National Institute on Drug Abuse

RESEARCH

MONOGRAPH SERIES

The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates

167

U.S. Department of Health and Human Services ○ National Institutes of Health

The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates

Editors:

Lana Harrison, Ph.D.

Arthur Hughes, M.S.

NIDA Research Monograph 167
1997

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

National Institute on Drug Abuse
Division of Epidemiology and Prevention Research
5600 Fishers Lane
Rockville, MD 20857

ACKNOWLEDGMENT

This monograph is based on the papers from a technical review on "The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates" held on September 8-9, 1994. The review meeting was sponsored by the National Institute on Drug Abuse.

COPYRIGHT STATUS

National Institute on Drug Abuse
NIH Publication No. 97-4147
Printed April 1997

NIDA Research Monographs are indexed in the *Index Medicus*. They are selectively included in the coverage of *American Statistics Index, BioSciences Information Service, Chemical Abstracts, Current Contents, Psychological Abstracts,* and *Psychopharmacology Abstracts.*

167339

Contents

Introduction—The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates

Lana Harrison and Arthur Hughes

ABSTRACT

Measuring levels and patterns of illicit drug use, their correlates, and related behaviors requires the use of self-report methods. However, the validity of self-reported data on sensitive and highly stigmatized behaviors such as drug use has been questioned. The goal of this monograph is to review current and cutting-edge research on the validity of self-reported drug use and to describe methodological advances designed to reduce total error in estimates of drug use and quantify sources of nonsampling error.

This monograph reviews a number of studies that use some presumably more accurate measure of drug use to validate self-reported use. In addition, evolving methods to improve a wide variety of procedures used in survey designs are explored, including computer-assisted interviewing, predictors of response propensity, measurement error models, and improved prevalence estimation techniques. Experimental manipulations of various survey conditions and situational factors also show promise in improving the validity of drug prevalence estimates in self-report surveys.

FOREWORD

The monograph arises from a technical review that was conducted on September 8 and 9, 1994 in Gaithersburg, MD, where papers were presented by 25 leading U.S. researchers on various aspects pertaining to the validity of self-reported drug use. The focus of the technical review was to examine recent research on validity using internal or external criteria, especially bioassays, as well as to examine methodological advances that can contribute to improved estimates of drug use in a survey environment. This monograph includes 20 of the 25 papers

presented. The loss of several papers addressing the validity of the biological assays to assess drug use, particularly using hair as the medium, are of particular concern. (Please refer to the Technical Note at the end of this Introduction.)

The Technical Review was broad based with two fairly distinct focuses. Hence, the first area of this monograph is an overview of what is known about the validity of self-report based on studies using internal and external validity criteria. Other chapters consider the importance of recanting earlier reports of drug use on longitudinal surveys and how ethnographic research methods may improve validity. The monograph includes overview chapters on several studies that attempt to determine the accuracy of self-reported drug use among criminal justice, treatment, and workplace populations by using urinalysis and/or hair analysis to validate recent drug use. Also included is a review article on the validity of biological assays to determine how accurately drug use is reported.

The second focus is on methodological advances that have been used or proposed as a means for understanding the extent of nonsampling error in surveys, and realizing further reductions in total error in estimates of drug use and associated behaviors. One promising method is the use of audio computer-assisted self-interviewing, which can allow complicated branching of questions to occur while permitting respondents with reading difficulties to complete the interview with minimum interviewer intervention. Other chapters deal with correlates of response propensity, cognitive laboratory procedures, privacy effects, sampling methods, measurement error models, and improved estimators of hardcore drug use. Overall, each chapter in this part of the monograph demonstrates where improvements can be achieved in the design of surveys collecting sensitive information, and how estimates of these behaviors can be improved.

As previously mentioned, the editors are concerned about the loss of several papers from the Technical Review that addressed the developing science of hair testing for drugs of abuse. These papers were based on the research of laboratory scientists and provided cautions about the state of the science with respect to the validity and reliability of methods to identify drugs in hair. The reader is referred to *Hair Testing for Drugs of Abuse: International Research on Standards and Technology*, edited by E. Cone, M. Welch, and M. Babecki, NIH Pub. No. 95-3727 (1995) for similar papers. The Technical Review and monograph include several papers detailing results of studies comparing drug use prevalence

based on self-report, urinalysis, and hair testing measures. Since hair analysis is still a developing science with unresolved issues, the results from these studies must be viewed with caution in light of the limitations of hair testing technology. (Please refer to the Technical Note at the end of this Introduction.)

Following is a review of the chapters in the order that they appear in the monograph.

MONOGRAPH OVERVIEW

Validity Studies

Harrison examines the research literature on validation studies to provide an overview of what is known about the accuracy of self-reported drug use. Before the mid-1980s, validation studies suggested that drug use was fairly accurately reported in self-report surveys. However, recent validation studies conducted with criminal justice and former treatment clients using improved urinalysis techniques and hair analyses suggest only about half or less of recent drug use is self-reported in confidential interviews. While this research has been used to criticize estimates of drug use generated from self-report surveys, there are limitations with the testing technology, as well as with the validity studies conducted to date. Harrison discusses these limitations, particularly with respect to urinalysis and the developing science of hair testing for drugs of abuse, but points out there is an accumulating body of research evidence that leads to some general conclusions about self-report. That is, self-report is less valid both for the more stigmatized drugs such as cocaine and for more recent rather than distant use. Self-report methods where respondents do not answer aloud increase reports of drug use. Also, the validity of self-report tends to be least reliable for those involved with the criminal justice system.

Former treatment clients, particularly narcotic users, have been the focus of much research on the validity of self-reported drug use. Harrell reports on a validity study conducted in 1985 in conjunction with the National Household Survey on Drug Abuse (NHSDA) sponsored by NIDA. NIDA chose not to publicly release the study because of an unacceptably low response rate. However, while the results must be viewed cautiously, they suggest variations in reporting by drug type, with the percentage of known users reporting their use highest for

marijuana, followed by cocaine and hallucinogens, and lowest for heroin. The pattern of inconsistent reporting was consistent with the social desirability hypothesis, with most admitting use of less stigmatized drugs but fewer admitting use of more stigmatized drugs.

In their chapter, Johnston and O'Malley examine the recanting of earlier reported lifetime use of several drugs from the Monitoring the Future study. Recanting rates are examined on nationally representative samples of high school seniors (18-year-olds) from the late 1970s as they are followed through age 32. Recanting rates were quite modest for the illegal drugs examined—marijuana, cocaine, and lysergic acid diethylamide (LSD)—but for the psychotherapeutic drugs examined (tranquilizers and barbiturates) they were more substantial. The larger differences for the psychotherapeutic drugs may be attributable to young adults correcting for earlier inaccurate reports of psychotherapeutic drug use due to difficulties in identifying the drugs. Consistent with earlier research, minorities—particularly African Americans—had somewhat higher rates of recanting on the illegal drugs. So did respondents in certain occupations, specifically the military and police/firefighting. In general, however, the evidence is quite good for validity of self-reported lifetime use of the illegal drugs gathered by mail in young adulthood.

