedures, which enabled the laboratory to lower the screening test cutoff for marijuana from 0.05 ng/mg to 0.005 ng/mg (i.e., still 5 times higher than the SAMHSA draft cutoff), and the GC/MS confirmation cutoff from 0.005 ng/mg to 0.001 ng/mg (i.e., still 20 times as high as the SAMHSA draft cutoff at the time of the study).

The Validity Study also used a higher screening test cutoff for opiates (0.5 ng/mg) than specified in the draft SAMHSA guidelines (0.2 ng/mg). This was lowered in year 2 of the study, but some positives may have been missed in the first year of the study.

Table B.6 SAMHSA June 2000 Draft Screening Cutoff Concentrations for Hair

Hair Screening Test	SAMHSA Draft Cutoff (ng/mg)[a]
Marijuana (Cannabinoids)	0.001
Cocaine	0.5
Opiates	0.2
Amphetamines	0.5

ng/mg = nanograms per milligram.

[a] Units of measurement are picograms per milligram (pg/mg) for hair in the SAMHSA draft *Mandatory Guidelines for Federal Workplace Drug Testing Programs*: 1.0 ng = 1,000 pg.

Table B.7 SAMHSA June 2000 Draft Confirmatory Cutoff Concentrations for Hair

Hair GC/MS Confirmation Drug Test	Analyte(s)	Draft Cutoff (ng/mg)[a]
Marijuana	THCA	0.00005
Cocaine and Metabolites	Cocaine	1.0
	Benzoylecgonine	0.1
Opiates	Codeine	0.2
	Morphine	0.2
	6-Acetylmorphine	0.2
Amphetamines	Amphetamine	0.3
	Methamphetamine	0.3

GC/MS = gas chromatography/mass spectrometry. ng/mg = nanograms per milligram. THCA = delta-9-tetrahydrocannabinol carboxylic acid.

[a] Units of measurement are picograms per milligram (pg/mg) for hair in the SAMHSA draft *Mandatory Guidelines for Federal Workplace Drug Testing Programs*: 1.0 ng = 1,000 pg.

Table B.8 Drugs and Drug Groups as Shown in Pill Cards for Pain Reliever and Stimulant Prescription Psychotherapeutic Drugs in NHSDA

Drug	Drug Groups at Top of Pill Card[1]	Drugs at Bottom of Pill Card[2]
Pain Relievers	(1) Darvocet®, Darvon®, or Tylenol® with Codeine (2) Percocet®, Percodan®, or Tylox® (3) Vicodin®, Lortab®, or Lorcet®/Lorcet Plus®	(4) Codeine (5) Demerol® (6) Dilaudid® (7) Fioricet® (8) Fiorinal® (9) Hydrocodone (10) Methadone (11) Morphine (12) OxyContin® (13) Phenaphen® with Codeine (14) Propoxyphene (15) SK-65® (16) Stadol® (17) Talacen® (18) Talwin® (19) Talwin® NX (20) Tramadol (21) Ultram®
Stimulants	(1) Methamphetamine ("crank," "crystal," "ice," or "speed"), Desoxyn®, or Methedrine® (2) Prescription Diet Pills (examples given: amphetamine, Benzedrine®, Biphetamine®, Fastin®, and phentermine) (3) Ritalin® or Methylphenidate	(4) Cylert® (5) Dexedrine® (6) Dextroamphetamine (7) Didrex® (8) Eskatrol® (9) Ionamin® (10) Mazanor® (11) Obedrin-LA® (12) Plegine® (13) Preludin® (14) Sanorex® (15) Tenuate®

[1] For drug groups shown at the top of a pill card, separate questions were asked for each drug group.
[2] For drugs shown at the bottom of a pill card, all drugs were asked in tandem with additional questions to identify the specific drug(s) used.

Table B.9 Location of Validity Study Questions Used to Produce 30-Day, 7-Day, and 3-Day Self-Reported Use of Various Substances

Substance	Substance Group	30-Day Self-Reported Use	7-Day Self-Reported Use	3-Day Self-Reported Use
Cigarettes	Tobacco	Core Repeat	Repeat	Follow-Up Repeat
Snuff	Tobacco	Core	Not Available	Follow-Up
Chewing Tobacco	Tobacco	Core	Not Available	Follow-Up
Cigars	Tobacco	Core Repeat	Repeat	Follow-Up Repeat
Pipe	Tobacco	Core	Not Available	Follow-Up
Marijuana/Hashish	Marijuana	Core Repeat	Repeat	Follow-Up Repeat
Cocaine	Cocaine	Core Repeat	Repeat	Follow-Up Repeat
Crack[1]	Cocaine	Core		
Heroin	Opiates	Core Repeat	Repeat	Follow-Up Repeat
Prescription Pain Relievers (Misuse)	Opiates	Core Repeat	Repeat	Repeat
Stimulants[2,3] (Misuse)	Amphetamines	Core Repeat	Repeat	Follow-Up Repeat
Methamphetamine, Desoxyn, or Methedrine[3] (Misuse)	Amphetamines	Core Repeat	Repeat	Follow-Up Repeat
Prescription Diet Pills[3-5] (Misuse)	Amphetamines	Follow-Up Repeat	Repeat	Follow-Up Repeat
Over-the Counter (OTC) Diet Pills or Stay-Awake Pills	Amphetamines	Repeat	Repeat	Repeat
OTC Cold Medicine	Amphetamines	Repeat	Repeat	Repeat
OTC or Prescription Nicotine Products to Help Quit Smoking[6]	Tobacco	Repeat	Repeat	Repeat
Prescription Pain Relievers[1] (Licit Use)	Opiates	Repeat	Repeat	Repeat
Prescription Diet Pills[4](Licit Use)	Amphetamines	Repeat	Repeat	Repeat
Marinol®[7] (Licit Use)	Marijuana	Repeat	Repeat	Repeat

[1] Repeat questions on 3-day and 7-day cocaine use (including crack) were asked, but questions on 3-day or 7-day use of crack specifically did not exist.

[1] Question asked about use of prescription pain relievers shown in Table B.8.

[2] Question asked about use of prescription stimulants shown in Table B.8.

[3] Follow-up questions on 3-day use were added in 2001.

[4] Question asked about use of prescription diet pills, such as amphetamines, Benzedrine®, Biphetamine®, Fastin®, or phentermine.

[5] Follow-up questions on 30-day use were added in 2001.

[6] Question asked about use of products, such as nicotine gum, nicotine nasal spray, nicotine patch, or nicotine inhaler.

[7] Marinol® (i.e., dronabinol*)* is a prescription drug used to treat appetite loss associated with weight loss in people diagnosed with AIDS. Dronabinol is a synthetic version of delta-9-tetrahydrocannabinol, which could cause a positive marijuana test.

Appendix C:
Sample Design

C.1 Design Overview

The design of the Validity Study closely reflects that of the main National Household Survey on Drug Abuse (NHSDA) fielded in 2000 and 2001.[12] This multistage approach began with the selection of predefined regions and ended with the selection of one eligible participant per household. An important difference between the design of the Validity Study and that of the NHSDA is that in the Validity Study an additional level of clustering was imposed at the first stage of selection. This additional clustering was introduced to minimize costs associated with Validity Study data collection.

The initial stage of selection entailed subselecting 200 field interviewer (FI) regions from among the 876 FI regions (excluding Hawaii and Alaska) for the NHSDA.[13] These FI regions were selected randomly within strata defined by categorizing past month marijuana use. This stratification was imposed in order to obtain more current drug users.

In the second stage, one segment per FI region was selected from previously "retired" 1999-2001 NHSDA segments two quarters earlier. For example, the 2000 quarter 1 Validity Study segments were used in the 1999 quarter 3 NHSDA, then "retired" from the NHSDA. This process is further explained in Section C.6.

In the third stage of sample selection, dwelling units (lines) were selected. Specifically, the primary objective for line selection was to determine the minimum number of dwelling units needed in each segment to meet the targeted sample size for the two age groups. Listing unit sample sizes for each segment were determined by using a combination of the sampling strata by age group with the largest sampling rate. We refer to this as the "driving" age group. Using more recent data from another source, Claritas, the 1990 census data were adjusted and age-specific sampling rates were computed. These rates were adjusted by the following factors: FI regions' probability of selection, the segment's probability of selection, the subsegmentation inflation factor[14] (if any), the probability of selecting a person in the age group (0.99 for the driving age group), and an adjustment for the "maximum-of-1" rule.[15] In addition to these factors, historical data from the 1998-2001 NHSDAs[16] were used to compute predicted screening and interviewing

[12] Prior to 2002, the National Survey on Drug Use and Health (NSDUH) was called NHSDA.

[13] A total of 900 regions was created for the NHSDA. However, the Validity Study excluded Hawaii and Alaska, which reduces the number of FI regions to 876. For more information on how the FI regions were created, see the 1999 NHSDA sample design report (Bowman, Penne, Chromy, & Odom, 2001).

[14] Segments found to be very large in the field are partitioned into equally representative *subsegments*. Then, one subsegment is chosen at random to be fielded. The subsegmentation inflation factor accounts for the narrowing down of the segment.

[15] Only one eligible person was selected per household. Probability proportional to size (PPS) sampling was used to perform this requirement. Thus, sampling rates are adjusted to satisfy this constraint.

[16] The 1998 NHSDA data were used for response rates due to the unavailability of 1999 data when the initial sample selection occurred.

response rate adjustments. For the Validity Study, additional biological specimen response rates were incorporated into an overall adjustment for each age group. The final adjusted sampling rate then was multiplied by the actual number of dwelling units found in the field during counting and listing activities. The product represents the segment's listing unit sample size.

Using random start point and interval-based (systematic) selection, the actual listing units were selected from the segment frame. Households selected in the earlier NHSDAs were not eligible for selection in the Validity Study. After dwelling unit selections were made, an interviewer visited each selected dwelling unit to obtain a roster of all persons residing in the dwelling unit. As in previous years, during the data collection period, if an interviewer encountered any new dwelling unit in a segment or found a dwelling unit that was missed during the original counting and listing activities, then the new/missed dwellings were selected into the 2000-2001 Validity Study using a half-open interval selection technique.[17] This technique eliminates any frame bias that might be introduced because of errors and/or omissions in the counting and listing activities and also eliminates any bias that might be associated with using "old" or "retired" segment listings.

Using the roster information obtained from an eligible member of the selected dwelling unit, 0 or 1 person was selected for the survey. Sampling rates were preset by age group and marijuana stratum. Roster information was entered directly into the electronic screening instrument, which automatically implemented this fourth stage of selection based on the stratum and age group sampling parameters.

C.2 Target Population

The respondent universe for the 2000-2001 Validity Study mirrored the 2000-2001 NHSDA. It consisted of the civilian, noninstitutionalized population, including noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes) and civilians residing on military bases. Persons excluded from this universe included those with no fixed household address (e.g., homeless transients not in shelters) and residents of institutional group quarters, such as jails and hospitals.

There were several differences between the target populations of the NHSDA and the Validity Study. The Validity Study respondents had to be between the ages of 12 and 25 and reside within the continental United States (including the District of Columbia). Statistics show there is a greater prevalence of recent drug use in this younger population than there is for the full age range of the general population surveyed in the NHSDA. As a means to increase the expected number of drug users in the Validity Study sample, the population was limited to this younger age group. Another major difference between the two surveys was the limitation placed on the number of participants from each household. To maintain the experimental design of the Validity Study, a maximum of one person per household was sampled as opposed to a maximum of two persons in the NHSDA. To obtain the required precision, the Validity Study sample size

[17] In summary, this technique states that if a dwelling unit is selected for the 1999 study and an interviewer observes any new or missed dwelling units between the selected dwelling unit and the dwelling unit appearing immediately after the selection on the counting and listing form, then all new/missed dwellings falling in this interval will be selected. If a large number of new/missed dwelling units are encountered (generally greater than six), then a sample of the missing dwelling units will be selected.

consisted of 1,000 per age group (12 to 17 and 18 to 25) per year. A sample of 2,000 per age group and 4,000 overall was to be available at the completion of the Validity Study.

C.3 General Optimal Allocation Procedures

The optimization procedures of the Validity Study were similar to those of the NHSDA and were specifically designed to address multiple precision and design requirements while minimizing the cost of data collection. Minimization was achieved by determining the smallest number of interviews and selected dwelling units necessary to achieve the various design requirements. This required a four-step process:

1. The first step was to determine the optimal number of interviews (i.e., fourth stage of selection sample sizes) by age groups needed to satisfy precision requirements for several drug use outcome measures. We sought to determine m_{wa} for each age group a. Details are given in Section C.4.

2. The second step was to determine the number of FI regions necessary to achieve these sample sizes using the m_{wa}, computed in the previous step. Because the survey design requires that only "retired" NHSDA segments be used (one per quarter), the cluster sizes available were limited to these "retired" NHSDA segments. Knowing this information, one can calculate the number of FI regions to sample from the NHSDA. Details are given in Section C.5.2.

3. The next step was by using m_{wa} to determine the optimal number of selected dwelling units (D_{wj}), for each segment selected in step 2 and with segments classified into one of the four strata necessary to meet our desired sample size and to minimize costs. Similar to the NHSDA, this was achieved quarterly through parameter constraints at the strata and segment level. In addition, only listed households not previously sampled in the NHSDA were eligible for Validity Study selection. Procedures for this step are further described in Section C.7.

4. The fourth and final step in the optimization process entailed determining age group–specific probabilities (S_{wa}) for each stratum. This was achieved by utilizing PPS sampling. Further details are presented in Section C.8. After the calculation of both selection probabilities and required dwelling unit sample sizes, sample size constraints at the screening level and interview level were applied.

C.3.1 Notation

a = Age group. a = 1 or 2 and represents the following groups: 12 to 17 year olds and 18 to 25 year olds, respectively.

j = Individual segment indicator (total of 800 per year; 200 per quarter). These segments are a subsample of the segments in the main study.

w = Marijuana use estimation stratum. This is classified at the FI region level, and subsequently all segments in the same FI region are within the same stratum. We defined FI regions into this stratum using 1999 data from quarters 1 and 2 questionnaire data. This definition was used for design purposes.

s = Marijuana use estimation stratum. The definition for each stratum is the same as in "w." However, because data on screening and interviewing rates were not available for the 1999 sample at the time of the first quarter allocation (quarter 1 of 2000), 1998 data were used instead. To use the 1998 data, we allocated the 1998 segments (FI regions were not available in the 1998 NHSDA sample) into the marijuana stratum using the same definitions. This definition was used for response and yield rates.[18]

m_{wa} = Number of completed interviews (respondents) desired in each stratum w and age group a. Computation of m_{wa} is discussed in Section C.8. For purposes of quarterly computation of selected dwelling unit sample size, 25 percent of the yearly estimate was used.

y_{wa} = Estimated number of persons in the target population in the stratum w and age group a. These values were computed using the adjusted census block counts. The block values were aggregated to the FI region level, then each FI region was classified into one of the four strata.

f_{wa} = m_{wa} / y_{wa}. Stratum and age group–specific sampling fraction.

F_w = $\text{Max}\{f_{wa}; \text{a}=1..2\}$.

P_{wj} = Inverse of the segment selection probability. To reduce costs, we visited a subsample of the NHSDA-selected FI regions/segments. This subsample consisted of one segment per quarter within a randomly selected FI region. In short, each segment's selection probability consisted of the FI region selection probability and the selection probability of the segment within the FI region. Because the NHSDA selected two segments per quarter, and the Validity Study further selected only one of these segments per quarter, and each quarter only consisted of 25 percent of the segment sample, selection probabilities were further adjusted by a factor of 8 to sum to the yearly totals.

I_{wj} = Subsegmentation inflation factor.

D_{wj} = Minimum number of dwelling units to select for screening in segment j to meet the targeted sample sizes for each stratum w and age group a.

L_{wj} = Final segment count of dwelling units available for screening.

S_{wja} = Stratum, segment-specific probability of selecting a person in age group a. As with the NHSDA, no single selection probability could exceed one.