The next chapter in the monograph presents an innovative approach to determining the reliability of self-reported drug use and drug dealing using both retrospective and prospective methods. Respondents were given a life history interview focusing on drug use history, involvement in drug sales, criminal history, violence history, and treatment history. They were also interviewed in detail about their activities over the past 7 days. Over the following 7 weeks, respondents were asked to report on activities in the preceding week. In general, Fendrich, Mackesy-Amiti, and Goldstein found the life history reports of current use for heroin, cocaine, marijuana, and alcohol were consistent with reports provided prospectively. However, subjects reported considerably higher use quantities and frequencies for substances in the life history reports than they did in the weekly interview reports. They also tended to underreport their alcohol use in the life history interviews compared to the weekly prospective interviews, suggesting they tended to minimize the importance of their alcohol use. However, with respect to heroin and cocaine, the phenomenon of overreporting was observed with respondents overestimating the volume and cost of cocaine and heroin they used in the life history interviews. On the other hand, nearly 20 percent of those reporting drug dealing in the weekly prospective interviews failed to

report drug dealing in the life history interviews. Preliminary inspection of the data suggests that some of the discrepancy in drug-dealing reports may be the result of discrepant definitions of dealing. This is especially applicable to low-level or sporadic dealers who, during the weekly prospective interviews, reported occasionally selling small quantities of drugs.

The next chapter, by Cone, assesses the strengths and limitations of biological assays to validate self-report. Over the past several decades, technologically sophisticated methods have been developed for analyzing drug metabolites in bodily fluids and tissues such as urine, blood, hair, saliva, semen, meconium, and perspiration. Each medium has advantages and disadvantages, and ongoing research is helping to further refine the tests. Drugs or their metabolites can generally be detected in urine for 2 to 4 days, although most illicit drugs are eliminated within 48 hours after use. Saliva offers advantages over urine, including a higher concentration of the parent drug than metabolites and a closer ratio to blood concentrations, but the window of detection is generally only 12 to 24 hours. Cone states this makes saliva most useful for the detection of recent drug use in accident victims, or testing employees before they engage in safety-sensitive activities. Research on sweat testing has been limited because of the difficulty of collecting sweat samples, but a sweat-collection device that is applied to the skin and worn for a period of several days to several weeks appears to have solved some of the collection problems and made sweat testing more feasible. The science of hair testing for drugs of abuse has improved in recent years, but there are still many unresolved issues. (Please refer to the Technical Note at the end of this Introduction.) Hair offers the potential for detecting drug use over much longer periods of time, which is very appealing; it can be easily stored and is less embarrassing to collect. Cone concludes that validation of self-report data by drug testing must be performed with careful consideration of the limitations imposed by the testing methodology and the biological specimen.

Preston, Silver, Schuster, and Cone discuss the innovative use of urinalysis to monitor treatment compliance in clinical trials. They report on their study of 37 patients who used cocaine consistently during the first 5 weeks of methadone treatment. Three days each week, subjects answered self-report questionnaires and submitted urine samples. Over the course of the 17-week clinical trial, subjects reported cocaine use on 20 percent of occasions, but tested positive for cocaine (qualitatively) on 68 percent of occasions. However, examination of the quantitative data

reveals that at least part of the differential rates of self-report and qualitative cocaine-positive urine specimen was due to carryover. A urine specimen collected several days after self-administration of a large amount of drug could have the same drug/metabolite concentration as a specimen collected just after self-administration of a small amount of drug. Concentrations of benzoylecgonine—a metabolite of cocaine—in urine specimens supported the suggestion that rates of drug use as determined by qualitative urinalysis were artificially high due to carryover. Preston and colleagues suggest that the effectiveness of substance abuse treatment programs can be monitored by frequently conducted urinalyses.

Miller, Donnelly, and Martz report on the forensic use of testing hair for drugs of abuse at the Federal Bureau of Investigation (FBI). Hair testing is only used by the FBI when other information exists that indicates drug use and the results can remove a person from suspicion or associate them with criminal activity. The detection of cocaine has been the FBI's first priority in hair testing for drugs of abuse because of its prevalence. Although the FBI does not routinely engage in testing hair for drugs of abuse, the chapter presents synopses of several cases where hair testing was used. Further, analysis of more than 100 samples was performed on hair obtained from a medical examiner's random autopsy collection. The results of the hair testing for drugs of abuse were found to be consistent with autopsy toxicology reports. Miller and colleagues conclude hair testing can be used in conjunction with urinalysis to give a more detailed drug history on a test subject.

In their chapter, Mieczkowski and Newel report on a study in which they compared urine and hair testing results among a population of Florida probationers. These probationers were already undergoing regular urinalysis, and were asked to participate in a confidential 6-month study that would also collect monthly hair samples. Of the 89 cases who had 6 complete sets of specimens, 36 were negative on all assays for all drugs, which was the most frequent finding. Focusing on the disconcordant hair and urine assay cases, the authors show that most of the disconcordant results were for cannabis and opiates. Mieczkowski and Newel have previously stated that hair testing is probably the best developed for cocaine, and their analysis helps to support their conclusion. They further posit that environmental contamination is not an unresolvable clinical problem for hair analysis of cocaine, provided one is willing to accept that marginal cocaine use, because of high cutoff values, may be classified as passive contamination. However, other research and

researchers would disagree with their assertion this is resolvable with the current state of technology (NIDA 1995). Mieczkowski and Newel caution that bioassays can create a false sense of certainty about the meaning and utility of biological testing of any kind. They suggest that hair assays should be used when the outcome cannot put the person undergoing the testing in jeopardy, and may be especially useful in epidemiological surveys.

Wish, Hoffman, and Nemes provide an overview of the research literature on the validity of self-report, making the point that as drug use became more stigmatized during the years of the War on Drugs, individuals may have become less willing to disclose past drug use. The research literature is replete with studies showing that individuals under criminal justice supervision are loath to report drug use on confidential and anonymous surveys. However, Wish and colleagues also suggest there is reason to question the validity of self-report among treatment clients—another group that has frequently been the focus of validity studies. Results are presented from a study of clients participating in the Washington, DC, Treatment Initiative study who were assessed for drug use by interview, urinalysis, and hair analysis. At intake, almost all clients who tested positive had reported their use of heroin (96 percent), but fewer clients had reported their cocaine use (82 percent). A subsample was followed posttreatment. Although information is not presented for urinalysis results, 62 percent tested hair-positive for opiates and 36 percent self-reported use, while 80 percent tested hair-positive for cocaine, and 52 percent self-reported their use in the past 90 days. One interesting finding was a strong association between the self-reported frequency of drug use and concentration of drugs found in the hair. Although this study can only be viewed as suggestive due to the limitations of hair testing technology and the small number of followup cases, Wish and colleagues assert that treatment evaluation studies that fail to validate their estimates of self-reported drug use should be interpreted with considerable caution. Clients may wish to show that the treatment they had participated in had some value.

In their chapter, Magura and Kang report the results of two validity studies conducted by the first author, one for a sample of patients in two methadone treatment programs in New York City and the other for a sample of criminally involved young adults. Self-report information and both urine and hair samples were obtained on all the clients. For the methadone sample, 60 percent self-reported recent cocaine use and

80 percent were hair positive. For the young adult sample, 23 percent self-reported recent cocaine use, but 67 percent were hair positive. Magura and Kang discount the sensitivity hypothesis because 75 percent reported lifetime drug dealing (41 percent in the past month). The curious finding, then, is the lower reports of cocaine use. Magura and Kang suggest that for the young adults, use of cocaine—or more specifically, crack—had become stigmatized, even though dealing of these drugs was not. However, there may be important explanations overlooked by the authors, based on the limitations of hair testing technology including issues of racial bias and passive contamination (see the Technical Note at the end of this Introduction).

Cook, Bernstein, and Andrews report on a study employing self-report, urinalysis, and hair analysis in a workplace sample. They selected a random sample of 1,200 employees of a steel plant in the western United States. Employees were randomly assigned to four different self-report methods of assessing illicit drug use: (1) individual interview in the workplace, (2) group-administered questionnaire in the workplace, (3) telephone interview, and (4) individual interview off the worksite. The group-administered questionnaire method produced prevalence rates that were roughly half those of the other self-report methods. However, perhaps surprisingly, Cook and colleagues found that self-reports produced higher prevalence rates than either urinalysis or hair analysis. For the entire sample, only 7.8 percent tested positive for any drug by urinalysis, while 9.4 percent reported recent drug use. For the subsample that had hair tests, 6.2 percent were positive for an illicit drug and 9.9 percent reported recent use. Nevertheless, Cook and colleagues found only about half of those positive for any drug on either test self-reported recent use. The authors concluded that the findings suggest the need for multiple assessment methods of estimating self-report. However, since most of those who tested positive by hair analysis were positive for marijuana, and hair analysis has been shown to be least reliable for detecting marijuana use, the need for multiple assessment methods does not appear a justifiable conclusion. In fact, the study results demonstrate that self-reports produced higher prevalence rates than either urinalysis or hair analysis.

Methodological Developments

Nonresponse error continues to be pervasive in surveys soliciting either sensitive or nonsensitive information. While surveys such as NHSDA and Monitoring the Future typically achieve response rates from the

upper 70s to mid 80s, little is known about what impact the nonrespondents (from 15 percent for high school seniors in Monitoring the Future to about 22 to 23 percent in NHSDA) have on estimates of drug use and other deviant behaviors. To gain a better understanding of nonresponse error in the NHSDA, Gfroerer, Lessler, and Parsley present results of the Census Match Study, a program where responding and nonresponding NHSDA households sampled in 1990 were matched to data from the 1990 Decennial Census. Information from the census on housing value, household composition, and other characteristics at the person, household, block, and interviewer level were examined, with a subset of these variables found to be related to response propensity. This effort led to the development of improved nonresponse adjustment procedures in NHSDA. A second and unrelated study in this chapter called the Skip Pattern Experiment was fashioned to compare drug use reporting from two questionnaires: an experimental questionnaire that allowed the respondent to skip out of a set of questions if no drug use is reported, and the conventional questionnaire designed to require the respondent to answer all questions regardless of use. Results indicate that the skip pattern questionnaire produced less reporting of drug use.

Large-scale drug use surveys such as NHSDA provide excellent coverage of the general population and many demographic and socioeconomic subdomains; however, a sufficient number of sample members who use heroin regularly, for example, can be difficult to obtain using conventional sampling methods. Thompson's chapter presents innovative ways to reach sufficient numbers of these and other similar types of individuals through the use of adaptive sampling and graph sampling techniques. Also included is a discussion of resultant estimators that are design unbiased.

Understanding the methods used by researchers to measure the quality of self-reported drug abuse and associated behaviors is crucial. Hser's chapter provides a review of techniques used to assess reliability and validity of self-reported drug use and presents an assessment of the quality of self-report data among people at sexually transmitted disease clinics, emergency rooms, jails, and from a sample of narcotics addicts. Hser shows that adjustments for underreporting in these subpopulations should vary by gender, race, population type, and other factors. For example, among cocaine users who were self-reported nonusers, factors such as being female, minority, in jail, having multiple arrests in the past year, not being in treatment, and being dependent in the past were significantly correlated with positive urine results.

Tourangeau, Jobe, Pratt, and Rasinski report findings from a methodological study of reporting differences of sensitive behaviors such as pregnancy outcome (including abortion), the number of sexual partners, presence of a sexually transmitted disease, and level of condom use from a sample of women. Four modes of data collection by method of administration procedures were examined to determine the combination that results in higher levels of reporting. Overall, self-report clearly produced higher levels of reporting among women. Reporting based on use of computer-assisted collection versus conventional paper-and-pencil methods appear to be mixed.

In two studies, Lessler and O'Reilly compare the performance of audio computer-assisted self-interviewing (audio-CASI) with other methods such as an in-home computer-assisted personal interview (CAPI), out-of-home CAPI, and the traditional paper-and-pencil self-administered questionnaire (SAQ). In the first study, results show that computer-assisted interviewing produced higher rates of drug use reporting compared to the traditional SAQ procedure. The second study compared results of abortion reporting from the National Survey of Family Growth (NSFG), a major source of data on pregnancy and related information in the United States. Compared to CAPI, audio-CASI interviews produced reports of a higher number of abortions. Currently, more than 10,000 women participating in the NSFG have been successfully interviewed using audio-CASI technology.

When designing and conducting surveys involving sensitive topics such as drug use, it is important to have a good understanding of how privacy (or lack of it) during the interview affects the veracity of reporting. In a household survey of adults aged 18 to 45, Aquilino examined the effects of third-party presence on respondents' willingness to report drug use. Results show that the presence of a spouse or living partner while the interview was taking place did not seem to deteriorate the validity of self-report. On the other hand, truthful response appeared to decrease when a parent was present, even though all respondents were over age 17. In addition, these findings do not seem to vary by the three modes of administration used (self-administered, interviewer administered, and telephone).

Among other purposes, cognitive laboratory procedures can be used to gain a better understanding of how sensitive questions are perceived by the respondent. Willis provides a comprehensive and indepth review of the literature on cognitive laboratory-based research on sensitive

topics such as drug use, reproductive behavior, and drinking history. Based on laboratory research conducted by Willis and others, several recommendations are made related to the survey administration process. Some recommendations include the continued utilization of self-report as the primary mode of administration, shortening questionnaires on drug use, and limiting of complex concepts such as self-assessment of cause-and-effect relationships between drug use and deleterious life events.

Beimer and Witt provide a review of measurement error terminology such as measurement bias, reliability, validity, and mean square error. They present the mathematical relationship between reliability and validity (under appropriate assumptions), and discuss why measurement bias and validity should be treated as very different concepts. The main focus of this chapter is to examine the use of the Hui-Walter method for estimating measurement bias of self-reported drug use from the NHSDA. Taking advantage of redundancies in questions on drug use (i.e., lifetime use based on the recency question versus lifetime use based on any other question), the authors used this method to estimate false positive and false negative rates of drug use based on two sets of model assumptions: independence versus dependence of false negative rates between trials, among other things. A comparison of NHSDA false negative rates with denial rates from the National Longitudinal Survey of Youth (NLSY) generally showed, for example, a high correlation between the two for cocaine across various socioeconomic groups. This means that at a minimum, one will be able to determine which groups are more likely to contribute to false negative error.