ε_s = Stratum, s, specific dwelling unit eligibility rate. Derived from 1998 NHSDA data by classifying segments into the marijuana stratum and computing the weighted eligibility

[18] Although it would have been ideal to use "w" in all computations, FI region level data (1999) were not initially available for the response rates. As a result, response rates were calculated using 1998 data. However, in 1998, only segments were defined and not FI regions, and as a result, "s" was created at the segment level. When FI region-level data (starting in 1999 through 2001) became available, response rates were recomputed using the "w" stratum definition.

rate by dividing the weighted number of sampled households by the weighted number of eligible households. These rates then were applied to stratum "*w*."

φ_s = Stratum, *s*, specific screening response rates. Derived from 1998 NHSDA data by classifying segments into the marijuana stratum and computing the weighted screening rate by dividing the weighted number of eligible households by the weighted number of successfully screened households. These rates then were applied to stratum "*w*."

λ_{sa} = Stratum *s* and age group *a* specific interview response rate. Derived from 1998 NHSDA data by classifying segments into the marijuana stratum and computing the weighted number of successfully interviewed sampled persons. These rates then were applied to stratum "*w*."

ξ_{wa} = Stratum *w* and age group *a* specific biological specimen response rate. This rate was originally estimated to be 75 percent and was later replaced with an actual rate of 89 percent based on experience from the field.

γ_{sa} = Expected number of persons within an age group per dwelling unit (yield). Calculated using 1998 NHSDA data by dividing the weighted total number of rostered persons in an age group *a* by the weighted total number of screener-complete dwelling units for each stratum *s*. These rates then were applied to stratum "*w*."

δ_{sa} = Stratum *s* and age group *a* specific "maximum-of-1" rule adjustment. The survey design restricted the number of interviews per dwelling unit to a total of one, which was achieved through PPS sampling. This resulted in a loss of potential interviews in dwelling units where selection probabilities summed to more than one. The adjustment was designed to inflate the number of required dwelling units to compensate for this loss. These rates then were applied to stratum "*w*." This procedure was iterative and utilized 1998 NHSDA data as follows:

1. *Determine the number of dwelling units (R_{wa})*. Determine the quantity necessary to obtain desired person sample sizes given the desired sample sizes in each stratum, age group.

$$R_{wa} = \frac{m_{wa}}{(\varepsilon_s * \varphi_s * \lambda_{sa} * \gamma_{sa} * \delta_{sa} * \xi_{wa})}, \text{ where } \delta_{sa} = 1 \text{ for first iteration.}$$

2. *Set S_{wa}* = .99 for the age group with the largest R_{wa}. All other age group probabilities are set in proportion to the largest:

$$S_{wa} = \frac{R_{wa}}{Max(R_{wa})}.$$

3. *Assign S_{wa} to respective person record in 1998 NHSDA data*. With PPS sampling, selection probabilities are now adjusted to reflect the total household composition. In short, if selection probabilities for all eligible dwelling unit members sum to

more than one, then probabilities are ratio adjusted to sum to one. This will be denoted as S^*_{wa}. However, sums less than one are unadjusted.

4. *Sum S_{wa} and S^*_{wa} within stratum.* The "maximum-of-1" rule (δ_{sa}) then is calculated as the ratio of the summed S^*_{wa} and S_{wa} (i.e., S^*_{wa} / S_{wa}).

5. *Insert new calculated δ_{sa} into step 1 and repeat steps 1 through 5.* Continue until the absolute difference between δ_{sa} of the current cycle and the previous cycle is less than .001, usually about two to three iterations.

C.4 Determining Person Sample Sizes, by Age Group

The age-specific person sample sizes were determined by meeting specified precision requirements. Because the planned analysis focused on the false-negative rates for the NHSDA computer-assisted interviewing (CAI) questionnaire, the precision requirements were placed on the relative standard error (RSE) for the false-negative rates. Table C.1 shows the projected RSE for the estimated false-negative rate for marijuana and cigarettes. It was believed that adequate estimates would be produced for the false-negative ratio if it were actually close to 0.50. At this level of false-negative reporting, the RSE for past month marijuana use for youths aged 12 to 17 with a sample of 1,000 persons per year was projected to be 0.073, and for past month cigarette use the RSE was projected to be 0.050. The RSEs were acceptable at this level, and thus the sample size of 1,000 persons per age group per year met the precision requirement for RSE of estimating false negatives near 0.50.

Table C.1 Projected Relative Standard Error for Estimated False-Negative Rate for Past Month Cigarette and Marijuana Use

Projected Relative Standard Error for the Estimated False-Negative Rate for Cigarette Use							
			Projected Relative Standard Error				
Age Group	Sample Size	1997 Self-Report	0.15	0.25	0.50	0.65	0.80
12 to 17	1,000	0.199	0.156	0.106	0.050	0.031	0.016
18 to 25	1,000	0.406	0.109	0.074	0.035	------	------
Projected Relative Standard Error for the Estimated False-Negative Rate for Marijuana Use							
			Projected Relative Standard Error				
Age Group	Sample Size	1997 Self-Report	0.15	0.25	0.50	0.65	0.80
12 to 17	1,000	0.094	0.226	0.155	0.073	0.045	0.023
18 to 25	1,000	0.128	0.194	0.133	0.063	0.038	0.020

Note: Self-report rates were based on 1997 preliminary estimates by age group for past month use. Only 1997 data were available at the time of defining the precision requirements (1998 and 1999 data were not available).

C.5 First-Stage Sample Allocation

This section discusses the computational procedures used to select the first-stage sampling units in the Validity Study. Unlike the NHSDA, which used FI region as a stratification variable, the supplemental sample defined FI region as its primary sampling unit (PSU). The

clustering of PSUs was implemented to achieve the desired precision while simultaneously reducing cost and interviewer burden. PPS and with-minimum-replacement sampling on the 876 NHSDA FI regions were implemented.

C.5.1 Initial Stratification and Formation of the Composite Size Measure

FI regions were explicitly stratified into four categories using raw 1999 NHSDA quarter 1 and 2 data based on past month marijuana use data (commonly referred to as marijuana strata). The four levels were defined at the FI region level by the lower bound of the 90 percent confidence interval (CI) of past month marijuana use and were categorized as follows: (1) no marijuana use, (2) 90 percent CI lower bound less than or equal to 0, (3) 90 percent CI lower bound greater than 0 but less than or equal to 0.10, and (4) 90 percent CI lower bound greater than 0.10. The FI region counts for each strata are listed in Table C.2.

Table C.2 Distribution of the Field Interviewer Regions for Each Marijuana Stratum

Marijuana Stratum	Description	FI Regions
1	No Marijuana Use	194
2	90% CI Lower Bound <= 0	208
3	90% CI Lower Bound in (0,0.10]	395
4	90% CI Lower Bound > 0.10	79

The composite size measure procedure was used to obtain self-weighting samples for multiple domains in multistage designs. The goal of using a composite size measure was as follows:

- Yield the targeted domain sample sizes in expectation (E_s) over repeated samples; that is, if m_{ds} is the domain-d sample size achieved by sample-s, then

$$E_s(m_{ds}) = m_d \text{ for } d=1,...,D.$$

- Constrain the maximum number of selections per dwelling unit at a specified value; specifically, limit the total number of within-dwelling unit selections across all age groups to a maximum of one.

- Minimize the number of sample dwelling units that must be screened to achieve the targeted domain sample sizes.

- Eliminate all variation in the sample inclusion probabilities within a domain except for the variation in the within-dwelling unit/within-domain probabilities of selection. The inverse probabilities of selection for each sample segment were used to determine the number of sample lines to select from within each segment. As a consequence, all dwelling units within a specific stratum were selected with approximately the same probability, and therefore, approximately equalized dwelling unit sampling weights. This feature minimizes variance inflation that results from unnecessary variation in sampling weights.

- Equalize the expected number of sample persons per cluster to balance the interviewing workload and to facilitate the assignment of interviewers to regions and segments. This feature also minimizes adverse effects on precision resulting from extreme cluster size variations.

- Simplify the size measure data requirements so that decennial census data (FI region counts) are adequate to implement the method.

Using the 1990 census data supplemented with revised population projections, a composite size measure was computed for FI regions defined within the United States excluding Hawaii and Alaska. The composite size measure began by defining the rate $f_w(d)$ at which we wished to sample each age group domain d (d=1,2) from marijuana stratum w.

Let $C_{wijk}(d)$ be the population count from domain d in census block k of segment j of FI region i within stratum w. The composite size measure for block k was defined as

$$S_{wij+} = \sum_{d=1}^{2} f_w(d) \sum_{j=1}^{N_{wij}} C_{wijk}(d).$$

The composite size measure for FI region i was calculated as

$$S_{w+} = \sum_{d=1}^{2} f_w(d) \sum_{w=1}^{N_w} C_{wijk}(d),$$

where N_{wij} equals the number of blocks within segment j of FI region i and stratum w and N_w equals the number of FI regions in stratum w.

C.5.2 FI Region Selection

Optimal allocation was used to allocate the 200 FI regions into the four marijuana strata. The allocation was defined as follows:

$$n_w \propto (N_w)\sqrt{p_w(1-p_w)},$$

where

$$p_w = \lambda p_w + (1-\lambda)p,$$

and

p_w = estimated weighted proportion of past month marijuana usage in each stratum,

p = overall weighted marijuana past month usage in the coterminous States and the District of Columbia, and

λ = the constant of proportionality.

Lambda (λ) equal to 0.5 was chosen due to its modest unequal weighting effect and the equality of using both the stratum-specific marijuana proportions and the overall marijuana proportion, while still oversampling the higher marijuana usage areas. Table C.3 summarizes the distribution of the 200 FI regions into the four marijuana strata.

Table C.3 Validity Study Number of Selected Field Interviewer Regions per Marijuana Stratum

Marijuana Stratum	Number of FI Regions	Number of Selected FI Regions
1	194	33
2	208	48
3	395	96
4	79	23
Total	876	200

Chromy's (1979, 1981) probability-minimum-replacement sequential-sampling procedure was used to select the desired number of FI regions within each marijuana stratum. The expected frequency of selection is given by:

$$P_{wi} = n_w S_{wi+} / S_{w++},$$

where n_w is the selected number of FI regions for each specific stratum, and S_{w++} is the sum of the composite size measure over FI regions in each stratum. To make 200 FI region selections from the frame of 876 FI regions, Chromy's procedure partitioned the FI regions, based upon their size measures $S_{wi+,}$ into *nw* zones of equal size (individual FI regions may have straddled zone boundaries) for each explicit stratum. The selection of FI regions was independent between strata. Exactly one sample FI region then was randomly selected from each zone. This zoned sequential selection made possible a deep implicit stratification of PSUs by a controlled ordering of the first-stage frame. Moreover, the zones were defined so that all pairs of PSUs had a chance of appearing together in the sample, a requirement for unbiased estimation of sampling variances.

The probability-minimum-replacement feature of Chromy's (1979, 1981) procedure refers to the treatment of PSUs for which the expected number of selections exceeds one (e.g., self-representing PSUs). The actual number of times a PSU can be selected for the sample differs from the expected number by less than one, and the average number of selections over all possible implementations of Chromy's procedure equals the expected number.

Using data estimated from the 1990 census supplemented with revised population counts, a serpentine ordering was implemented by race/ethnicity, census region, and a rural/urban indicator for each FI region. The serpentine nature of the sort maximizes the similarity of adjacent FI regions in the ordered list. More specifically, FI regions were ordered first by race/ethnicity. Therefore, the list is ordered such that FI regions with similar levels of racial composition are next to each other. The next level of ordering is by census region. This ordering within census region places FI regions in similar regions of the United States adjacent to each

other. This provides near proportional distribution across second and third levels of stratification within the slightly disproportionate allocation to the first level of stratification.

Several miscellaneous topics arose during first-stage sample selection. For management purposes, 400 FI regions were selected instead of 200. Then out of these 400 FI regions, 200 were selected with equal probability. This measure was taken in case an FI did not agree to work with biological specimens. Although this did not occur, it was a possibility. Also, due to the nature of the probability-minimum-replacement procedure, one FI region was selected twice. Due to limited availability of unused lines in Validity Study segments, it was decided to replace this duplicate FI region with a previously unselected FI region. The duplicate FI region was replaced with the FI region in the zone immediately before the duplicated FI region.

C.6 Second-Stage Sample Allocation

Segments, the second-stage sampling units, are defined as adjacent census blocks within FI regions. The Validity Study used one segment per FI region per quarter unlike the NHSDA, which used two segments. To minimize costs, the second-stage sampling unit consisted of previously selected NHSDA segments. More specifically, the Validity Study used retired NHSDA segments with a two-quarter lag. For example, in quarter 1 of 2000, segments ending in "05" were used. These were previously utilized in the quarter 3 of 1999 NHSDA sample. Detailed segment identification number suffixes (i.e., last two digits of the segment identification number) can be viewed in Table C.4. One exception to this rule existed. If the Validity Study segment had zero remaining lines after being used in the NHSDA, then it was replaced with a segment ending in "03" (retired segment from quarter 2 of 1999) from the same FI region as the exhausted segment. This occurred in 2000 during quarters 3 and 4 and in 2001 during quarters 1 and 4 for the Validity Study. Table C.5 lists the exhausted segments and their replacements.

Table C.4 Validity Study Segment Identification Number Suffixes

Year	Quarter	Segment Suffix	Used in NHSDA
2000	1	05	1999 Quarter 3
	2	07	1999 Quarter 4
	3	02	2000 Quarter 1
	4	04	2000 Quarter 2
2001	1	06	2000 Quarter 3
	2	08	2000 Quarter 4
	3	09	2001 Quarter 1
	4	10	2001 Quarter 2

For specifics on how the segments were formed and selected for the NHSDA, refer to the *1999 National Household Survey on Drug Abuse: Sample Design Report* (Bowman et al., 2001).

Table C.5 Validity Study Exhausted Segments and Their Replacements

Year	Quarter	Exhausted Segments	Replacement Segments
2000	1	None	---
	2	None	---
	3	CA0802, CA4002, GA0302, MA1202	CA0803, CA4003, GA0303, MA1203
	4	CA2104, GA0804, NC0304, VA1004	CA2103, GA0803, NC0303, VA1003
2001	1	CA0406, CA3906, CA4106, CA4206, GA0206, NJ0206	CA0403, CA3903, CA4103, CA4203, GA0203, NJ0203
	2	None	---
	3	None	---
	4	ND0110, NE0410, NJ1110	ND0103, NE0403, NJ1103

C.7 Third-Stage Sample Allocation

The third-stage sampling units consisted of dwelling units that were selected based on the required number of respondent sample sizes for each stratum and age group. These third-stage units were selected quarterly to take advantage of differences in segment characteristics within each quarter and to allow for design parameter adjustments. The allocation was selected by computing the minimal number of dwelling units needed for each segment. The procedures described below are quarter 1 specific, and any deviations are described in Section C.8.3.

The formula utilized to optimally minimize the required number of dwelling units is as follows:

$$f_{wa} = P_{wj} * \mathrm{I}_{wj} * \left(\frac{D_{wj}}{L_{wj}} \right) * S_{wa} * \varphi_s * \lambda_{sa} * \delta_{sa} * \xi_{wa} .$$

At this point in the procedure, only two components in the formula are unknown: D_{wj} and S_{wja}. Selection probabilities are segment and age group specific and, to maximize the number of selected persons within a dwelling unit, the age group whose sampling fraction (f_{wa}) = F_w , known as the "driving age group," was set to the largest allowable selection probability (S_{hwja}) of .99. Thus, D_{wj} was solved for each of the pre-specified driving age groups:

$$D_{wj} = \frac{f_{wa}}{(P_{wj} * I_{wj} * S_{wja} * \varphi_s * \lambda_{sa} * \delta_{sa} * \xi_{wa})} * L_{wj}.$$

In addition to these formulas, another constraint was implemented. As stated earlier, the Validity Study used retired NHSDA segments with a two-quarter lag. Therefore, a number of dwelling units had already been sampled from the Validity Study segments. Consequently, the Validity Study was only allowed to select dwelling units that were not selected previously in the NHSDA. If no remaining dwelling units were left, the segment was replaced (see Section C.6).

The process by which the dwelling unit frame was constructed is called counting and listing.[19] A certified lister visited the selected area and listed a detailed and accurate address (or description if no address was available) for each dwelling unit within the segment boundaries. The lister was given a series of maps on which he or she also made note of the location of these dwelling units. The resulting list of dwelling units was entered into a database and served as the frame from which the third-stage sample was drawn.