Estimating hardcore use of drugs such as cocaine and heroin is a particularly challenging problem in major surveys due to the relatively small segment of the population involved in this behavior and the increased likelihood of underreporting very frequent use of these drugs. This will often lead to estimates with unacceptable sampling errors and measurement errors that may be much higher than those associated with other drugs and lower levels of use. In an attempt the address these shortcomings in the NHSDA, Wright, Gfroerer, and Epstein present a more sophisticated ratio estimator (than the one currently employed in the NHSDA) that incorporates population estimates from the Uniform Crime Report and the National Drug and Alcoholism Treatment Unit Survey. Results indicate that the alternative ratio estimator generated higher estimates of hardcore use of cocaine and heroin with higher levels of relative precision.

SUMMARY

In summary, the Technical Review sent a clear signal to the field that NIDA is supportive of the development and continuation of research on techniques to improve the validity of self-report and the accuracy of drug use estimates. Self-report will remain the primary mode of administration in drug use surveys, and is critical to obtain as valid information as possible. A developing body of important research information has been presented here and elsewhere about the successes and limitations of self-reported data collected from criminal justice and treatment populations. Researchers are also beginning to see validity studies conducted with general population groups, such as studies in the workplace. However, much more research needs to be conducted with more general populations such as households and school students, the major source of drug use pattern and trend data for the Nation. There is also a growing body of research on validity studies that vary data-collection methodologies, but a much more systematic approach to determining the impact of various factors that may be manipulated in a survey environment needs to be employed.

With regard to improvements in sample design and estimation, it is hoped that at least some of these chapters will encourage those in the survey community to continue to pursue and develop better ways to collect sensitive data (e.g., via results obtained from laboratory procedures and through computerization), measure nonresponse error and measurement error in a quantitative manner, and develop better estimators. It is also hoped these chapters will encourage the design of improved survey procedures used on rare and hard-to-reach populations that result in significant reductions in both sampling and nonsampling error.

Appreciation goes out to the more than 100 individuals from the public and private sector who attended the 2-day technical review. Thanks also are due to Mary Beth Babecki, Marc Brodsky, James Colliver, Peter Delaney, Andrea Kopstein, and Elizabeth Lambert of NIDA, and Joseph Gfroerer and Doug Wright from the Substance Abuse and Mental Health Services Administration for serving as reviewers of earlier versions of selected chapters.

TECHNICAL NOTE: ISSUES PERTAINING TO THE USE OF HAIR AS A MEDIUM TO ANALYZE DRUG USE

The central focus of this monograph is the accuracy of self-reported drug use. Using hair as a medium to analyze drug use has been receiving increased attention because of the less embarrassing circumstances of collection, and because hair does not decompose like other body fluids/tissues. Hair testing also offers a wider time window of detection of drug exposure than conventional urine testing. However, a review of the current state of the science of hair testing technology demonstrates the unresolved issues with hair testing.

The first caution is that the studies using hair testing reported in this monograph are not state-of-the-art at the time of the publication of the monograph. Increased research on hair testing has led to even more questions about this developing science. The mechanism of how drugs enter the hair remains unknown (Cone and Wang 1995; Kidwell and Blank 1995). Understanding the pathway of drug entry into hair is important for interpretation of results, i.e., if drugs get into hair only from blood there is less risk of contamination and more likelihood of dose-concentration and time-location relationships existing; however, if sweat or sebum are important contributors, then these relationships are expected to be much less reliable and introduce the risk of environmental contamination. Research has demonstrated that passive contamination occurs, and that procedures to remove external contamination are not effective (cf., Kidwell and Blank 1995).

The basic pharmacological relationship between drug dose and concentration in hair has not been demonstrated; the amount of drugs incorporated into the hair depends on a variety of factors (Kidwell and Blank 1995). The relationship between time of drug exposure and location of drug in the hair strand has not been clearly established. Studies with labeled cocaine have found only a limited dose and time relationship (Cone 1994*a*; Henderson et al. 1993; Kidwell and Blank 1995).

There is considerable developing evidence that cocaine selectively accumulates in darkly colored (black) hair compared to brown or blonde hair (Cone 1994*b*; Henderson et al. 1993; Kidwell and Blank 1994). Commercial companies assert that their techniques "remove the melanin fraction" prior to analysis nullifying concerns about racial biases. However, NIDA research mimicking their techniques and in vitro

binding experiments demonstrated a complete lack of effectiveness for removing color bias from hair testing (Cone, personal communication, 1996).

Variability also exists across and within drug types in terms of the accuracy of detection. Cocaine has been shown to readily bind to hair (Kidwell and Blank 1994), but uptake and washout rates of cocaine in hair vary extensively between individuals. Hair testing for marijuana is the most difficult test to perform and the least reliable. Research shows much smaller amounts of cannabis are incorporated into the hair. Consequently, a positive finding for marijuana for a subject who denies use increases the likelihood that the report is not accurate, but should not be used as an absolute indicator. A negative test is even more unreliable and should not be used to conclude that marijuana was not used. Hair is not yet considered a good medium to test for cannabis use (Hindin et al. 1994).

Despite these limitations, hair is increasingly being used in prevalence studies as a measure of drug use. Several of these studies are reported in chapters in the monograph. These studies generally show higher rates of drug use obtained by hair analysis as compared to self-report. The chapters describing these studies may provide a few cautions about the science of hair testing, but results are generally presented as if the hair test results were totally accurate. The "absoluteness" of positive or negative findings is disputable. Since many of these studies included a large proportion of black subjects, the issue of racial bias in hair testing is especially salient to consider, as well as other limitations of the current state of hair testing technology. While the studies reported herein are clearly valuable and add to the accumulating knowledge on the developing science of hair testing, they must clearly be evaluated as suggestive rather than definitive. Although hair testing of subjects who provide self-report data increases the information base on these subjects, hair test results should not be regarded as the absolute reference criteria determining whether the subject is truthful or not.

Several controversial aspects of hair testing remain unresolved, although the technology has progressed rapidly over the last decade. Unfortunately, few clinical studies have been conducted that resolve important issues needed for interpretation of hair test results. The Office of Workplace Programs within the Substance Abuse and Mental Health Services Administration (SAMHSA) plans a Technical Review and resulting publication examining the current state of the science with respect to

bioassay testing of bodily fluids and tissues in the spring of 1997. Hair testing will be a major focus. Readers should refer to the resulting monograph for up-to-date information on hair testing technology.

REFERENCES

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

Cone, E. "Potential for Ethnic Bias in Hair Testing." Paper presented at the SOFT Conference on Drug Testing in Hair, Tampa, FL, October 29-30, 1994*b*.

Cone, E.J., and Wang, W.L. How environmental drug exposure can affect hair testing for drugs of abuse. In: Cone, E.; Welch, M.; and Babecki, M., eds. *Hair Testing for Drugs of Abuse: International Research on Standards and Technology*. USDHHS, NIDA, NIH Pub. No. 95-3727, 1995.

Henderson, G.L.; Harkey, M.R.; and Jones, R. Hair Analysis for Drugs of Abuse, Final Report, Grant No. NIJ90-NIJ-CX-0012. National Institute of Justice, National Institute on Drug Abuse, September 1993.

Hindin, R.; McCusker, J.; Vickers-Lahti, M.; Bigelow, C.; Garfield, F.; and Lewis, L. Radioimmunoassay of hair for determination of cocaine, heroin, and marijuana exposure: Comparison with self-report. *Int J Addict* 29(6):771-789, 1994.

Kidwell, D.A., and Blank, D.L. "Incorporation of Drugs of Abuse Into and Their Removal from Hair." Paper presented at the SOFT Conference on Drug Testing in Hair, Tampa, FL, October 29-30, 1994.

Kidwell, D.A., and Blank, D.L. Mechanisms of incorporation of drugs into a hair and the interpretation of hair analysis data. In: Cone, E.; Welch, M.; and Babecki, M., eds. *Hair Testing for Drugs of Abuse: International Research on Standards and Technology*. USDHHS, NIDA, NIH Pub. No. 95-3727, 1995.

National Institute on Drug Abuse. Cone, E.; Welch, M.; and Babecki, M., eds. *Hair Testing for Drugs of Abuse: International Research on Standards and Technology.* USDHHS, NIDA, NIH Pub. No. 95-3727, 1995.

AUTHORS

Lana Harrison, Ph.D.
Associate Director
Center for Drug and Alcohol Studies
University of Delaware
77 East Main Street
Newark, DE 19716

(Formerly Statistician
Division of Epidemiology and Prevention Research
National Institute on Drug Abuse)

Arthur Hughes, M.S.
Chief
Epidemiology Research Branch
Division of Epidemiology and Prevention Research
National Institute on Drug Abuse
Room 9A-53
5600 Fishers Lane
Rockville, MD 20857

The Validity of Self-Reported Drug Use in Survey Research: An Overview and Critique of Research Methods

Lana Harrison

ABSTRACT

Since illicit drug use is by definition illegal, the tasks of measuring incidence and prevalence and charting the course of the epidemic have fallen to survey researchers over the past 30 years. Although survey methods have obvious advantages over indirect measures such as arrests, seizures, and treatment admissions, they are frequently criticized because they rely on valid self-reporting of sensitive and highly stigmatized behavior. Validation studies conducted before the mid-1980s involving known samples of drug users or urinalysis techniques suggested that drug use was fairly accurately reported in self-report surveys. However, more recent validation studies conducted with criminal justice and former treatment clients using improved urinalysis techniques and hair analyses demonstrate that self-report methods miss a lot of recent drug use. A review of the research literature suggests that neither self-reports nor bioassays are wholly accurate, and both have inherent problems. However, because self-report measures are necessary to understand the complexity of causal and correlational attributes of drug abuse, it is necessary to determine what can be done to improve valid self-reporting. This chapter examines the research literature on validation studies to provide an overview of what is known about the accuracy of self-reported drug use.

INTRODUCTION

How accurately can illicit drug use be measured in society? Drug use is an illegal activity and illicit drugs are illegal commodities; therefore, use cannot be measured by normal marketing procedures. Routinely compiled indicators such as police and court data on arrests and seizures, as well as clinical data on treatment admissions or drug-related medical emergencies provide a wealth of information, but can provide little

information about the incidence and prevalence of use. Surveys using self-report measures were initially developed as an alternative to anecdotal information or measures obtained from clinical, police, or court records. They were designed to assess the prevalence and frequency of illicit drug use in representative samples of the population. They were also designed to examine the correlates of illicit drug use and inform prevention and intervention efforts.

The sensitivity of collecting data on drug use has always made validity and reliability important issues. Survey research on drugs, where questions are asked about socially disapproved and illegal behaviors or socially marginal attitudes, may well generate inaccurate reporting and bias in survey estimates. Survey researchers recognize the need to design methods that elicit accurate and truthful reporting of drug use experience and attitudes. However, not much research has been conducted on the factors that improve an individual's reporting of sensitive information on questions about potentially embarrassing or self-incriminating behavior.

Even with their limitations, surveys are still a good measure of the nature and extent of drug use in a population, and provide information on the characteristics of drug users in a society. Treatment data provide very important information on the characteristics of people presenting themselves for treatment, but tell nothing about the characteristics of the pool of individuals from which those people are drawn or how those who enter treatment differ from those who do not. Likewise, drug users who become involved with the criminal justice system are not representative of drug users in general. Even ethnography, which frequently employs a loosely structured interview conducted in a more naturalistic setting, is restricted by its lack of generalizability to a known population, although it can provide a wealth of detailed information. Survey research can provide a more thorough profile of drug use and abuse among a broader cross-section of the population, and it can also provide a much greater range of information for use in designing intervention strategies. But the challenge is how to convince survey respondents to provide accurate information. Guarantees of anonymity and confidentiality are now standard fare in survey research on drugs. However, the research evidence suggests this is not enough to allay fears of some respondents in reporting recent drug use. Several of the articles in this monograph report on studies, particularly among criminal justice populations and treatment clients, in which recent drug use is underreported (see Mieczkowski and

Newel, this volume; Wish et al., this volume; Magura and Kang, this volume).

METHODS USED TO TEST THE VALIDITY OF SELF-REPORT

Validating self-report requires comparison to some method that is presumably more accurate. Over the past several decades, technologically sophisticated methods have been developed for analyzing drug metabolites in bodily fluids. Urine is most often used, but drugs have also been detected in blood, saliva, semen, meconium, perspiration, and hair (Smith and Liu 1986; Cone, this volume). Each biological specimen is unique and offers a somewhat different pattern of information regarding drug use over time. Also, each specimen has unique strengths and weaknesses regarding the type of information obtained from drug testing. The same testing methods are generally applied to the various bodily fluids. Testing methods fall into either screening or confirmations assays. Screening assays are generally valid—usually erring on the side of not identifying specimens that may contain drugs or their metabolites rather than identifying specimens as positive that do not contain drugs or drug metabolites. Confirmations assays are more expensive, but they are also more specific in identifying drug use.

Aside from biological assays, other methods have been explored for their potential to validate self-reports of drug use. Official record checks, such as checks of criminal justice or treatment records, have frequently been used to validate self-reports of drug use. Reports by family, close friends, or counselors have been used (Stephens 1972). Even polygraph tests have been used to validate self-reports of drug use (Clark and Tifft 1966). These types of validation procedures, which rely on checking validity against external criteria, are examining external or empirical validity.