In some situations, the number of dwelling units within the segment boundaries was much larger than the specified maximum (based on census information). To obtain a reasonable number of dwelling units for the frame, the lister first counted the dwelling units within the segment boundaries. The sampling staff at RTI then partitioned the segment into smaller pieces or subsegments and randomly selected one to be listed.

During counting and listing, the lister moved about the segment in a prescribed fashion called the "continuous path of travel." In short, the lister attempted to move in a clockwise fashion, making each possible right turn, making U-turns at segment boundaries, and not breaking street sections. Following these defined rules and always looking for dwelling units on the right-hand side of the street, the lister minimized the chance of not listing a dwelling unit within the segment. Also, using this defined path of travel made it easier for the FI to locate the sampled dwelling units. Finally, the continuous path of travel laid the groundwork for the half-open interval procedure for recovering missed dwelling units.

C.8 Fourth-Stage Sample Allocation

Having solved for D_{wj}, the selection probabilities for the remaining age groups were determined:

$$S_{wja} = \frac{f_{wa}}{P_{wj} * I_{wj} * \left(\frac{D_{wj}}{L_{wj}} \right) * \varphi_s * \lambda_{sa} * \delta_{sa}}.$$

At this point, another constraint was implemented. Only one person per dwelling unit was selected to participate in the Validity Study. This differed from the NHSDA, which selected up to two persons per dwelling unit. As noted previously, only individuals aged 12 to 25 were eligible for selection. This differed from the NHSDA, whose population consisted of those aged 12 or older.

C.8.1 Sample Size Constraints: Remaining Sample and Reducing Field Interviewer Burden

The NHSDA was designed to ensure that an adequate sample of eligible dwelling units remained within each segment. This was implemented to provide for a yearly 50 percent overlap across segments, as well as to allow the Substance Abuse and Mental Health Services

[19] All counting and listing tasks were performed for the NHSDA, then borrowed for the Validity Study. A summary of the counting and listing procedures for the NHSDA follows.

Administration (SAMHSA) to implement supplemental studies. The Validity Study took advantage of this by using the sample remaining from the retired NHSDA segments.

Concerns were noted about guarantees that FIs would be able to complete the amount of work assigned to them within the quarterly timeframe. These concerns prompted the following constraints to be placed on the D_{wj} sample size:

1. Number of selected dwelling units for screening: < 50 or < ½*$L_{wj.}$ Adjustments were made by adjusting the D_{hj} counts to equal the minimum of 50 or ½*L_{wj} .

2. Expected number of interviews: < 20.

This expected number of interviews (m^*_{wja}) was computed for the Validity Study as follows:

$$m^*_{wja} = D^*_{wj} * \varepsilon_s * \varphi_s * \gamma_{sa} * S_{wja} * \lambda_{sa} * \delta_{sa},$$

where D^*_{hj} has been adjusted for constraint 1. This value was the total number of interviews expected within each segment. The calculation of the adjustment was as follows:

$$20 / (m^*_{wja}).$$

This adjustment was applied to D_{wj} under the assumption of an equal number of screened dwelling units for each completed interview.[20]

Both constraints 1 and 2 reduced the third-stage sample. This, in turn, could have reduced the expected fourth-stage sample size. Therefore, the reduction in the third-stage sample was reallocated back to the FI region by applying a marginal adjustment to the fourth-stage sample size (m_{wa}) at the stratum and age group level. As a result, segments that were not subject to these constraints could be affected. This adjustment to reallocate the dwelling unit sample was iterative until the expected person sample sizes were met.

C.8.2 Dwelling Unit Selection and Release Partitioning

After derivation of the required dwelling unit sample size (D_{wj}), the sample was selected from the frame of counted and listed dwelling units for each segment (L_{wj}). The frame was ordered in the same manner as described in the third-stage sample allocation procedures (see Section C.7). Selection was completed using systematic sampling with a random start value.

Some early problems with field staff and inaccurate sample projections prompted a sample partitioning procedure. The entire sample (D_{wj}) would still be selected, but only certain percentages of the total would be released into the field. An initial percentage, based on

[20] The optimization procedures implemented for the derivation of D_{wj} assigned the larger dwelling unit samples to FI regions with better response rates. Often, such FI regions were the first to be affected by the sample size constraints. Hence, when forced to reallocate the reduction in dwelling unit sample size to an FI region with poorer response rates, the overall dwelling unit sample size would increase in nonlinear amounts. In short, FI regions with lower response rates required more screened dwelling units per completed interview.

interquarter work projections, was released to all FI regions at the beginning of the quarter. Additional percentages would be released if it was concluded that the field staff could handle the added workload. Each partitioning of the sample was a valid sample and helped to control the amount of nonresponse without jeopardizing the validity of the study. In every quarter, a reserve sample also was selected, over and above the required D_{wj} sample, to try to compensate for any shortcomings in previous quarters. Quarterly sample data are summarized in Table C.6.

Table C.6 Quarterly Dwelling Unit Sample Sizes and Percent Released

Strata	2000 Quarter 1			2000 Quarter 2		
	# Sampled	# Released	%	# Sampled	# Released	%
Total	4,181	3,800	91	4,747	3,951	83
No Marijuana Use	691	631	91	792	661	83
90% CI Lower Bound ≤ 0	976	885	91	1,154	963	83
90% CI Lower Bound (0,0.1]	2,004	1,821	91	2,273	1,889	83
90% CI Lower Bound > 0.10	510	463	91	528	438	83
Strata	**2000 Quarter 3**			**2000 Quarter 4**		
	# Sampled	# Released	%	# Sampled	# Released	%
Total	3,480	3,170	91	3,152	3,152	100
No Marijuana Use	596	543	91	526	526	100
90% CI Lower Bound ≤ 0	812	739	91	761	761	100
90% CI Lower Bound (0,0.1]	1,668	1,518	91	1,497	1,497	100
90% CI Lower Bound > 0.10	404	370	91	368	368	100
Strata	**2001 Quarter 1**			**2001 Quarter 2**		
	# Sampled	# Released	%	# Sampled	# Released	%
Total	3,468	3,150	91	3,319	3,319	100
No Marijuana Use	574	521	91	549	549	100
90% CI Lower Bound ≤ 0	855	777	91	831	831	100
90% CI Lower Bound (0,0.1]	1,629	1,480	91	1,579	1,579	100
90% CI Lower Bound > 0.10	410	372	91	360	360	100
Strata	**2001 Quarter 3**			**2001 Quarter 4**		
	# Sampled	# Released	%	# Sampled	# Released	%
Total	3,603	3,603	100	3,825	3,186	83
No Marijuana Use	602	602	100	629	526	83
90% CI Lower Bound ≤ 0	892	892	100	907	756	83
90% CI Lower Bound (0,0.1]	1,720	1,720	100	1,834	1,524	83
90% CI Lower Bound > 0.10	389	389	100	455	380	83

CI = confidence interval.

C.8.3 Quarter-by-Quarter Deviations

This section describes modifications implemented in the process of design optimization. *Design* refers to deviations from the original proposed plan of design. *Procedural* refers to changes made in the calculation methodologies. Finally, *dwelling unit selection* addresses changes that occurred after sample size derivations—specifically, corrections implemented during fielding of the sample (i.e., sample partitioning, as described in Section C.8.2).

C.8.3.1 2000 Quarter 1

Design: No changes.

Procedural: Used a 75 percent biological specimen response rate.

Dwelling Unit Selection: In addition to the 100 percent sample, a supplemental 10 percent was selected.

Release 1: 100 percent of original sample.
Release 2: 100 percent of the supplemental sample (10 percent of original sample).

C.8.3.2 2000 Quarter 2

Design: Because of the higher than expected yield in quarter 1, only 97 percent of the original sample was targeted (i.e., 485 completed interviews with at least one biological specimen).

Procedural: The 1998 data were used for the following rates: yield rates and "Max of 1" adjustment. The 1999 quarters 3 and 4 NHSDA data from FI regions that correspond to the marijuana strata for eligibility, screening, and interview response rates were used. A 75 percent biological specimen response rate was used.

Dwelling Unit Selection: In addition to the 100 percent sample, a supplemental 20 percent was selected.

Release 1: 75 percent of original sample.
Release 2: 25 percent of original sample.
Release 3: 100 percent of the supplemental sample (20 percent of original sample).

C.8.3.3 2000 Quarter 3

Design: No changes.

Procedural: The 1998 data were used for the following rates: yield rates and "Max of 1" adjustment. The 1999 quarters 3 and 4 NHSDA and 2000 quarter 1 NHSDA data from FI regions that correspond to the marijuana strata for eligibility, screening, and interview response rates were used. An 89 percent biological specimen response rate, which was obtained from quarter 1 field experience, was used.

Dwelling Unit Selection: In addition to the 100 percent sample, a supplemental 10 percent was selected.

Release 1: 100 percent of original sample.
Release 2: 100 percent of the supplemental sample (10 percent of original sample).

C.8.3.4 2000 Quarter 4

Design: Because of the higher than expected yield in quarters 1 and 2, only 92 percent of the original sample was targeted (i.e., 460 completed interviews with at least one biological specimen).

Procedural: The 1998 data were used for the following rates: yield rates and "Max of 1" adjustment. The 1999 quarters 3 and 4 NHSDA and 2000 quarters 1 and 2 NHSDA data from FI regions that correspond to the marijuana strata for eligibility, screening, and interview response rates were used. An 89 percent biological specimen response rate, which was obtained from quarters 1 and 2 field experience, was used.

Dwelling Unit Selection: In addition to the 100 percent sample, a supplemental 10 percent was selected.

Release 1: 100 percent of original sample.
Release 2: 100 percent of the supplemental sample (10 percent of original sample).

C.8.3.5 2001 Quarter 1

Design: No changes.

Procedural: The 1999 quarter 4 NHSDA and 2000 quarters 1 through 3 NHSDA data from FI regions that correspond to the marijuana strata were used for all sampling parameters, including response rates and "Max of 1" adjustment. An 89 percent biological specimen response rate, which was obtained from field experience, was used.

Dwelling Unit Selection: In addition to the 100 percent sample, a supplemental 10 percent was selected.

Release 1: 100 percent of original sample.
Release 2: 100 percent of the supplemental sample (10 percent of original sample).

C.8.3.6 2001 Quarter 2

Design: No changes.

Procedural: The 2000 quarters 1 through 4 NHSDA data from FI regions that correspond to the marijuana strata were used for all sampling parameters, including response rates and "Max of 1" adjustment. An 89 percent biological specimen response rate, which was obtained from field experience, was used.

Dwelling Unit Selection: In addition to the 100 percent sample, a supplemental 10 percent was selected.

Release 1: 100 percent of original sample.
Release 2: 100 percent of the supplemental sample (10 percent of original sample).

C.8.3.7 2001 Quarter 3

Design: No changes.

Procedural: The 2000 quarters 2 through 4 NHSDA and 2001 quarter 1 NHSDA data from FI regions that correspond to the marijuana strata were used for all sampling parameters, including response rates and "Max of 1" adjustment. In addition, an adjustment was made to the interview response rate to compensate for a lower Validity Study interview response rate as compared with the NHSDA. This adjustment was based on 2000 data. An 89 percent biological specimen response rate, which was obtained from field experience, was used.

Dwelling Unit Selection: In addition to the 100 percent sample, a supplemental 15 percent was selected.

Release 1: 100 percent of original sample.
Release 2: 66 percent of supplemental sample (10 percent original sample).
Release 3: 33 percent of supplemental sample (5 percent of original sample).

C.8.3.8 2001 Quarter 4

Design: No changes.

Procedural: The 2000 quarters 3 and 4 NHSDA and 2001 quarters 1 and 2 NHSDA data from FI regions that correspond to the marijuana strata were used for all sampling parameters, including response rates and "Max of 1" adjustment. In addition, an adjustment was made to the interview response rate to compensate for a lower Validity Study interview response rate as compared with the NHSDA. This adjustment was based on 2000 data. An 89 percent biological specimen response rate, which was obtained from field experience, was used.

Dwelling Unit Selection: In addition to the 100 percent sample, a supplemental 20 percent was selected.

Release 1: 100 percent of original sample.
Release 2: 50 percent of supplemental sample (10 percent original sample).
Release 3: 50 percent of supplemental sample (10 percent of original sample).

C.9 Creation of Variance Estimation Strata

C.9.1 Validity Study

Because of the nature of the stratified clustered sampling design, it is essential that the design structure is taken into consideration when computing variances of survey estimates. The

Validity Study sample followed similar procedures for creating variance estimation strata that were used in the earlier NHSDAs. Adjacent design strata were collapsed into pairs to create pseudostrata with primarily two replicates each.

For the Validity Study supplemental sample, PSUs were grouped into sets based on their sequential order of selection. These variance strata were comprised of two or three FI regions that were selected consecutively in the selection algorithm. Each variance stratum should be identical in respect to the explicit stratification and similar with respect to the implicit stratification that was utilized in the PSU selection. As a result, each explicit stratum has unique pseudostrata. Each set of PSUs defined a pseudostratum with two or three replicates. More specifically within a variance stratum, the first FI region to be selected of the two or three replicates would be designated at the first variance replicate, and the next would be the second replicate. A third replicate occurred if the last variance stratum for the explicit stratum turned out to be comprised of only one replicate. In this case, the last FI region would be added as the third replicate to the previous variance stratum in the specific explicit stratum. This exception occurs because at least two replicates per variance stratum are required to compute a variance.

All weighted statistical analyses for which variance estimates are needed should use the pseudostrata and replicate identifying variables to identify nesting. Variance estimates can be computed by using clustered data analysis software packages, such as SUDAAN® (Shah et al., 1997). The SUDAAN® software package computes variance estimates for nonlinear statistics using procedures such as a first-order Taylor series approximation of the deviations of estimates from their expected values. The approximation is unbiased for sufficiently large samples.

C.9.2 NHSDA (Main Study)

To take advantage of the positive covariance between the Validity Study and the NHSDA estimates, special variance estimation strata and replicates were required for the main study. Because FI regions serve as strata for the NHSDA sample and as PSUs for the Validity Study, there was no direct way of capturing the covariance and using the entire NHSDA sample. A heuristic approach was followed in developing a new design structure that could be used to analyze both samples simultaneously. The design structure for the combined analyses was based on the Validity Study sample structure, which included 97 variance estimation strata with two variance estimation replicates and two variance estimation strata with three variance estimation replicates each as described above. The comparable main study sample (excluding Alaska and Hawaii) structure consisted of 876 variance estimation strata defined as FI regions. Two main study variance estimation replicates were defined within each FI region; each replicate consisted of four area segments (one from each quarterly sample). To facilitate a matching of the main study sample structure to the Validity Study sample structure, the following steps were carried out:

1. Validity Study strata with three replicates were partitioned randomly into two replicates (one with two FI regions and one with one FI region), resulting in 99 variance estimation strata with two variance estimation replicates each.

2. The 200 matching main study FI regions were assigned the same variance estimation stratum and replicate numbers (replicate 1 and replicate 2) as the amended Validity Study numbers (step 1).

3. The remaining 676 main study FI regions were associated with the Validity Study variance estimation strata based on their proximity to Validity Study strata in the circularly ordered list of FI regions within each Validity Study sample stratum. For the combined analysis, these 676 FI regions were assigned the variance estimation stratum number of the associated Validity Study variance estimation stratum.

4. In these 676 main study FI regions, all main study replicate 1 segments were assigned to the combined replicate 1, and all main study replicate 2 segments were assigned to the combined replicate 2.

Using these new design structure variables, the denominator degrees of freedom is 99. In contrast, for the main study sample, the corresponding degrees of freedom for the national estimates is 900.

C.10 Sample Weighting Procedures

The analysis weights for both the 2000 and 2001 Validity Study surveys were developed in a similar way. As in the main NHSDA, sampling weights were calibrated via the generalized exponential model (GEM) of Folsom and Singh (2000) for adjusting at various levels of screener dwelling unit (SDU) and person for nonresponse, poststratification, and possibly extreme weights. After the main sample in the Validity Study was calibrated, the subsample with the biological specimens also was calibrated by viewing it as a nonresponse adjustment problem so that the adjustment factor was constrained to be at least 1. In the GEM application for this adjustment, drug use variables (e.g., past month, past year, and lifetime use of alcohol, cigarettes, and marijuana) also were included in the model's set of predictor variables.