Internal validity procedures are performed in a cross-sectional survey to determine the amount of internal consistency across survey items. Internal validity checks may also be employed to assess internal consistency across survey items on repeated administrations of a longitudinal survey. For example, both the National Household Survey on Drug Abuse (NHSDA) (Substance Abuse and Mental Health Administration 1995*a*, 1995*b*) and the Monitoring the Future study (Johnston et al. 1995) of high school students demonstrate a high amount of internal consistency. In the Monitoring the Future survey, theoretically

predicted relationships among a number of deviant behaviors have been demonstrated; estimates of friends' drug use closely parallel cumulative estimates of overall drug use (O'Malley et al. 1984). Analyses of NHSDA data show consistent patterns of self-reported friendship with users of specific drugs, opportunity to use these drugs, and actual drug use (Harrell 1985). Analyses of the longitudinal followups of the Monitoring the Future data have also shown relationships among variables to persist over time (O'Malley et al. 1984). Drug use in the years following high school is highly consistent with and predictable from senior year drug use. Analyses also showed that past-year marijuana and alcohol use were more reliably measured than use in the past 30 days. Marijuana use was more reliably measured than the use of other illicit drugs (O'Malley et al. 1984).

Urine Testing and Recent Challenges to Validity

There are obvious limitations to internal consistency checking and record checks, which is why the field has looked to bioassays to test the validity of self-report. Urine testing in particular has become more widespread and is considered to be quite valid, particularly with recent technological improvements. Earlier urine tests such as thin layer chromatography (TLC) have been found to be much less valid than the more recently developed tests such as enzyme multiplied immunoassay (EMIT) or fluorescence polarization immunoassay (FPIA). In a comparison of the three urine tests, the U.S. Justice Department found EMIT and FPIA to have false positive rates of about 0.2 to 2.5 percent (incorrectly identifying a negative specimen as positive), and false negative rates of 2.4 to 40.8 percent (incorrectly identifying a positive specimen as negative) (Visher 1991). The highest false negative rates were found for marijuana. Radioimmunoassay (RIA) was found to have a false positive rate of 0.1 to 4.1 percent, with the highest false positive rate (4.1 percent) associated with cocaine. The false negative rates for RIA ranged from 5.8 percent to 37.1 percent. The highest false negative rate was for marijuana. The validity of these urine tests was determined by comparing the EMIT and FPIA technologies to gas chromatography/mass spectrometry (GC/MS), which is presumed to be virtually 100 percent accurate. TLC showed a 0.3 to 3.1 percent false positive rate, but a false negative rate of 52 to 92 percent (Visher and McFadden 1991). All of the tests err on the side of not identifying a negative specimen as positive, which means they sacrifice the ability to correctly identify many specimens. However, the TLC test performance is significantly poorer than the other tests in terms of failing to identify positive specimens correctly.

Since the late 1980s, several studies using the improved urinalysis technology have disputed the accuracy of self-report drug use surveys. It had generally been believed that estimates of self-reported drug use were reasonably valid. In a review of self-report validation studies, Mieczkowski (1990) found that researchers reported validity rates of generally 70 percent or higher, and some even 90 percent or higher. The first large-scale study to cast doubt on the validity of self-reported drug use was the Drug Use Forecasting (DUF) study (Harrison 1989; National Institute of Justice 1990). Begun in several major U.S. cities in 1987, the study grew to include most of its current 23 sites by 1989. The DUF study employs urinalysis to measure drug use among those recently arrested and charged for serious crimes. Interviews are conducted in a central booking facility in several large U.S. cities, where privacy is not always available. Respondents are informed that the study is anonymous and confidential, and their participation will not have a bearing on their case. They are asked for a urine specimen at the end of the interview.

The DUF study has fairly consistently found that only about half of those who test positive for a drug report use in the past 2 to 3 days. Figure 1 compares drug use prevalence rates measured by urinalysis and self-report for the entire sample of arrestees participating in the DUF study in 1991. The most common way to interpret the congruence of urinalysis and self-report is to focus on just those with positive urinalyses and determine the percentage who accurately report their drug use. Of the 17.3 percent testing positive for marijuana, 9.3 percent report use in the past 3 days and 8.1 percent do not. Considering those who test positive, only about one-third to one-half admit their drug use. But this interpretation must be balanced against the interpretation that takes into account a fuller range of information. For example, notice that for marijuana 74.2 percent tested negative and (self-) reported no use of the drug in the past 3 days. Looking at the lower right hand corner of the marijuana grid, notice that 9.3 percent tested positive and admitted recent marijuana use. Therefore, in 83.5 percent of the cases, there was congruence between self-report and urinalysis. The rate of congruence for cocaine is 77.2 percent, and 95.3 percent for opiates. However, this measure of congruence is heavily influenced by the prevalence of drug use. The less likely the use, the higher the congruence rate.

Another way to look at the data is to compare the percentage who self-report use of a drug to the percentage who test positive for that drug. For example, 21.1 percent of the arrestees told the DUF interviewer that they

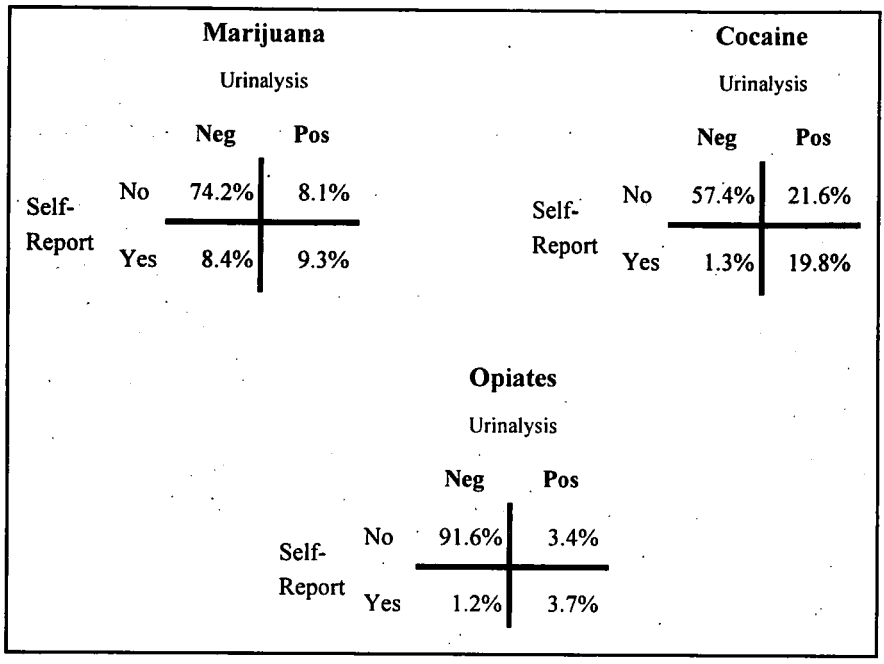


FIGURE 1. *Three-day self-report versus urinalysis among arrestees in major U.S. cities. Total DUF sample N = 34,720.*

SOURCE: National Institute of Justice 1990.

used cocaine in the past 3 days, but twice as many—41.3 percent—tested positive. About 4.9 percent reported recent opiate use, but 7.1 percent were found positive by urinalysis. The numbers are much closer in the case of marijuana, for which 17.7 percent admitted use in the past 3 days and 17.4 percent tested positive. The relative comparability between self-report and urinalysis for marijuana is largely impacted by those admitting use but not testing positive. Therefore, at least in the case of marijuana, self-report of use in the past 3 days appears to detect as much marijuana use as urinalysis—and this among individuals being interviewed in a jail setting. The fact that they underreport is probably not as surprising as the fact that many report validly. Despite assurances that they are participating in a confidential study, the respondents are interviewed in jail while awaiting arraignment; many are concerned about the outcome of their arraignment and anxious to talk to anyone. Perhaps it is not so surprising that recent cocaine and opiate use are not more validly reported among this population. Moreover, the congruence

rate between self-report and urinalysis even among deviant groups in a less-than-ideal interviewing environment is noteworthy.

Another factor to consider is that earlier analyses also assume that the EMIT test is totally accurate. Recall that a Justice Department study found EMIT's false positive rate to fall between 2.1 and 2.5 percent for marijuana, opiates, and cocaine (Visher 1991). However, a more recent large scale study conducted by the National Institute on Drug Abuse (NIDA) found higher false positive rates in urine samples collected primarily in workplace settings, analyzed by EMIT, and for which positives were confirmed by GC/MS. Because the study involved laboratories that used GC/MS to verify samples that tested positive by EMIT, it can therefore only report on the false positive and negative rate of EMIT-screened positive specimens, and not on how well EMIT correctly identifies positive samples. In these tests, between 95 and 96 percent of the cocaine and marijuana samples found positive primarily by EMIT were confirmed by GC/MS, but only 53 to 55 percent of the opiates samples were confirmed (Stephenson 1992; Harrison 1995).

Urinalysis is not an exact science. EMIT can detect cocaine for 2 to 3 days in the urine. Opiates are detectable for 2 to 4 days, although detection time is generally limited to 2 days. Cannabis may be detected for up to 4 weeks. Further, the window of detectability is not a constant that applies in all cases, but is rather dependent upon the particular type of drug, the physical condition of the individual (i.e., state of hydration and fluid balance), the route of drug ingestion (i.e., intranasally, intravenously), the amount of drugs used, and the individual's frequency of use (American Medical Association 1987).

Urinalysis has limitations in terms of what it can reveal about the validity of self-report. In addition, factors inherent to interviewing incarcerated people recently arrested for serious offenses limit the generalizability of the results to other samples. Validation studies employing urinalysis techniques have frequently been conducted on criminal justice populations, but the results are not generalizable to other populations because they oversample heavy drug users. Likewise, studies that use record checks to validate drug use may introduce bias simply because the characteristics of people likely to have records may differ significantly from those in the general population. Nevertheless such studies, and particularly the DUF study, have been used to call into question the validity of self-reported drug use in all surveys—regardless of the characteristics of the populations being surveyed. Findings on the validity of self-report from criminal

justice populations have led to concern about the validity of self-reported drug use from several policymakers (e.g., U.S. Senate 1990; General Accounting Office 1993).

Hair Testing

Hair analysis is a newly developing technology being used as a check on the accuracy of self-reported drug use. It is increasingly being performed in numerous laboratories, some of which offer commercial drug testing services. Hair analysis is being used by private employers in pre-employment drug screening and tested in criminal justice settings. A pretrial diversionary program in New Orleans relies on hair testing, in addition to urinalysis and self-report, to monitor compliance with program rules of abstinence from illicit drugs (Mieczkowski et al. 1995). Hair testing has been receiving increased attention because of the less embarrassing circumstances of collection. Further, hair can be easily stored. Hair samples are generally collected from the vertex of the scalp, and then washed and dissolved with an acid or a strong base. With the hair sample prepared in this manner, the same types of analytical principles and technology used to analyze urine can be applied.

Although the technology of hair testing has progressed rapidly over the last decade, several highly controversial aspects of the procedure remain unresolved. It is still unclear how drugs enter the hair, creating concerns about contamination via exposure to cocaine dust particles, smoke, vapor, or drug solutions. At least two studies have found cocaine in the hair of children, suggesting that contamination is an important consideration (see Randall 1992; Smith et al. 1994). Cocaine has been shown to readily bind to hair, but binding depends on several physicochemical variables such as pH of exposure, ionic strength, and hair type (Kidwell and Blank 1994). Research shows much smaller amounts of cannabis are incorporated into the hair, and hair is not yet considered a good medium to test for cannabis use. Hair testing appears most valid for testing cocaine use (Mieczkowski and Newell, this volume; Hindin et al. 1994).

Another controversial issue in hair testing 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. Further, because hair grows at the rate of approximately 1.5 ± 0.15 centimeters per month, it is believed that hair can be segmented to provide a record of an individual's drug use equivalent to the length of the hair. But studies

with labeled cocaine have found only a limited dose and time relationship (Henderson et al. 1993; Cone 1994*a*). Studies have shown that at any one time, about 85 percent of head hairs are growing (Hindin et al. 1994). There is also evidence of ethnic differences in hair test results; coarse, dark hair retains more of the drug than other hair types (Henderson et al. 1993; Cone 1994*b*; Kidwell and Blank 1994). Despite these limitations, hair is increasingly being used to detect drug use. Several studies comparing hair, urine, and self-report results are reported in chapters in this monograph. These studies suggest that hair analyses disclose more recent drug users than can be found through either urinalysis or self-reports.

One of the early studies comparing hair, urine, and self-report was conducted by Mieczkowski and colleagues (1991*b*) in Florida using a prototype of the DUF study. Hair was analyzed by RIA, and both EMIT and FPIA were used to test urine. Mieczkowski and colleagues concluded that about four times as many arrestees had a positive hair assay as self-reported cocaine use within the previous 30 days. There was a ninefold increase in the number who had hair positive for opiates as compared to self-reported opiate use in the past 30 days (Mieczkowski et al. 1991*b*). Mieczkowski and colleagues also 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). They determined that self-report was least reliable for cocaine (Mieczkowski et al. 1991*b*). There were many inconsistencies in comparisons of the urine and hair samples, which to some extent was expected because the hair was analyzed for the past 60 days. However, not all the inconsistencies can be explained by the differences in the time frames of the tests employed. In fact, in a study of probationers in which an average of 5.9 urine samples were obtained per month, only 46 percent of positive RIA for hair (RIAH) tests were confirmed by urinalysis for cocaine and only 60 percent for heroin (Baumgartner et al. 1989; also reported in Hindin et al. 1994).

Some discrepancies between urine and hair analysis results cannot easily be explained. Recall that RIA, which has been extensively employed in analyzing hair samples, was found in the U.S. Justice Department studies mentioned earlier to have a 4.1 percent false positive rate for cocaine based on analysis of urine. Nevertheless, researchers tend to conclude that hair analysis is most accurate in detecting cocaine use. The host of unresolved issues surrounding hair analyses give reason for concern in attempting to validate drug use. Most studies that have

been conducted analyzing drugs in hair have used relatively small numbers of subjects and have failed to include proper controls. Washing may be problematic because drugs may be removed by this procedure. The effects of shampooing and cosmetic treatments such as dyeing, perming, or bleaching, and exposure to ultraviolet light or other external contaminants may alter the presence of the drug in the hair shaft (Harkey and Henderson 1989; Henderson et al. 1993). Research has shown that uptake and washout rates of cocaine in hair, for example, vary extensively between individuals and may be related in part to differences in hygiene. It is critical that validation of self-report data by drug testing be performed with careful consideration of the limitations imposed by the testing methodology and the biological specimen.