The calculation of the sampling weights was based on the stratified, four-stage design of the study. Specifically and for each year, the analysis weights comprised four stagewise sampling weight components, each equal to the inverse of the selection probability for that stage.

- Stage 1: Selection of field interviewer (FI) region.
- Stage 2: Selection of segment.
- Stage 3: Selection of dwelling unit (DU), which included four possible adjustments:
  - quarter segment adjustment;
  - subsegmentation inflation adjustment, a by-product of counting and listing;
  - added DU adjustment, which results from the half-open interval rule when subsampling is needed; and
  - DU percent release adjustment.
- Stage 4: Selection of person within a DU.

In addition to these eight weighting components, seven weighting adjustments were necessary to calculate the final analysis sample weight for the Validity Study respondents, and eight weighting adjustments were needed to calculate the final analysis weight for Validity Study respondents with biological specimens. All of these weight adjustments are in the form of nonresponse adjustment, poststratification, or extreme weight adjustment using the GEM modeling technique. These adjustments are listed below in the order in which they were implemented:

1. *Nonresponse adjustment at the DU level.* This adjustment was used to account for the failure to complete the within-DU roster.

2. *DU-level poststratification.* This adjustment involved using screener data of demographic information (e.g., age group, race, gender). DU weights were adjusted to the census population estimates derived from the 1990 census for various demographic domains. Explanatory variables used in the modeling consisted of counts of eligible persons with each DU that fell into the demographic categories. These counts, multiplied by the newly adjusted DU weight and summed across all DUs for various domains, added to the census population estimates. This adjustment was useful for providing more stable control totals for subsequent adjustments.

3. *Extreme weight adjustment at the DU level.* If it was determined that design-based weights (stages 1, 2, and 3) along with any of their respective adjustments resulted in an unsatisfactory unequal weighting effect (i.e., variance of the DU-level weights was too high, with high frequency of extreme weights), then the extreme weights were further adjusted. This adjustment was implemented by doing another weight calibration. The control totals were the DU-level poststratified weights, and the same explanatory variables as in DU-level poststratification were used so that the extreme weights were controlled and all the distributions in various demographic groups were preserved. This step was not implemented due to low extreme weight proportion after DU level poststratification adjustment.

4. *Selected person weight adjustment for poststratification to roster data.* This step used control totals derived from the DU roster that were already poststratified to the census population estimates. This adjustment assisted in bias reduction and improved precision by taking advantage of the properties of a two-phase design. Selected person sample weights (i.e., those adjusted at the DU level and that accounted for fourth-stage sampling) were adjusted to the DU weight sums of all eligible rostered persons. Any demographic information used in modeling was based solely on screener information because this was the only information available for all rostered persons.

5. *Person-level nonresponse adjustment.* This step allowed for the adjustment of weights resulting from the failure of selected sample persons to complete the interview. Respondent sample weights were adjusted to the weight of all selected persons. Again, demographic information used in modeling was based solely on screener information.

6. *Person-level poststratification.* This step adjusted the final person sample weights to the census population estimates derived from the 1990 census. These were the same outside control totals used in the second adjustment. However, demographic variables for this adjustment were based on questionnaire data, not on screener data as in adjustments 2, 4, and 5.

7. *Extreme weight treatment at the person level.* This adjustment served the same purpose as described in adjustment 3, except the weights reflected the fourth stage of selection. This step was not implemented due to low extreme weight proportion after person-level poststratification adjustment.

8. *Nonresponse adjustment for respondents with biological specimens.* This adjustment accounted for the failure to obtain biological specimens for respondents. In addition to the demographic variables, some drug use variables also were used in the modeling.

All weight adjustments were derived from GEM. Model variable selection at each adjustment was done using a combination method of forward and backward selection processes. The forward selection was used for the model enlargement, and within each enlargement, backward selection was used. The final adjusted weight, which was the product of weight components 1 through 15, was the analysis weight (VANALWT) used in the estimates for Validity Study respondents. The final analysis weight (BVANALWT) for respondents with biological specimens was the product of weight components 1 through 16. Exhibit C.1 presents all of the individual weight components for 2000 and 2001 Validity Study.

Table C.7 displays the weighted sum and the census population counts for some domains of interest for the respondents. Table C.8 displays the weighted sum and the census population counts for some domains of interest for respondents with biological specimens.

Weighted estimates based on the full sample of 4,465 respondents over the 2-year period were constructed using the weight VANALWT divided by two. Similarly, weighted estimates based on the 4,000 respondents providing biological specimens were constructed using the weight BVANALWT divided by two.

For more details about the 2000 and 2001 Validity Study sampling weight, please see Chen, Emrich, Gordek, and Singh (2006).

Exhibit C.1 Summary of the 2000 and 2001 Validity Study Sample Weight Components

Phase I Dwelling Unit–Level

#1 Inverse of Probability of Selecting Field Interviewer (FI) Region
#2 Inverse of Probability of Selecting Segment
#3 Quarter Segment Weight Adjustment
#4 Subsegmentation Inflation Adjustment
#5 Inverse of Probability of Selecting Dwelling Unit
#6 Added Dwelling Unit Adjustment
#7 Dwelling Unit Percentage Release Adjustment

#8 Dwelling Unit Nonresponse Adjustment ***(res.sdu.nr)****
#9 Dwelling Unit Poststratification Adjustment ***(res.sdu.ps)****
#10 Dwelling Unit Extreme Weight Adjustment ***(res.sdu.ev)***

Phase II Person Level

Design Weight Components
#11 Inverse of Probability of Selecting a Person within a Dwelling Unit

#12 Selected Person-Level Poststratification to Screener Data Controls ***(sel.per.ps)****
#13 Person-Level Nonresponse Adjustment ***(res.per.nr)****
#14 Person-Level Poststratification Adjustment ***(res.per.ps)****
#15 Person-Level Extreme Weight Adjustment ***(res.per.ev)***
#16 Respondents with Biological Specimens Nonresponse Adjustment ***(bio.per.nr)****

***** These adjustments use the generalized exponential model (GEM), which also involves pre- and postprocessing in addition to running the GEM macro.

Table C.7 2000 and 2001 Validity Study Weighted Sum and Census Population Estimates for VANALWT

Domain	2000 Validity Study			2001 Validity Study		
	n	Weighted Sum	Census Total	*n*	Weighted Sum	Census Total
Total	2,342	52,007,950	52,007,950	2,123	52,738,680	52,738,680
Quarter						
Quarter 1	715	12,929,584	12,929,584	571	13,103,344	13,103,344
Quarter 2	617	12,976,641	12,976,641	513	13,153,584	13,153,584
Quarter 3	472	13,029,007	13,029,007	542	13,211,927	13,211,927
Quarter 4	538	13,072,718	13,072,718	497	13,269,825	13,269,825
Age Group						
12-17	1,160	23,210,079	23,210,079	1,143	23,446,770	23,446,770
18-25	1,182	28,797,871	28,797,871	980	29,291,910	29,291,910
Race[1]						
White	1,801	41,508,449	41,508,449	1,663	42,038,846	42,038,846
Black	372	7,795,454	7,795,454	309	7,933,466	7,933,466
Other	169	2,704,047	2,704,047	151	2,766,368	2,766,368
Hispanicity						
Hispanic	360	7,618,325	7,618,325	313	7,861,569	7,861,569
Non-Hispanic	1,982	44,389,625	44,389,625	1,810	44,877,111	44,877,111
Gender						
Male	1,146	26,201,674	26,201,674	1,015	26,547,155	26,547,155
Female	1,196	25,806,276	25,806,276	1,108	26,191,525	26,191,525

[1]Note that the race domain includes both Hispanic and non-Hispanic.

Table C.8 2000 and 2001 Validity Study Weighted Sum and Census Population Estimates for BVANALWT

Domain	2000 Validity Study			2001 Validity Study		
	n	Weighted Sum	Census Total	*n*	Weighted Sum	Census Total
Total	2,095	52,007,950	52,007,950	1,905	52,738,680	52,738,680
Quarter						
Quarter 1	639	12,929,584	12,929,584	516	13,103,344	13,103,344
Quarter 2	548	12,976,641	12,976,641	448	13,153,584	13,153,584
Quarter 3	414	13,029,007	13,029,007	503	13,211,927	13,211,927
Quarter 4	494	13,072,718	13,072,718	438	13,269,825	13,269,825
Age Group						
12-17	1,034	23,210,079	23,210,079	1,032	23,446,770	23,446,770
18-25	1,061	28,797,871	28,797,871	873	29,291,910	29,291,910
Race[1]						
White	1,617	41,508,449	41,508,449	1,502	42,038,846	42,038,846
Black	335	7,795,454	7,795,454	273	7,933,466	7,933,466
Other	143	2,704,047	2,704,047	130	2,766,368	2,766,368
Hispanicity						
Hispanic	329	7,618,325	7,618,325	227	7,861,569	7,861,569
Non-Hispanic	1,766	44,389,625	44,389,625	1,628	4,4877,111	44,877,111
Gender						
Male	1,030	26,201,674	26,201,674	913	26,547,155	26,547,155
Female	1,065	25,806,276	25,806,276	992	26,191,525	26,191,525

[1]Note that the race domain includes both Hispanic and non-Hispanic.

Appendix D:
Establishing Drug Testing Cutoffs

Because the cutoffs established by the Substance Abuse and Mental Health Services Administration (SAMHSA) for Federal workplace drug testing are used in many drug testing programs, the appropriateness of using these cutoffs for validating self-reported drug use by youths aged 12 to 17 and young adults aged 18 to 25 in the general population was examined.

If the goal of drug testing is to deter drug use in a population, the cutoff must be set higher to minimize the possibility of false positives and to provide test results that are scientifically and legally defensible. Workplace drug testing is an example of this type of application, where a positive drug test may initiate adverse consequences for an employee or job applicant. If the goal of drug testing is to detect any presence of drug in the specimen, the lowest cutoff should be used. Testing in drug rehabilitation treatment and some clinical and postmortem drug testing are examples of drug testing applications requiring low cutoffs.

Any numerical value between an assay's lower limit of quantitation and upper limit of linearity may be used theoretically as an analytical cutoff for drug testing procedures. From a statistical point of view, the lower the cutoff concentration, the higher the probability that a positive result may not be reproducible analytically. Conversely, the higher the cutoff concentration, the higher the probability that a negative result may not reflect the true absence of a specific drug analyte. As cutoff concentrations increase above the lower limit of quantitation, the probability of encountering irreproducible positive drug test results decreases, while the probability of false negative drug test results increases. The choice of a cutoff depends on the specific application of the drug testing.

The Validity Study was designed to compare self-reported drug use within a defined period of time with drug test results. Therefore, the goal was to maximize the detection and identification of drug presence in the specimens. With this in mind, screening and confirmatory test cutoffs were set at concentrations that were low enough to maximize the detection of drug use within the timeframes of the survey questions. In fact, confirmatory test cutoffs were set at the lower limits of quantitation established by the testing laboratory for its gas chromatography/mass spectrometry (GC/MS) assays.

Commercial workplace drug testing laboratories routinely use cutoffs higher than those used in the Validity Study, often the initial and confirmatory drug test cutoffs established by SAMHSA for Federal workplace drug testing programs (SAMHSA, 2004a). The concern of using SAMHSA's cutoffs in the Validity Study was that the presence of drug may be missed (i.e., false negatives) within these defined periods of time, therefore defining that the lowest concentrations for detecting drug presence should be established.

As stated above, workplace drug testing is used to deter illegal drug use by employees or prospective employees. SAMHSA's *Mandatory Guidelines for Federal Workplace Drug Testing Programs* include strict testing requirements, including the established cutoffs, to support this goal while preventing any individual from being falsely accused of illegal drug use (SAMHSA,

2004a, 2004b). For example, scientific studies support that these cutoffs exceed drug analyte amounts that a donor could claim were due to unknowing passive exposure. Because some drugs subject to abuse are used commonly in medical treatment in the United States, SAMHSA also has implemented additional measures to distinguish illegal drug use from legitimate use. These measures include the requirement for a licensed physician trained as a medical review officer (MRO) to review drug test results and interview donors with positive tests in order to verify any alternative explanation for the positive test result. Because the Validity Study was used to compare self-reported drug use with drug test results, the results were not verified by an MRO.

The current mandatory guidelines specify relatively high cutoffs for codeine and morphine (i.e., 2,000 nanograms per milliliter [ng/mL] for both the initial and confirmatory tests) to reduce the number of positive opiate results from legitimate use of codeine and from ingestion of poppy seeds, which contain codeine and morphine. Specimens with a positive confirmatory test for morphine, concentrations $\geq$ 2,000 ng/mL, are subjected to confirmatory testing for 6-acetylmorphine (6-AM), a specific metabolite of heroin. The Validity Study did not test for this additional metabolite used to distinguish heroin use.

The Validity Study and SAMHSA urine cutoffs and analytes are shown in Tables D.1 and D.2.

Table D.1 Immunoassay Test Cutoffs for Urine

Urine Immunoassay Test	Validity Study Screening Test Cutoff (ng/mL)	SAMHSA Initial Test Cutoff (ng/mL)
Marijuana (Cannabinoids)	30	50
Cocaine	50	300
Opiates	50	2,000
Amphetamines	500	1,000
Cotinine[1]	100	N/A

ng/mL = nanograms per milliliter.

[1] Cotinine is not tested in Federal workplace drug testing programs.

Table D.2 Gas Chromatography/Mass Spectrometry (GC/MS) Test Cutoffs for Urine

Urine GC/MS Drug Test	Analyte(s)	Validity Study Confirmatory Test Cutoff (ng/mL)	SAMHSA Confirmatory Test Cutoff (ng/mL)
Marijuana	Delta-9-Tetrahydrocannabinol Carboxylic Acid (THCA)	2	15
Cocaine	Benzoylecgonine	5	150
Opiates	Codeine	5	2,000
	Morphine	5	2,000
	6-Acetylmorphine[1]	N/A	10
Amphetamines	Amphetamine	25	500
	Methamphetamine	25	500

ng/mL = nanograms per milliliter.

[1] In Federal workplace drug testing, any specimen confirmed positive for morphine is tested for the specific heroin metabolite, 6-acetylmorphine (6-AM). Testing for 6-AM was not done in the Validity Study.

Table D.3 compares self-reports with urine drug test results based on the Validity Study cutoffs. Table D.4 compares self-reports with urine drug test results based on the higher SAMHSA cutoffs for Federal workplace drug testing programs.

Table D.3 Comparison of Responses to Self-Reports Versus Urinalyses Based on Validity Study Cutoffs

Substance	Sample Size	Percent Negative Agreement[1]	Percent Positive Agreement[2]	Percent Under-reporting[3]	Percent Over-reporting[4]	Sensitivity[5]	Specificity[6]	Kappa[7]	Screening Cutoff (ng/mL)	Confirmation Cutoff (ng/mL)
Marijuana[8]									30	2
30-day	3,752	82.2	7.7	3.7	6.5	0.673	0.927	0.543		
7-day	3,748	86.0	6.5	4.8	2.7	0.579	0.969	0.594		
3-day	3,749	86.9	6.1	5.2	1.7	0.540	0.980	0.601		
Cocaine									50	5
30-day	3,753	97.7	0.3	1.1	0.9	0.229	0.991	0.241		
7-day	3,753	98.2	0.3	1.2	0.3	0.185	0.997	0.257		
3-day	3,753	98.4	0.2	1.2	0.1	0.163	0.999	0.255		
Opiates									50	5
30-day[9]	3,755	96.5	0.1	0.8	2.6	0.119	0.974	0.045		
7-day	3,754	97.9	0.0	0.8	1.2	0.034	0.988	0.018		
3-day[10]	3,754	98.4	0.0	0.8	0.8	0.034	0.992	0.027		
Amphetamines									500	25
30-day	3,750	98.0	0.1	0.9	1.1	0.102	0.989	0.083		
7-day	3,748	98.8	0.0	0.9	0.2	0.051	0.998	0.075		
3-day[10]	3,748	98.9	0.0	0.9	0.1	0.051	0.999	0.083		

NOTE: Estimates reported are based on all available self-reported data. Unless otherwise noted, 30-day use was based on positive responses to the core or repeat questions, 7-day use was based on positive responses to repeat questions only, and 3-day use was based on positive responses to the follow-up or repeat questions. Respondents with unknown information from the repeat or follow-up questions were excluded. ng/mL = nanograms per milliliter.

[1] Percent Negative Agreement = % Reporting no use and testing negative = 100a/(a+b+c+d).

[2] Percent Positive Agreement = % Reporting use and testing positive = 100d/(a+b+c+d).

[3] Percent Underreporting = % Reporting no use and testing positive = 100b/(a+b+c+d).

[4] Percent Overreporting = % Reporting use and testing negative = 100c/(a+b+c+d).

[5] Sensitivity = Proportion of positive self-reports among persons who test positive = d/(b+d).

[6] Specificity = Proportion of negative self-reports among persons who do not test positive = a/(a+c).

[7] Kappa = Measure of interrater agreement = $(P_o-P_e)/(1-P_e)$, where P_o = observed agreement and P_e = expected agreement.
1 = Perfect, $0.75 \leq k < 1$ = good, $0.4 \leq k < .75$ = moderate, $k < 0.4$ = poor.

[8] Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

[9] Based on responses to core questions only.

[10] Based on responses to repeat questions only.

In the two-by-two table below, a, b, c, and d are weighted cell counts.

		Urine Test Result -	Urine Test Result +	
Self-Report	-	a	b	a+b
	+	c	d	c+d
		a+c	b+d	a+b+c+d

Table D.4 Comparison of Responses to Self-Reports Versus Urinalyses Based on SAMHSA Workplace Cutoffs

Substance	Sample Size	Percent Negative Agreement[1]	Percent Positive Agreement[2]	Percent Under-reporting[3]	Percent Over-reporting[4]	Sensitivity[5]	Specificity[6]	Kappa[7]	Screening Cutoff (ng/mL)	Confirmation Cutoff (ng/mL)
Marijuana[8]									50	15
30-day	3,752	82.7	6.3	3.2	7.8	0.666	0.914	0.476		
7-day	3,748	86.8	5.5	3.9	3.8	0.584	0.958	0.545		
3-day	3,749	87.8	5.1	4.3	2.7	0.544	0.970	0.556		
Cocaine									300	150
30-day	3,753	98.1	0.3	0.7	0.9	0.269	0.991	0.236		
7-day	3,753	98.7	0.2	0.8	0.4	0.223	0.996	0.275		
3-day	3,753	98.8	0.2	0.8	0.2	0.191	0.998	0.273		
Opiates									2,000	2,000
30-day[9]	3,755	97.3	0.0	0.0	2.7	0.554	0.973	0.020		
7-day	3,754	98.7	0.0	0.0	1.2	0.554	0.988	0.044		
3-day[10]	3,754	99.2	0.0	0.0	0.8	0.554	0.992	0.067		
Amphetamines									1,000	500
30-day	3,750	98.2	0.1	0.7	1.1	0.126	0.989	0.093		
7-day	3,748	99.0	0.0	0.7	0.2	0.063	0.998	0.089		
3-day[10]	3,748	99.1	0.0	0.7	0.1	0.063	0.999	0.099		

NOTE: Estimates reported are based on all available self-reported data. Unless otherwise noted, 30-day use was based on positive responses to the core or repeat questions, 7-day use was based on positive responses to repeat questions only, and 3-day use was based on positive responses to the follow-up or repeat questions. Respondents with unknown information from the repeat or follow-up questions were excluded. ng/mL = nanograms per milliliter.

[1] Percent Negative Agreement = % Reporting no use and testing negative = 100a/(a+b+c+d).

[2] Percent Positive Agreement = % Reporting use and testing positive = 100d/(a+b+c+d).

[3] Percent Underreporting = % Reporting no use and testing positive = 100b/(a+b+c+d).

[4] Percent Overreporting = % Reporting use and testing negative = 100c/(a+b+c+d).

[5] Sensitivity = Proportion of positive self-reports among persons who test positive = d/(b+d).

[6] Specificity = Proportion of negative self-reports among persons who do not test positive = a/(a+c).

[7] Kappa = Measure of interrater agreement = $(P_o-P_e)/(1-P_e)$, where P_o = observed agreement and P_e = expected agreement.
1 = Perfect, $0.75 \leq k < 1$ = good, $0.4 \leq k < .75$ = moderate, $k < 0.4$ = poor.

[8] Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

[9] Based on responses to core questions only.

[10] Based on responses to repeat questions only.

In the two-by-two table below, a, b, c, and d are weighted cell counts.

		Urine Test Result -	Urine Test Result +	
Self-Report	-	a	b	a+b
	+	c	d	c+d
		a+c	b+d	a+b+c+d

Appendix E:
Technical Issues Concerning Drug Test Data: A Retrospective Analysis

Only the respondents who completed the Validity Study interview (i.e., 4,465 individuals) were given the opportunity to provide hair and urine specimens. Of those, 89.4 percent provided one or both biological specimens. This includes 80.5 percent who provided both a urine and hair specimen, 4.7 percent providing only urine specimens, and 4.3 percent providing only hair specimens (see Table F.2 in Appendix F).

The loss of urine specimens due to insufficient quantity was minimal. Among the urine specimens that were obtained, only 0.9 percent had insufficient quantity for the laboratory to conduct the urine screening tests for marijuana, cocaine, amphetamines, and opiates. Another 0.1 to 0.4 percent had insufficient quantity to conduct confirmatory testing for these drugs. For tobacco, 1.0 percent of specimens had insufficient quantity for screening tests, and there were no confirmation tests.

The Validity Study was designed to test the first 1.3 centimeters (cm) of hair (segment A) to approximate a 30-day time period. As noted previously, unlike the urine specimens, there were many hair specimens with insufficient quantity for testing. For those specimens with insufficient quantity, such that the first 1.3 cm of hair (segment A) could not be separated from the remaining hair specimen (segment B), the hair segment was tested in total (segment C). Over half of the hair specimens collected (53.6 percent) did not have a sufficient quantity of hair in the first 1.3 cm to conduct screening tests for drugs. Another 2.2 percent of the specimens had insufficient quantity for confirmatory testing, so that a total of 55.8 percent of the specimens could not be analyzed for drug use in the first 1.3 cm of hair. Overall, 13.7 percent of the hair specimens collected could not be tested for drugs.

Table E.1 shows the screening test cutoffs used for hair specimens in the Validity Study.

Tables E.2 through E.5 reflect a retrospective analysis of what has been learned in a detailed analysis of the specimen data. The numbers in these tables for urine specimens are based on the urine specimens tested positive using the Validity Study cutoffs listed in Tables D.1 and D.2. The numbers for hair specimens are based on the hair specimens screened positive using the Validity Study screening test cutoffs listed in Table E.1 and shown to contain the indicated analyte by confirmatory testing. For this retrospective analysis, specimens were identified as confirmed based on the presence of drug analyte, not based on the Validity Study confirmatory test cutoffs listed in Table B.5. This approach was used to maximize the number of hair specimens for which there was evidence that a drug was present. For example, the number of hair specimens confirmed to contain methamphetamine increased by 50 percent, and the number of hair specimens confirmed to contain amphetamine increased by 250 percent. The analysis of the specimen types, total specimens screened and confirmed by gas chromatography/mass spectrometry (GC/MS), and specimens with insufficient quantity are

presented in table format by drug class (e.g., marijuana metabolites, cocaine metabolites, amphetamines, opiates). The rates of drugs detected in hair compared with urine demonstrate differences in the usefulness of hair testing in the general population for most of the drug classes that were analyzed in this study.

Table E.1 Validity Study Hair Testing Cutoffs

Drug Class	Screening Test Cutoff (ng/mg)
Marijuana (Cannabinoids)	0.05 in Year 1 0.005 in Year 2[1]
Cocaine and Metabolites	0.5
Opiates	0.5 in Year 1 0.2 in Year 2[1]
Amphetamines	0.5

ng/mg = nanograms per milligram.

AMP = amphetamine; BZE = benzoylecgonine; COC = cocaine; COD = codeine; METH = methamphetamine; MOR = morphine; THCA = delta-9-tetrahydrocannabinol carboxylic acid.

[1] The marijuana (cannabinoids) and opiates screening test cutoffs were lowered in year 2 (2001).

Table E.2 Test Data for Marijuana Metabolites in Hair and Urine

Specimen Source	Analyte Identified by GC/MS Testing	Total Screened	Total Screened (Adjusted)[1]	Number Screened Positive	Number Screened Positive and Tested by GC/MS	Number Screened Positive and Not Tested by GC/MS	Number Confirmed Positive	Positivity Rate[2] (%)
Urine	THCA	3,775	3,763	450	438	12	420	11.16
Hair Segment A	THCA	1,768	1,421	69	22	47	4	0.23
Hair Segment B	THCA	1,499	1,469	93	63	30	7	0.48
Hair Segment C	THCA	1,473	1,416	106	49	57	7	0.49

GC/MS = gas chromatography/mass spectrometry; THCA = delta-9-tetrahydrocannabinol carboxylic acid.

Note: Hair segment A is no longer than 1.3 centimeters (cm) from the proximal end of the hair specimen. Hair segment B is greater than 1.3 cm from the proximal end but less than 6.5 cm in length. Hair segment C is hair segments A and B combined when there is insufficient quantity of hair to test in segment A.

[1] These values excluded specimens that screened positive but contained insufficient urine or hair for confirmatory testing with GC/MS.

[2] Positivity rate was calculated by dividing the number confirmed positive by the Total Screened (Adjusted)

Table E.3 Test Data for Cocaine Metabolites in Hair and Urine

Specimen Source	Analyte Identified by GC/MS Testing	Total Screened	Total Screened (Adjusted)[1]	Number Screened Positive	Number Screened Positive and Tested by GC/MS	Number Screened Positive and Not Tested by GC/MS	Number Confirmed Positive	Positivity Rate[2] (%)
Hair Segment A	COC	1,768	1,749	60	41	19	21	1.20
Hair Segment B	COC	1,499	1,477	60	38	22	24	1.62
Hair Segment C	COC	1,473	1,464	41	32	9	22	1.50
Urine	BZE	3,775	3,761	209	195	14	52	1.38
Hair Segment A	BZE	1,768	1,748	60	40	20	10	0.57
Hair Segment B	BZE	1,499	1,476	60	37	23	16	1.08
Hair Segment C	BZE	1,473	1,464	41	32	9	18	1.23

BZE = benzoylecgonine; COC = cocaine; GC/MS = gas chromatography/mass spectrometry.

Note: Hair segment A is no longer than 1.3 centimeters (cm) from the proximal end of the hair specimen. Hair segment B is greater than 1.3 cm from the proximal end but less than 6.5 cm in length. Hair segment C is hair segments A and B combined when there is insufficient quantity of hair to test in segment A.

[1] These values excluded specimens that screened positive but contained insufficient urine or hair for confirmatory testing with GC/MS.

[2] Positivity rate was calculated by dividing the Number Confirmed Positive by the Total Screened (Adjusted).

Table E.4 Test Data for Amphetamines in Hair and Urine

Specimen Source	Analyte Identified by GC/MS Testing	Total Screened	Total Screened (Adjusted)[1]	Number Screened Positive	Number Screened Positive and Tested by GC/MS	Number Screened Positive and Not Tested by GC/MS	Number Confirmed Positive	Positivity Rate[2] (%)
Urine[3]	AMP	3,775	3,762	256	243	13	42	1.12
Hair Segment A	AMP	1,768	1,754	69	55	14	3	0.17
Hair Segment B	AMP	1,499	1,488	79	68	11	5	0.34
Hair Segment C	AMP	1,473	1,463	74	64	10	2	0.14
Urine	METH	3,775	3,762	256	243	13	19	0.51
Hair Segment A	METH	1,768	1,754	69	55	14	9	0.51
Hair Segment B	METH	1,499	1,488	79	68	11	7	0.47
Hair Segment C	METH	1,473	1,463	74	64	10	2	0.14
Urine[3]	AMPS	3,775	3,762	256	243	13	42	1.12
Hair Segment A	AMPS	1,768	1,754	69	55	14	9	0.51
Hair Segment B	AMPS	1,499	1,488	79	68	11	8	0.54
Hair Segment C	AMPS	1,473	1,463	74	64	10	3	0.21

AMP = amphetamine; AMPS = amphetamine and/or methamphetamine; GC/MS = gas chromatography/mass spectrometry; METH = methamphetamine.

Note: Hair segment A is no longer than 1.3 centimeters (cm) from the proximal end of the hair specimen. Hair segment B is greater than 1.3 cm from the proximal end but less than 6.5 cm in length. Hair segment C is hair segments A and B combined when there is insufficient quantity of hair to test in segment A.

[1] These values excluded specimens that screened positive but contained insufficient urine or hair for confirmatory testing with GC/MS.

[2] Positivity rate was calculated by dividing the Number Confirmed Positive by the Total Screened (Adjusted).

[3] The use of amphetamine (i.e., Adderall®) by the youths and young adults aged 12 to 25 in this sample population may explain the presence of amphetamine in all positive AMPS specimens.

Table E.5 Test Data for Opiates in Hair and Urine

Specimen Source	Analyte Identified by GC/MS Testing	Total Screened	Total Screened (Adjusted)[1]	Number Screened Positive	Number Screened Positive and Tested by GC/MS	Number Screened Positive and Not Tested by GC/MS	Number Confirmed Positive	Positivity Rate[2] (%)
Urine	MOR	3,775	3,771	190	186	4	30	0.80
Hair Segment A	MOR	1,768	1,767	25	24	1	1	0.06
Hair Segment B	MOR	1,499	1,498	15	14	1	0	0.00
Hair Segment C	MOR	1,473	1,468	22	17	5	1	0.07
Urine	COD	3,775	3,771	190	186	4	8	0.21
Hair Segment A	COD	1,768	1,767	25	24	1	3	0.17
Hair Segment B	COD	1,499	1,498	15	14	1	2	0.13
Hair Segment C	COD	1,473	1,468	22	17	5	1	0.07
Urine	OPI	3,775	3,771	190	186	4	32	0.85
Hair Segment A	OPI	1,768	1,767	25	24	1	4	0.23
Hair Segment B	OPI	1,499	1,498	15	14	1	2	0.13
Hair Segment C	OPI	1,473	1,468	22	17	5	2	0.14

COD = codeine; GC/MS = gas chromatography/mass spectrometry; MOR = morphine; OPI = opiate.

Note: Hair segment A is no longer than 1.3 centimeters (cm) from the proximal end of the hair specimen. Hair segment B is greater than 1.3 cm from the proximal end but less than 6.5 cm in length. Hair segment C is hair segments A and B combined when there is insufficient quantity of hair to test in segment A.

[1] These values excluded specimens that screened positive but contained insufficient urine or hair for confirmatory testing with GC/MS.

[2] Positivity rate was calculated by dividing the Number Confirmed Positive by the Total Screened (Adjusted).

Table E.6 compares the drug positivity rates in urine and hair from Tables E.2 through E.5 in the context of the window of detection for each specimen matrix. For hair, 1.3 cm represents 1 month of growth. Thus, segment A represents a 0- to 30-day time period for drug use, and segment B represents a 31- to 180-day time period. Segment C represents a 0- to 180-day time period or longer, depending on total hair length (i.e., greater than 7.8 cm).

Table E.6 Comparison of Drug-Positive Rates in Urine and Hair: Percentages

	Urine	Hair[1]		
Drug Class	**0-7 Days[2]**	**Segment A, 0-30 Days[2]**	**Segment B, 31-180 Days[2]**	**Segment C, 0-180 Days[2]**
Marijuana	11.16	0.28	0.48	0.49
Cocaine[3]	1.38	1.20	1.62	1.50
Amphetamines	1.12	0.51	0.54	0.21
Opiates	0.85	0.23	0.13	0.14

[1] Populations contained hair of varying lengths.

[2] Indicated timeframes are the estimated minimum and maximum time periods represented by specimens in each population.

[3] Rates are based on the major cocaine analytes for the specimen types: benzoylecgonine (BZE) for urine and cocaine (COC) for hair.

Table E.6 clearly shows a profound difference in the marijuana positivity rate between hair and urine. The confirmed drug positivity rates for marijuana metabolite (delta-9-tetrahydrocannabinol carboxylic acid, THCA) in hair (i.e., 0.2 to 0.5 percent) are 20 to 50 times less than that of urine. This was likely due to the selection of the marijuana screening cutoff used for this study and the laboratory's methods at the time of the study. Over the 5 years since this study began, improved testing protocols and instrumentation sensitivity and methodology have been developed and implemented for marijuana detection in hair testing laboratories (Musshoff & Madea, 2006).

Cocaine detection in hair was nearly equivalent to that in urine. This was unexpected because the timeframe for detection of cocaine metabolite (i.e., benzoylecgonine) in urine is 0 to 7 days. Hair segments A, B, and C represented significantly longer time periods for the detection of cocaine and its metabolites (Table E.6). The lower than expected positivity rates in hair may be due to the large number of hair specimens that screened positive, but could not be confirmed due to insufficient quantity.

The positivity rates for amphetamines and opiates presented in Table E.6 suggest that the positivity rates in hair may be much lower in a general population of this age range (12 to 25 years) than has been shown in a number of studies that focus on criminal justice populations. This was likely due to two factors: (1) the age of the respondents in this study may be significant, in that youths aged 12 to 17 may not be included in criminal justice studies; and (2) the youth and young adult respondents aged 12 to 25 in this general population study may be more representative of occasional drug use, while the criminal justice population may be more representative of chronic drug use.

As shown in Tables E.2 through E.5, only half as many hair specimens had a sufficient amount for testing when compared with urine. This was likely due to two factors: (1) the collectors' misperception of the quantity of hair required for proper collection, and (2) the quantity of specimen required by the testing laboratory to perform both screening and confirmatory testing. This first factor is evidenced by the differences in the number of hair specimens screened and the number of urine specimens screened (i.e., the "Total Screened" column in Tables E.2 through E.5). The second factor is evidenced by the number of hair specimens screened positive that could not be tested by GC/MS due to insufficient quantity (i.e., the "Number Screened Positive and Not Tested by GC/MS" column).

The differences identified in these data may help to focus a future specialized survey to incorporate some aspects of drug testing that better identify the most appropriate and cost-effective testing approach.

It is easier to look backward and see what can be learned than to project forward with what is already known. This methodological study illustrates the complementary but not necessarily equivalent nature of the hair and urine drug tests. In future studies, there may be an opportunity to further explore and define observations from this methodological study. Observations from this study support other scientific studies suggesting that the incorporation and later detection of drugs and/or metabolites in hair are dependent on the drug class. With proper segmental analysis, sampling procedures to identify more drug users (particularly users of opiates and amphetamines), and the inclusion of additional survey questions on the use of specific opiates and stimulants, those characteristics hypothesized above (age and frequency of use) could be studied in more detail using hair specimens as compared with urine in certain populations of study participants.

Appendix F:
Detailed Tables

Table F.1 Summary of Validity Study Sample Sizes, by Data Collection Year and Quarter

	Year 2000				Year 2001				Total 2000 and 2001 Combined
	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Quarter 1	Quarter 2	Quarter 3	Quarter 4	
Number of Respondents	715	617	472	538	571	513	542	497	**4,465**[1]
Respondents Submitting Urine or Hair Sample	639	548	414	494	516	448	503	438	**4,000**
FINAL INTERVIEW CODE									
Total Selected Persons	**928**	**796**	**638**	**732**	**788**	**689**	**712**	**702**	**5,985**
Interview Completed	714	618	471	538	569	512	540	497	**4,459**[2]
No One at Housing Unit	17	18	15	17	22	21	23	30	**163**
Respondent Unavailable	47	43	38	39	41	28	36	48	**320**
Breakoff	8	0	1	3	4	1	2	2	**21**
Physically/Mentally Incompetent	4	2	9	5	6	3	3	11	**43**
Language Barrier									
Spanish	11	21	15	27	29	12	14	24	**153**
Other	2	1	1	4	3	1	0	5	**17**
Refusal	83	58	59	60	72	76	60	54	**522**
Parental Refusal	36	28	22	33	38	31	28	24	**240**
Other (Eligible)[3]	6	7	7	6	4	4	6	7	**47**

Note: The survey's completed case rule requires that respondents have reported lifetime use or nonuse for cigarettes and at least nine other drugs. For drug categories that are defined based on answers to multiple lead questions (hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives), lifetime use or nonuse of at least one lead question had to be reported in order for this requirement to be considered for that drug.

[1] Of the 4,465 respondents (i.e., those satisfying the completed case rule), 10 were breakoffs.

[2] Of the 4,459 interviews completed, 4 did not satisfy the completed case rule.

[3] Includes cases of denied access to buildings, PC problems, interviewed wrong roster member, and fraudulent cases (no reinterview).

.

Table F.2 Sample Size and Weighted Percentage of Respondents Providing Urine or Hair Specimens during the Interview

	Number	%
Total Respondents	**4,465**	**100.0**
Neither Specimen	465	10.6
Urine Only	188	4.7
Hair Only	190	4.3
Both Urine and Hair	3,622	80.5
Total Providing Urine	3,810	85.1
Total Not Providing Urine	655	14.9
Total Providing Hair	3,812	84.7
Total Not Providing Hair	653	15.3
Total Providing Urine, Hair, or Both	4,000	89.4

Table F.3 Sample Size and Weighted Percentage of Respondents Providing Urine or Hair Specimens during the Interview, by Demographic Characteristics

	Provided Urine Specimen		Provided Hair Specimen	
Demographic Characteristic	**Number**	**%**	**Number**	**%**
Total (*n* = 4,465)	3,810	85.1	3,812	84.7
Age				
12-17 Years (*n* = 2,303)	1,977	85.9	1,983	85.5
18-25 Years (*n* = 2,162)	1,833	84.5	1,829	84.1
Gender				
Male (*n* = 2,161)	1,873	86.6	1,801	82.9
Female (*n* = 2,304)	1,937	83.6	2,011	86.6
Race/Ethnicity				
White, Non-Hispanic (*n* = 2,866)	2,447	85.1	2,514	86.8
Black, Non-Hispanic (*n* = 603)	528	86.5	465	76.8
Other, Non-Hispanic (*n* = 323)	260	85.4	260	83.2
Hispanic (*n* = 673)	575	84.0	573	83.6
Population Density				
Large Metropolitan (*n* = 1,788)	1,486	83.2	1,468	81.9
Small Metropolitan (*n* = 1,659)	1,411	84.5	1,437	85.5
Nonmetropolitan (*n* = 1,018)	913	90.1	907	89.1
Region				
Northeast (*n* = 796)	656	82.4	670	84.2
Midwest (*n* = 1,112)	957	85.4	964	85.7
South (*n* = 1,617)	1,391	85.9	1,379	84.9
West (*n* = 940)	806	85.8	799	83.9
Education[1]				
Less Than High School (*n* = 454)	405	88.8	388	85.3
High School Graduate (*n* = 827)	703	85.5	712	85.4
Some College (*n* = 636)	533	82.5	534	82.7
College Graduate (*n* = 245)	192	79.0	195	80.8

[1]Education is reported only for persons 18 years old or older.

Table F.4 Number of Persons Interviewed in Validity Study (Unweighted *n*), by Age Group and Demographic Characteristics Compared with 2000-2001 NHSDA

Demographic Characteristic	12 to 17 Years		18 to 25 Years		Total 12 to 25 Years	
	Validity Study	NHSDA	Validity Study	NHSDA	Validity Study	NHSDA
Total	2,303	47,616	2,162	44,090	4,465	91,706
Gender						
Male	1,186	24,068	975	20,918	2,161	44,986
Female	1,117	23,548	1,187	23,172	2,304	46,720
Race/Ethnicity						
White, Non-Hispanic	1,510	32,012	1,356	29,468	2,866	61,480
Black, Non-Hispanic	284	6,393	319	5,517	603	11,910
Other, Non-Hispanic	159	2,567	164	2,549	323	5,116
Hispanic	350	6,644	323	6,556	673	13,200
Population Density						
Large Metropolitan	945	17,500	843	15,575	1,788	33,075
Small Metropolitan	832	16,795	827	16,535	1,659	33,330
Nonmetropolitan	526	13,321	492	11,980	1,018	25,301
Region						
Northeast	394	10,064	402	9,100	796	19,164
Midwest	594	13,036	518	12,468	1,112	25,504
South	811	14,828	806	13,943	1,617	28,771
West	504	9,688	436	8,579	940	18,267
Education[1]						
Less Than High School	--	--	454	9,080	454	9,080
High School Graduate	--	--	827	15,990	827	15,990
Some College	--	--	636	13,804	636	13,804
College Graduate	--	--	245	5,216	245	5,216
Marital Status[2]						
Married	2	84	361	8,317	363	8,401
Widowed	0	0	1	28	1	28
Divorced/Separated	0	3	56	1,066	56	1,069
Never Married	1,128	23,420	1,744	34,679	2,872	58,099

Note: Alaska and Hawaii are excluded.

-- Not applicable.

[1] Education is reported only for persons 18 years old or older.

[2] Marital status is available only for persons 15 years old or older.

Table F.5 Tobacco Product Use: Responses to Lifetime, Past Year, and Past Month Self-Report *Core* Questions from the Validity Study and from the 2000-2001 NHSDA, by Age Group: Percentages

Tobacco Product Use	Validity Study (n = 4,465)			NHSDA Main Study[1] (n = 91,706)		
	Lifetime	Past Year	Past Month	Lifetime	Past Year	Past Month
Any Tobacco Use[2]						
12-25 Years	57.6	41.1	32.5	57.1	39.8	30.9
12-17 Years	37.1	24.2	16.4	37.5	23.9	15.3
18-25 Years	74.1	54.6	45.5	72.9	52.6	43.4
Cigarettes						
12-25 Years	53.6	35.8	29.2	53.0	34.8	27.3
12-17 Years	33.8	20.5	14.5	34.1	20.5	13.2
18-25 Years	69.5	48.1	41.0	68.2	46.3	38.7
Smokeless Tobacco[3]						
12-25 Years	18.0	6.5	3.9	16.8	6.5	3.8
12-17 Years	10.1	5.0	2.4	8.4	4.4	2.1
18-25 Years	24.4	7.7	5.1	23.6	8.3	5.2
Chewing Tobacco						
12-25 Years	13.7	4.2	2.3	12.7	4.3	2.1
12-17 Years	7.6	3.2	1.4	6.2	3.0	1.3
18-25 Years	18.6	5.0	2.9	18.0	5.3	2.8
Snuff						
12-25 Years	12.5	5.1	3.2	11.4	4.8	3.0
12-17 Years	6.8	3.9	2.0	5.3	3.0	1.5
18-25 Years	17.1	6.0	4.1	16.2	6.2	4.1
Cigars						
12-25 Years	31.9	17.7	8.1	31.4	16.5	7.7
12-17 Years	16.7	9.9	4.1	16.8	10.1	4.4
18-25 Years	44.1	23.9	11.3	43.1	21.7	10.4
Pipes						
12-25 Years	6.2	NA	0.9	6.0	--	1.0
12-17 Years	3.4	NA	0.7	2.9	--	0.7
18-25 Years	8.5	NA	1.1	8.4	--	1.3

Note: *Core* questions asked about drug use and replicated the NHSDA format.

-- Not available.

[1] Alaska and Hawaii are excluded.

[2] Lifetime use and past month use include users who reported use of any tobacco product. Past year use includes those who reported use of cigarettes, chewing tobacco, snuff, or cigars in the past year (no data available for the use of pipes in the past year).

[3] Smokeless tobacco is defined as chewing tobacco or snuff.

Table F.6 Tobacco Use: Responses to Self-Report *Core* Questions on Recency of Any Tobacco Use among Those Tested for Cotinine in Their Urine: Percentages

Recency of Any Tobacco Use[1]	Positive for Cotinine in Urine (*n* = 1,085)	Negative for Cotinine in Urine (*n* = 2,674)
Past Month	80.4	13.6
More Than 1 Month, but in Past Year	3.7	10.4
More Than 1 Year, but in Lifetime	5.2	21.3
No Use	10.8	54.7

Note 1: *Core* questions asked about drug use and replicated the NHSDA format.

Note 2: Percentages are based on those who provided a valid urine specimen. A valid urine specimen has sufficient volume for laboratory testing as defined by the laboratory's procedures and as specified in the collection protocol.

Note 3: Respondents who used quitting smoking products in the past month were excluded.

Note 4: Cutoff for positive urine screen is greater than or equal to 100 nanograms per milliliter (ng/mL).

[1] Lifetime use and past month use include users who reported use of any tobacco product. Past year use includes users who reported use of cigarettes, chewing tobacco, snuff, or cigars in the past year (no data available for the use of pipes in the past year).

Table F.7 Tobacco Use: Responses to Self-Report *Core* Questions on the Number of Days Used in the Past 30 Days and Average and Median Cotinine Screening Concentrations for Various Tobacco Products among Past Month Tobacco Users Tested for Cotinine in Their Urine

	Cigarettes (*n* = 1,036)			Smokeless Tobacco[1] (*n* = 134)			Cigars (*n* = 282)		
Days Used	%	Average Cotinine Screening Concentration[2,3] (ng/mL)	Median Cotinine Screening Concentration[2,4] (ng/mL)	%	Average Cotinine Screening Concentration[2,3] (ng/mL)	Median Cotinine Screening Concentration[2,4] (ng/mL)	%	Average Cotinine Screening Concentration[2,3] (ng/mL)	Median Cotinine Screening Concentration[2,4] (ng/mL)
1 to 9 Days	30.4	1,363	34	50.6	2,705	824	85.6	2,631	466
10 to 19 Days	10.4	3,282	1,643	10.2	3,852	2,367	9.0	2,937	1,597
20 to 29 Days	9.6	3,414	1,903	12.2	4,791	3,032	3.2	2,428	2,427
Everyday	49.6	4,702	3,383	27.0	5,769	6,192	2.2	2,754	3

Note 1: *Core* questions asked about drug use and replicated the NHSDA format.

Note 2: Percentages are based on those who provided a valid urine specimen and reported past 30 day use of tobacco on the core questions. A valid urine specimen has sufficient volume for laboratory testing as defined by the laboratory's procedures and as specified in the collection protocol.

Note 3: Respondents who used quitting smoking products in the past month were excluded.

[1] Smokeless tobacco is defined as chewing tobacco or snuff. The number of days used is defined as the maximum number of days in the past month a person reported using either chewing tobacco or snuff.

[2] For the 2000 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 100,000 nanograms per milliliter (ng/mL). For the 2001 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 10,000 ng/mL. Thus, for analytical purposes, all concentrations in 2000 and data greater than 10,000 ng/mL were recoded and analyzed as 10,000 ng/mL.

[3] Average screening concentration is calculated by adding all cotinine screening results and dividing the sum by the number of respondents.

[4] Median concentration is the exact middle concentration in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.8 Tobacco Use: Responses to Self-Report *Core* Questions on Number of Cigarettes Smoked Per Day in the Past 30 Days and Average and Median Cotinine Screening Concentrations among Past Month Cigarette Users Tested for Cotinine in Their Urine

Number of Cigarettes Smoked Per Day in the Past 30 Days	Cigarettes (*n* = 1,031)		
	%	Average Cotinine Screening Concentration[1,2] (ng/mL)	Median Cotinine Screening Concentration[1,3] (ng/mL)
Less Than 1 Cigarette	14.3	1,067	30
1 Cigarette	12.7	1,958	123
2 - 5 Cigarettes	29.9	3,418	1,331
6 - 15 Cigarettes (about ½ Pack)	23.4	4,156	2,735
16 - 25 Cigarettes (about 1 Pack)	16.6	4,930	4,113
26 - 35 Cigarettes (about 1 ½ Packs)	2.7	5,928	5,362
More Than 35 Cigarettes (about 2 Packs or More)	0.5	7,202	10,000

Note 1: *Core* questions asked about drug use and replicate the NHSDA format.

Note 2: Percentages are based on those who provided a valid urine specimen and reported past 30 day use of cigarettes on the core questions. A valid urine specimen has sufficient volume for laboratory testing as defined by the laboratory's procedures and as specified in the collection protocol.

Note 3: Respondents who used quitting smoking products in the past month were excluded.

[1] For the 2000 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 100,000 nanograms per milliliter (ng/mL). For the 2001 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 10,000 ng/mL. Thus, for analytical purposes, all concentrations in 2000 and data greater than 10,000 ng/mL were recoded and analyzed as 10,000 ng/mL.

[2] Average screening concentration is calculated by adding all cotinine screening results and dividing the sum by the number of respondents.

[3] Median concentration is the exact middle concentration in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.9 Tobacco Use: Responses to 30-Day Self-Report *Follow-Up* Questions on Recency of Any Tobacco Use and Average and Median Cotinine Screening Concentrations among Past Month Tobacco Users Tested for Cotinine in Their Urine

Recency of Any Tobacco Use[1]	Positive for Cotinine in Urine (*n* = 741)			Negative for Cotinine in Urine (*n* = 260)		
		Screening Concentrations[2] (ng/mL)			Screening Concentrations[2] (ng/mL)	
	%	Average[3]	Median[4]	%	Average[3]	Median[4]
Past 3 Days	93.2	4,667	3,397	43.3	22	4
More Than 3 Days, but in Past 30 Days	6.8	3,752	2,367	56.7	11	3

Note 1: In the *core* portion of the study interview, respondents were asked about their past 30 day use of tobacco products. If they responded "yes" to using in the past 30 days, they were asked about their use in the past 3 days in the *follow-up*.

Note 2: Percentages are based on those who provided a valid urine specimen and reported past 30 day use of tobacco on the core questions. A valid urine specimen has sufficient volume for laboratory testing as defined by the laboratory's procedures and as specified in the collection protocol.

Note 3: Respondents who used quitting smoking products in the past month were excluded.

Note 4: Cutoff for positive urine screen is greater than or equal to 100 nanograms per milliliter (ng/mL).

[1] Any tobacco use includes those who reported use of any tobacco product in the past 30 days.

[2] For the 2000 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 100,000 ng/mL. For the 2001 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 10,000 ng/mL. Thus, for analytical purposes, all concentrations in 2000 and data greater than 10,000 ng/mL were recoded and analyzed as 10,000 ng/mL.

[3] Average screening concentration is calculated by adding all cotinine screening results and dividing the sum by the number of respondents.

[4] Median concentration is the exact middle concentration in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.10 Cigarette or Cigar Use: Responses to Self-Report *Repeat* Questions on Recency of Use: Percentages

Recency of Use	Type of Tobacco Use (*n* = 4,465)	
	Cigarettes	Cigars
Past 3 Days	22.3	2.9
More Than 3 Days, but in Past 7 Days	2.3	1.2
More Than 7 Days, but in Past 30 Days	3.0	3.1
More Than 30 Days, but in Past 6 Months	4.5	5.6
More Than 6 Months, but in Past Year	4.8	5.5
More Than Past Year, but in Lifetime	15.6	12.4
No Use	46.8	68.9
Don't Know/Refusal	0.7	0.5

Note: The *core* and *follow-up* portions of the study interview asked respondents to report their recency of use. Respondents then listened to one of two possible introductions to the next series of questions in which respondents again reported their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions were asked only about cigarette and cigar use.

Table F.11 Cigarette or Cigar Use: Comparison of Responses to Self-Report *Core* and *Repeat* Questions on Lifetime, Past Year, and Past Month Use, by Age Group: Percentages

	Age Group in Years					
	12 to 17			18 to 25		
	Lifetime	Past Year	Past Month	Lifetime	Past Year	Past Month
Core						
Cigarettes	33.8	20.5	14.5	69.5	48.1	41.0
Cigar	16.7	9.9	4.1	44.1	23.9	11.3
Cigarettes or Cigars	36.1	22.9	15.3	72.7	53.6	43.9
Total Repeat Questions						
Cigarettes	34.1	22.1	14.1	68.4	49.4	38.8
Cigar	16.1	10.6	3.5	42.7	24.6	10.2
Cigarettes or Cigars	36.1	23.9	14.8	71.0	54.3	41.2
Repeat Questions with Appeal						
Cigarettes	33.2	20.8	13.6	66.4	49.7	40.7
Cigar	16.3	10.6	2.9	41.9	24.3	10.4
Cigarettes or Cigars	36.0	23.3	14.3	68.8	54.3	43.1
Repeat Questions without Appeal						
Cigarettes	34.8	23.3	14.5	70.4	49.0	36.8
Cigar	16.0	10.6	3.9	43.5	24.9	10.0
Cigarettes or Cigars	36.1	24.5	15.2	73.3	54.4	39.3

Note: The *core* and *follow-up* portions of the study interview asked respondents to report their recency of use. Respondents then listened to one of two possible introductions to the next series of questions in which respondents were asked again to report their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions were asked only about cigarette and cigar use. One of the introduction scenarios very broadly introduced the next series of questions. This is defined as *repeat questions without appeal*. The second introduction scenario gave a broad overview of the study and emphasized the importance of the respondent's responses. An appeal was made to the respondent to answer the questions as honestly as he or she could. This is defined as *repeat questions with appeal.*

Table F.12 Tobacco Use: Responses to Self-Report *Repeat* Questions on Exposure to Passive Tobacco Smoke in the Past 6 Months and Average and Median Cotinine Screening Concentrations among Those Who Denied Use of Any Tobacco Product in the Past 30 Days but Tested Positive for Cotinine in Their Urine

Exposure to Passive Tobacco Smoke (*n* = 319)	Number of Respondents	Percent	Average Cotinine Screening Concentration[1,2] (ng/mL)	Median Cotinine Screening Concentration[1,3] (ng/mL)
Daily	155	45.0	2,827	977
Frequently, but Not Every Day	85	28.8	4,001	2,039
Seldom	56	17.5	3,517	1,300
Never	23	8.6	3,625	1,341

Note 1: The *core* and *follow-up* portions of the study interview asked respondents to report their recency of use. Respondents then listened to one of two possible introductions to the next series of questions in which respondents again reported their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions were asked only about cigarette and cigar use.

Note 2: In the 2000 Validity Study, respondents were asked only about their exposure to cigarette smoke. The 2001 Validity Study asked respondents about their exposure to smoke from cigarettes or any other tobacco product.

Note 3: Percentages are based on responses to *repeat* questions among those who denied past 30 day use of any tobacco product on the *repeat* questions but tested positive for cotinine in their urine.

Note 4: Respondents who used quitting smoking products in the past month were excluded.

Note 5: Cutoff for positive urine screen is greater than or equal to 100 nanograms per milliliter (ng/mL).

[1] For the 2000 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 100,000 ng/mL. For the 2001 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 10,000 ng/mL. Thus, for analytical purposes, all concentrations in 2000 and data greater than 10,000 ng/mL were recoded and analyzed as 10,000 ng/mL.

[2] Average screening concentration is calculated by adding all cotinine screening results and dividing the sum by the number of respondents.

[3] Median concentration is the exact middle concentration in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.13 Tobacco Use: Responses to Self-Report *Repeat* Questions on Exposure to Passive Tobacco Smoke in the Past 6 Months and Average and Median Cotinine Screening Concentrations among Those Who Denied Use of Any Tobacco Product in the Past 30 Days but Tested Positive for Cotinine in Their Urine, by Age Group

	12 to 17 Years			
Exposure to Passive Tobacco Smoke (*n* = 172)	**Number of Respondents**	**Percent**	**Average Cotinine Screening Concentration[1,2] (ng/mL)**	**Median Cotinine Screening Concentration[1,3] (ng/mL)**
Daily	92	49.2	2,490	899
Frequently, but Not Every Day	37	20.8	3,077	1,560
Seldom	28	18.8	3,472	646
Never	15	11.3	4,640	2,829
	18 to 25 Years			
Exposure to Passive Tobacco Smoke (*n* = 147)	**Number of Respondents**	**Percent**	**Average Cotinine Screening Concentration[1,2] (ng/mL)**	**Median Cotinine Screening Concentration[1,3] (ng/mL)**
Daily	63	41.3	3,186	1,132
Frequently, but Not Every Day	48	36.0	4,477	3,190
Seldom	28	16.4	3,562	2,636
Never	8	6.2	1,987	325

Note 1: The *core* and *follow-up* portions of the study interview asked respondents to report their recency of use. Respondents then listened to one of two possible introductions to the next series of questions in which respondents again reported their recency of use. This series of recency questions is defined as *repeat* questions. *Repeat* questions were asked only about cigarette and cigar use.

Note 2: In the 2000 Validity Study, respondents were asked only about their exposure to cigarette smoke. The 2001 Validity Study asked respondents about their exposure to smoke from cigarettes or any other tobacco product.

Note 3: Percentages are based on responses to *repeat* questions among those who denied past 30 day use but tested positive for cotinine in their urine.

Note 4: Cutoff for positive urine screen is greater than or equal to 100 nanograms per milliliter (ng/mL).

[1] For the 2000 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 100,000 ng/mL. For the 2001 Validity Study data, the laboratory did not report an exact concentration when the test result was greater than 10,000 ng/mL. Thus, for analytical purposes, all concentrations in 2000 and data greater than 10,000 ng/mL were recoded and analyzed as 10,000 ng/mL.

[2] Average screening concentration is calculated by adding all cotinine screening results and dividing the sum by the number of respondents.

[3] Median concentration is the exact middle concentration in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.14 Marijuana Use: Responses to Self-Report *Core* Questions on Recency of Use among Those Tested for Marijuana in Their Urine: Percentages

Recency of Marijuana Use	Positive for Marijuana in Urine (*n* = 418)	Negative for Marijuana in Urine (*n* = 3,342)
Past Month	60.9	6.5
More Than 1 Month, but in Past Year	9.6	8.7
More Than 1 Year, but in Lifetime	11.6	17.4
No Use	17.9	67.3

Note 1: *Core* questions asked about drug use and replicated the NHSDA format.

Note 2: Percentages are based on those who provided a valid urine specimen. A valid urine specimen has sufficient volume for laboratory testing as defined by the laboratory's procedures and as specified in the collection protocol.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: Metabolite concentration cutoff for a positive gas chromatography/mass spectrometry (GC/MS) confirmation urine test is greater than or equal to 2 nanograms per milliliter (ng/mL).

Table F.15 Marijuana Use: Responses to Self-Report *Core* Questions on the Number of Days Used in the Past 30 Days and Average and Median Marijuana Metabolite Concentrations among Past Month Marijuana Users Who Tested Positive for Marijuana in Their Urine

	Marijuana (*n* = 258)		
Days Used Marijuana	%	Average Concentration[1] (ng/mL)	Median Concentration[2] (ng/mL)
1 to 2 Days	11.8	78.0	85
3 to 5 Days	14.5	88.3	100
6 to 9 Days	9.8	75.6	84
10 to 19 Days	15.1	71.7	80
20 to 29 Days	26.0	89.1	96
All 30 Days	22.8	89.9	98

Note 1: *Core* questions asked about drug use and replicated the NHSDA format.

Note 2: Percentages are based on those who tested positive for marijuana in their urine and reported past 30 day use of marijuana on the *core* questions.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: Metabolite concentration cutoff for a positive gas chromatography/mass spectrometry (GC/MS) confirmation urine test is greater than or equal to 2 nanograms per milliliter (ng/mL).

[1] Average concentration is calculated by adding all screening concentrations and dividing the sum by the number of respondents.

[2] Median concentration is the exact middle of all screening concentrations in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.16 Marijuana Use: Responses to Self-Report *Follow-Up* Questions on Recency of Use and Average and Median Marijuana Metabolite Concentrations among Those Tested for Marijuana in Their Urine

Recency of Marijuana Use	Marijuana (*n* = 3,760)		
	%	Average Concentration[1] (ng/mL)	Median Concentration[2] (ng/mL)
Past 3 Days	7.1	69.3	86
More Than 3 Days, but in Past 30 Days	5.5	24.4	7
More Than 1 Month, but in Past 6 Months	4.7	16.7	2
More Than 6 Months, but in Past Year	2.1	10.4	3

Note 1: In the *core* portion of the study interview, respondents were asked about their past 30 day use of marijuana. If they responded positive to using in the past 30 days, they were asked about their use in the past 3 days on the *follow-up* questions. *Core* questions also asked respondents about their past 12 months use of marijuana. If they reported positive to using in the past 12 months, they were asked about their use in the past 6 months on the *follow-up* questions.

Note 2: Percentages are based on those who provided a valid urine specimen. A valid urine specimen has sufficient volume for laboratory testing as defined by the laboratory's procedures and as specified in the collection protocol.

Note 3: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 4: Metabolite concentration cutoff for a positive gas chromatography/mass spectrometry (GC/MS) confirmation urine test is greater than or equal to 2 nanograms per milliliter (ng/mL).

[1] Average concentration is calculated by adding all screening concentrations and dividing the sum by the number of respondents.

[2] Median concentration is the exact middle of all screening concentration in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.17 Marijuana Use: Responses to Self-Report *Follow-Up* Questions on Recency of Use and Average and Median Metabolite Concentrations among Those Who Tested Positive for Marijuana in Their Urine

Recency of Marijuana Use	Marijuana (*n* = 418)		
	%	Average Concentration[1] (ng/mL)	Median Concentration[2] (ng/mL)
Past 3 Days	49.0	84.5	95
More Than 3 Days, but in Past 30 Days	11.6	81.0	98
More Than 1 Month, but in Past 6 Months	5.8	97.1	101
More Than 6 Months, but in Past Year	1.1	98.9	101

Note 1: In the *Core* portion of the study interview, respondents were asked about their past 30 day use of marijuana. If they responded positive to using in the past 30 days, they were asked about their use in the past 3 days on the *follow-up* questions. *Core* questions also asked respondents about their past 12 months use of marijuana. If they reported positive to using in the past 12 months, they were asked about their use in the past 6 months on the *follow-up* questions.

Note 2: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 3: Metabolite concentration cutoff for a positive gas chromatography/mass spectrometry (GC/MS) confirmation urine test is greater than or equal to 2 nanograms per milliliter (ng/mL).

[1] Average concentration is calculated by adding all screening concentrations and dividing the sum by the number of respondents.

[2] Median concentration is the exact middle of all screening concentration in a distribution of ranked concentration or the concentration at the 50th percentile.

Table F.18 Marijuana Use: Responses to Self-Report *Follow-Up* Questions on Recency of Use among Those Who Tested Negative for Marijuana in Their Urine: Percentages

Recency of Marijuana Use	Marijuana (*n* = 3,342)
	%
Past 3 Days	1.7
More Than 3 Days, but in Past 30 Days	4.7

Note 1: In the *core* portion of the study interview, respondents were asked about their past 30 day use of marijuana. If they responded positive to using in the past 30 days, they were asked about their use in the past 3 days on the *follow-up* questions. *Core* questions also asked respondents about their past 12 months use of marijuana. If they responded positive to using in the past 12 months, they were asked about their use in the past 6 months on the *follow-up* questions.

Note 2: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 3: Metabolite concentration cutoff for a positive gas chromatography/mass spectrometry (GC/MS) confirmation urine test is greater than or equal to 2 nanograms per milliliter (ng/mL).

Table F.19 Marijuana Use: Responses to Self-Report *Repeat* Questions on Recency of Use: Percentages

Recency of Marijuana Use	Marijuana (*n* = 4,465)
Past 3 Days	6.9
More Than 3 Days, but in Past 7 Days	2.1
More Than 7 Days, but in Past 30 Days	3.8
More Than 30 Days, but in Past 6 Months	4.3
More Than 6 Months, but in Past Year	3.8
More Than Past Year, but in Lifetime	13.7
No Use	65.0
Don't Know/Refusal	0.5

Note: The *core* and *follow-up* portions of the study interview asked respondents to report their recency of use. Respondents then listened to one of two possible introductions to the next series of questions in which respondents again reported their recency of use. This series of recency questions is defined as *repeat* questions.

Table F.20 Marijuana Use: Comparison of Responses to *Core* or *Repeat* Questions on Lifetime, Past Year, and Past Month Use: Percentages

	Age Group in Years					
	12 to 17			18 to 25		
	Lifetime	Past Year	Past Month	Lifetime	Past Year	Past Month
Core	20.6	15.2	9.0	51.3	26.2	15.5
Total Repeat Questions	19.1	14.2	8.4	47.4	26.4	16.3
Repeat Questions with Appeal	18.7	13.7	8.6	47.4	24.9	15.9
Repeat Questions without Appeal	19.4	14.7	8.2	47.4	27.8	16.8

Note: The *core* and *follow-up* portions of the study interview asked respondents to report their recency of use of marijuana. Respondents then listened to one of two possible introductions to the next series of questions in which respondents were asked again to report their recency of use. This series of recency questions is defined as *repeat* questions. One scenario very broadly introduced the next series of questions. This is defined as *repeat questions without appeal*. The second scenario gave a broad overview of the study and emphasized the importance of the respondent's responses. An appeal was made to the respondent to answer the questions as honestly as he or she could. This is defined as *repeat questions with appeal*.

Table F.21 Marijuana Use: Responses to Self-Report *Repeat* Questions on Recency of Use and Average and Median Metabolite Concentrations among Those Who Tested Positive for Marijuana in Their Urine

Recency of Marijuana Use	Marijuana (n = 418)			
	Number of Respondents	% Positive	Average Concentration[1] (ng/mL)	Median Concentration[2] (ng/mL)
Past 3 Days	203	50.6	85.6	95
More Than 3 Days, but in Past 7 Days	34	6.7	81.2	96
More Than 7 Days, but in Past 30 Days	26	6.0	77.6	87
More Than 30 Days, but in Past 6 Months	18	4.4	95.5	100
More Than 6 Months, but in Past Year	23	4.6	80.9	94
More Than Past Year, but in Lifetime	34	8.1	85.9	98
No Use	76	18.7	84.8	95
Don't Know/Refusal	4	0.9	91.7	91

Note 1: The *core* and *follow-up* portions of the study interview asked respondents to report their recency of use. Respondents then listened to one of two possible introductions to the next series of questions in which respondents again reported their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Those who reported using Marinol® (i.e., dronabinol) in the past 30 days were excluded.

Note 3: Metabolite concentration cutoff for a positive gas chromatography/mass spectrometry (GC/MS) confirmatory urine test is greater than or equal to 2 nanograms per milliliter (ng/mL).

[1] Average concentration is calculated by adding all screening concentrations and dividing the sum by the number of respondents.

[2] Median concentration is the exact middle of all screening concentration in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.22 Marijuana Use: Responses to Self-Report *Repeat* Questions on Exposure to Passive Marijuana Smoke and Average and Median Metabolite Concentrations among Those Who Denied Use but Tested Positive for Marijuana in Their Urine

Exposure to Passive Marijuana Smoke	Denied Use in Past 3 Days (n = 211)			Denied Use in Past 7 Days (n = 177)			Denied Use in Past 30 Days (n = 151)		
		Concentration (ng/mL)			Concentration (ng/mL)			Concentration (ng/mL)	
	Percent	Average[1]	Median[2]	Percent	Average[1]	Median[2]	Percent	Average[1]	Median[2]
Daily	9.9	69.9	59	9.4	66.9	58	7.0	76.4	68
Frequently, but Not Everyday	25.7	81.7	94	22.6	86.2	95	19.2	84.8	94
Seldom	40.2	91.5	100	39.9	90.8	100	40.9	92.1	100
Never	24.3	80.8	93	28.2	80.8	93	32.9	80.8	93

Note 1: The *core* and *follow-up* portions of the study interview asked respondents to report their recency of use. Respondents then listened to one of two possible introductions to the next series of questions in which respondents again reported their recency of use. This series of recency questions is defined as *repeat* questions.

Note 2: Percentages are based on responses to *repeat* questions among those who denied use in specified time period but tested positive for marijuana in their urine.

Note 3: Metabolite concentration cutoff for a positive gas chromatography/mass spectrometry (GC/MS) urine test is greater than or equal to 2 nanograms per milliliter (ng/mL).

[1] Average concentration is calculated by adding all screening concentrations and dividing the sum by the number of respondents.

[2] Median concentration is the exact middle of all screening concentration in a distribution of ranked concentrations or the concentration at the 50th percentile.

Table F.23 Cocaine Use: Comparison of Responses to Lifetime, Past Year, and Past Month Self-Report *Core* Questions between the Validity Study and the 2000-2001 NHSDA: Percentages

Cocaine	Validity Study (*n* = 4,465)			NHSDA Main Study (*n* = 91,706)		
	Lifetime	Past Year	Past Month	Lifetime	Past Year	Past Month
Cocaine (Powder) Only						
12-25 Years	5.1	2.9	0.9	5.6	2.9	1.0
12-17 Years	1.8	1.1	0.4	1.7	1.2	0.4
18-25 Years	7.7	4.3	1.2	8.8	4.3	1.4
Crack						
12-25 Years	1.8	0.4[a]	0.0[b]	2.0	0.6	0.2
12-17 Years	0.5	0.2	0.0	0.6	0.4	0.1
18-25 Years	2.9	0.5	0.0[b]	3.1	0.8	0.2
Any Cocaine Use						
12-25 Years	6.9	3.3	0.9	7.6	3.5	1.1
12-17 Years	2.3	1.3	0.5	2.3	1.6	0.5
18-25 Years	10.7	4.9	1.3	11.9	5.1	1.6

Note 1: *Core* questions asked about drug use and replicated the NHSDA format.

Note 2: Percentages are based on those reporting use on *core* questions.

Note 3: Alaska and Hawaii are excluded.

Note 4: Cocaine (powder) only excludes those persons who also reported using crack.

[a] Difference between the 2000-2001 Validity Study and the 2000-2001 NHSDA is statistically significant at the 0.05 level.

[b] Difference between the 2000-2001 Validity Study and the 2000-2001 NHSDA is statistically significant at the 0.01 level.

Table F.24 Cocaine Use: Responses to Self-Report *Core* Questions on Lifetime, Past Year, and Past Month Use of Cocaine Products

Cocaine	Type of Cocaine Use			
	Cocaine		Crack	
	No.	%	No.	%
Lifetime Use	304	100	83	26.6
Past Year Use	144	100	22	11.5
Past 30 Day Use	48	100	3	3.7

Note 1: *Core* questions asked about drug use and replicated the NHSDA format.

Note 2: Percentages are based on those reporting use on *core* questions.

Note 3: All reporting crack use also reported cocaine use. Self-reported crack use is the percentage of those reporting cocaine use.

Table F.25 Opiate Use: Comparison of Responses to Lifetime, Past Year, and Past Month Self-Report *Core* Questions between the Validity Study and the 2000-2001 NHSDA: Percentages

Opiate	Validity Study (n = 4,465)			NHSDA Main Study (n = 91,706)		
	Lifetime	Past Year	Past Month	Lifetime	Past Year	Past Month
Heroin						
12-25 Years	0.8	0.4	0.2	1.0	0.3	0.1
12-17 Years	0.5	0.3	0.3	0.3	0.2	0.1
18-25 Years	1.0	0.5	0.2	1.5	0.4	0.2
Any Opiate Use[1]						
12-25 Years	12.5	7.4[a]	3.3[a]	13.3	7.5	2.9
12-17 Years	8.0	5.7[a]	2.6	9.0	6.0	2.5
18-25 Years	16.1	8.8	3.8	16.8	8.6	3.3

Note 1: *Core* questions asked about drug use and replicated the NHSDA format.

Note 2: Percentages are based on those reporting use on *core* questions.

Note 3: Alaska and Hawaii are excluded.

[a] Difference between the 2000-2001 Validity Study and the 2000-2001 NHSDA is statistically significant at the 0.05 level.

[1] Any opiate use consists of the use of heroin or the use of prescription pain relievers (including codeine and morphine).

Table F.26 Stimulant Use: Comparison of Responses to Lifetime, Past Year, and Past Month Self-Report *Core* Questions between the Validity Study and the 2000-2001 NHSDA: Percentages

Stimulant	Validity Study (*n* = 4,465)			NHSDA Main Study (*n* = 91,706)		
	Lifetime	Past Year	Past Month	Lifetime	Past Year	Past Month
Prescription Diet Pills[1]						
12-25 Years	1.5	--	--	1.6	--	--
12-17 Years	0.9	--	--	1.1	--	--
18-25 Years	2.0	--	--	2.0	--	--
Methamphetamines						
12-25 Years	3.1	1.0	0.2[b]	3.2	1.2	0.4
12-17 Years	1.5	0.8	0.2	1.4	0.8	0.2
18-25 Years	4.3	1.2	0.2[b]	4.6	1.4	0.5
Nonmedical Use of Any Prescription Stimulant[2]						
12-25 Years	6.9	2.6	1.0	6.5	2.6	0.9
12-17 Years	3.9	2.2	0.9	3.9	2.3	0.7
18-25 Years	9.4	2.9	1.0	8.6	2.9	1.0

-- Data not available.

Note 1: *Core* questions asked about drug use and replicated the NHSDA format.

Note 2: Percentages are based on those reporting use on *core* questions.

Note 3: Alaska and Hawaii are excluded.

[a] Difference between the 2000-2001 Validity Study and the 2000-2001 NHSDA is statistically significant at the 0.05 level.

[b] Difference between the 2000-2001 Validity Study and the 2000-2001 NHSDA is statistically significant at the 0.01 level.

[1] Respondents were asked about their use of prescription diet pills and were given the following as examples: amphetamines, Benzedrine®, Biphetamine®, Fastin®, or phentermine. However, respondents were not given an exhaustive list of examples of prescription diet pills.

[2] Nonmedical use of any prescription-type stimulant; does not include over-the-counter drugs.

Appendix G:
Key Definitions

Analytes
The term "analytes" is used to refer to those chemical substances (i.e., drugs and drug metabolites) that are specifically tested for by laboratory analysis.

Appeal
An "appeal" was developed to serve as an introduction to the repeat questions in the Validity Study questionnaire and was designed to increase a respondent's willingness to provide valid responses. Half of the respondents received this introduction that concerned the need to accurately measure the prevalence of drug use and the need to report truthfully as well as assurances of the confidentiality and anonymity of survey responses. The other half of respondents received a brief introduction asking them what they thought about the survey and announcing the repeat questions.

Biological Specimen Matrices
Biological specimen matrices are parts of an organism (e.g., hair, urine, plasma) that may serve as a storage place for substances (e.g., drugs, pollutants). In this study, the term refers to the biological material used as the drug testing specimen (i.e., urine, hair).

Cochran-Mantel-Haenszel (CMH) Chi-Square
The Cochran-Mantel-Haenszel (CMH) chi-square is a chi-square test of homogeneity or independence on stratified two-way tables and is based on observed minus expected values.

Confirmation Test
A confirmation or confirmatory test is used to identify samples from a screened population that possess a specified characteristic. In this study, a confirmation test was used as a second, independent test to identify specimens containing a drug analyte.

Core Questions
Core questions in the Validity Study questionnaire included introductory demographic questions (age, gender, race, ethnicity, marital status, educational attainment, overall health status, and personal and household income), as well as questions on the use of tobacco, alcohol, marijuana/hashish, cocaine, crack cocaine, heroin, hallucinogens, inhalants, and the nonmedical use of prescription pain relievers, tranquilizers, stimulants, and sedatives.

Debriefing Questions
Debriefing questions were asked of respondents at the end of the survey regarding the accuracy of their responses and how they thought "most people" would respond to the questions.

Degrees of Freedom
Degrees of freedom are a measure of the number of independent pieces of information on which a parameter estimate is based to determine how much precision an estimate of variability has.

The total degrees of freedom for an estimate equals the number of observations (values) minus the number of additional parameters estimated for that calculation.

Enzyme-Linked Immunosorbent Assay (ELISA)
ELISA is a type of immunoassay that uses an antibody bound to a substrate and an enzyme bound to either a second antibody or antigen to detect the presence of a specific drug or drug class.

Enzyme-Multiplied Immunoassay Technique (EMIT)
EMIT is a type of immunoassay that uses antibody binding and an enzyme reaction to detect the presence of a specific drug or drug class.

Fluorescence Polarization Immunoassay (FPIA)
FPIA is a type of immunoassay that uses antibody binding of fluorescein-tagged antigens and polarized vertical light to detect the presence of a specific drug or drug class.

Follow-Up Questions
Follow-up questions were asked of those reporting past month or past year use of cigarettes or cigars, marijuana, cocaine, heroin, prescription pain relievers (including opiates), and stimulants (including amphetamines). Those reporting past month use in the core were asked whether use occurred in the past 3 days. Those reporting past year use in the core were asked whether use occurred in the past 6 months. Those questions were followed by questions on perceptions of friends' drug use and parents' attitudes about respondents' drug use.

Gas Chromatography/Mass Spectrometry (GC/MS)
GC/MS is a test method that uses two analytical techniques to separate chemical mixtures and identify the components of the mixture with a very sensitive detector. The GC/MS instrument system includes a data collector (a computer component) to process the results of the test. GC/MS was used as the confirmation test for drugs in the biological specimens in this study.

Immunoassay
An immunoassay is a biochemical test that measures the concentration of a substance in a biological specimen using the reaction of an antibody or antibodies to its antigen. The test uses antibodies that react only with the particular drug or drug class for which the sample is being tested. The antibodies attach themselves to the drug if it is present in the sample.

Kappa
Kappa is a measure of the degree of agreement between two items, adjusting for any agreement occurring by chance. Perfect agreement is indicated by $\kappa = 1$ and chance agreement by $\kappa = 0$. A kappa value greater than 0.75 is considered "good." A kappa value of 0.40 to 0.75 indicates moderate agreement, and a kappa value less than 0.40 indicates poor agreement.

Metabolites
The metabolites of a substance are the intermediates and products resulting from the chemical and physical changes occurring in the body (i.e., metabolism).

_P_ Value
The probability value (p value) of a statistical hypothesis test is the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone, if the null hypothesis H_0 is true. It is equal to the significance level of the test for which we would only just reject the null hypothesis. The p value is compared with the desired significance level of our test, and the result is significant if it is smaller. That is, if the null hypothesis were to be rejected at the 5 percent significance level, this would be reported as "$p < 0.05$."

Percent Agreement
The "percent agreement" is the percentage of respondents with consistent statuses on self-reports and urine testing (i.e., positive self-reports and positive urine testing or negative self-reports and negative urine testing).

Percent Overreporting
The "percent overreporting" is the percentage of respondents who self-reported use but tested negative.

Percent Underreporting
The "percent underreporting" is the percentage of respondents who self-reported no use but tested positive.

Pharmacokinetics
Pharmacokinetics refers to the characteristic interactions of a drug and the body in terms of its absorption, distribution, metabolism, and excretion.

Recency of Use
Recency refers to the period of time associated with a particular estimate. The recency of use question for each drug was essentially the same for all classes of drugs. The question was "How long has it been since you last used [drug name]?" The recency question was the source for the past month, past 7 day, and past 3 day estimates provided in this report.

Repeat Questions
Repeat questions were asked of respondents at the end of the survey about their use of cigarettes and cigars, marijuana, cocaine, stimulants, and opiates in the past 6 months, past 7 days, and past 3 days. Questions about passive exposure from drugs smoked by other people during the past 6 months also were included.

Screening Test
A screening test quickly and efficiently identifies samples that may possess a specified characteristic. In this study, the screening test was used to distinguish negative specimens from those that could contain drugs.

Sensitivity
Sensitivity is the proportion of positive self-reports (i.e., those reporting drug use) among those with a positive drug test.

Specificity
Specificity is the proportion of negative self-reports (i.e., those reporting no use) among those with a negative drug test.

Urinalysis
Urinalysis is a physical and/or chemical examination of the urine to identify a specific property of the urine or a substance contained in the urine. In this study, the term refers to the drug tests.

Valid Hair Specimen
A valid hair specimen has sufficient weight and length for laboratory testing as defined by the laboratory's procedures and as specified in the collection protocol. In this study, the minimum weight required for testing was 16 milligrams (mg), and the minimum length was 1 centimeter (cm). The collection protocol required a specimen about the area of a pencil eraser, at least 1 cm long. After the first quarter of the study, the specifications were changed to two specimens of this size.

Valid Urine Specimen
A valid urine specimen has sufficient volume for laboratory testing as defined by the laboratory's procedures and as specified in the collection protocol. In this study, the collection protocol specified 30 milliliter (mL) to be collected. The laboratory required 4 mL total to conduct screening tests for cotinine and drugs and an additional 2 to 7 mL for each confirmatory drug test.