Validation Studies

There are studies in the literature that suggest relatively high rates of self-reported drug use. For example, Zanis and associates (1994) found that for a sample of patients in methadone treatment for at least 6 months, 13 percent and 19 percent of those testing positive by EMIT for opiates and cocaine respectively, failed to self-report use in the previous month. Additionally, 58 percent and 28 percent of those with negative urines for opiates and cocaine, respectively, reported use of the drug during the previous month. 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. This was a highly drug-involved sample; 76 percent reported injecting drugs in the past 30 days. Self-reports and urinalysis results agreed for 85 percent of the heroin users. 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 in 5.2 percent and 5.8 percent, respectively, of respondents who did not self-report heroin or cocaine use in the past 48 hours (Weatherby et al. 1994). Therefore, self-report produced much higher rates of drug use than obtained by urinalysis. Likewise, analyses by Weatherby and colleagues (1994) suggest that heavily drug-involved individuals can self-report recent drug use fairly validly.

In a validity study conducted among a workplace population, which more closely resembles a general population than treatment or criminal justice populations, self-reports were found to quite reliably measure drug use. Cook and colleagues (this volume) found that self-reports

produced higher prevalence rates than either urinalysis or hair analysis. For the entire sample (N = 800) of employees from a large steel plant, only 7.8 percent tested positive for any drug by urinalysis, while 9.4 percent reported recent drug use. For the subsample that had hair tests, 6.2 percent were positive for an illicit drug and 9.9 percent reported recent use. The most frequently detected drug was marijuana, and little cocaine use was found by testing. It was also anticipated that some of the subjects who tested positive for morphine or sedatives failed to report prescription drug use. Another important caveat is that the study used much lower cutoff levels than recommended by NIDA for determining a urine specimen as positive (because the analyses were being conducted for research purposes only). Because of the small number with positive drug assays despite the lowered cutoff levels, the validity analyses were combined across all drug types. Although self-report methods produced higher prevalence rates than testing, Cook and colleagues found only about half of those positive for any drug by either urine or hair test self-reported recent use. They concluded that it is necessary to use multiple assessment methods to estimate self-report. Cook and colleagues also varied the method of data-collection setting between telephone interview, personal interview in the workplace, group interview in the workplace, and personal interview offsite. They found rates of drug use self-report were highest in the individual workplace interview and lowest in the group interview in the workplace. The results nevertheless suggest that self-report methods appear to provide good measures of prevalence—in this study, higher than those generated by the assay tests.

It seems clear that although drug use may vary substantially among different populations such as household members, students, and arrestees, the accuracy of their self-report may also vary substantially. The research literature suggests that self-report may be the least reliable among criminal justice clients. Magura and Kang (this volume) report the results of two validity studies conducted using similar methods in New York City. One study included a sample of patients in two methadone treatment programs recruited because clinic records showed they had tested positive by urinalysis for cocaine. The other study included a sample of criminally involved young 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 both urine and hair samples were obtained on all the clients at followup interviews. For the methadone sample, 60 percent self-reported recent cocaine use and 80 percent were positive by RIAH. For the young adult sample, 23 percent self-reported recent cocaine use, but almost three

times as many—67 percent—were positive by RIAH. This led the researchers to conclude that self-report is more valid for treatment clients than for criminal justice clients. 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. This would indicate that the young adults were not afraid of self-reporting sensitive information. Magura and Kang suggest that for the young adults, cocaine (or more specifically, crack) had become stigmatized, even though dealing of these drugs was not; while the young adults reported drug dealing, they were more reluctant to report cocaine use. There may also 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 youth's handling of cocaine/crack for sale.

With respect to the validity of self-report among treatment clients, the research literature suggests that self-report is more accurate at intake. That is, clients are more likely to provide self-reports that are congruent with urine or hair test results in the beginning stages of treatment than they are at followup posttreatment (Wish, this volume; Hindin et al. 1994). For example, Hindin and colleagues (1994) found that among 109 entrants to two New England treatment facilities, 89 percent of the 87 found positive for cocaine by hair and 96 percent of 45 heroin positives were confirmed by self-report. However, among the 86 followed up, only 51 percent of the 43 found positive for cocaine by hair and 67 percent of 18 heroin positives were confirmed by self-report. This is an important finding because it suggests the importance of validating self-report in studies measuring treatment outcome.

Limitations of Validity Research

Validity research is still in its early stages. It is hampered by the limitations of technology, but also by the lack of sophisticated knowledge about critical elements conducive to the honest reporting of sensitive information in a survey environment. There are undoubtedly multiple influences on respondents in terms of their ability and desire to provide a valid response. These factors include setting, real or perceived consequences of reporting use, literacy, clarity of questions, and memory. Studies validating self-report have frequently not taken advantage of procedures to maximize accurate self-reporting of illicit drug use. Although studies typically promise anonymity and confidentiality,

confidentiality can be compromised by administration procedures that require respondents to provide their answers aloud to interviewers.

In a series of methodological studies undertaken in conjunction with the NHSDA, significantly higher rates of drug use were found using self-administered answer sheets as opposed to having respondents answer aloud to interviewers. The methodological field test found respondents were 1.6 times more likely to report cocaine use in the past year and 2.4 times more likely to report use in the past 30 days on the self-administered answer sheets. The increase in reported drug use was a function of the recency of the event, with few differences in lifetime rates, more difference in past-year rates, and the greatest difference in rates of past month drug use. A more recent national field test of even the cigarette questions revealed significantly higher rates of cigarette use reported using a self-administered as opposed to an interviewer-administered answer sheet (SAMHSA 1995*a*).

Similar findings about the impact of self- versus interviewer-administered questions on drug use have been found in the National Longitudinal Survey of Labor Market Experience, Youth Cohort (NLSY). Substantially less cocaine and marijuana use were reported in the interviewer-administered conditions. Respondents were more likely to report marijuana as opposed to cocaine use (Schober et al. 1992). Mensch and Kandel (1988) also found self-administered answer sheets to yield higher reports of drug use than interviewer-administered questions. Likewise, a study of prison inmates found more reports of drug use using self-administered versus interviewer-administered questionnaires (McElrath 1994).

Another finding from the series of methodological studies conducted in conjunction with the NHSDA that has been replicated in a number of studies: The more stigmatized the drug, the more prevalence rates are suppressed. Marijuana use is reported more validly than cocaine use. This finding has been replicated in several studies (Harrison 1992, 1995; Fendrich and Xu 1994; Mieczkowski et al. 1991*a*). Currently the most stigmatized drug appears to be cocaine (or more specifically, crack). In the DUF study, comparing the percentages who self-report use of the respective drugs to the percentages who test positive leads to the conclusion that arrestees are most willing to admit marijuana use, followed by opiates, amphetamines, and then cocaine (Harrison 1992).

To maximize reporting the use of stigmatized drugs, it is vital to use procedures that maximize confidentiality. This includes using self-

administered rather than speak-aloud interviews. Most of the research that has been done has involved paper-and-pencil questionnaires, but research is also beginning on the impact of computer-assisted self-interviewing (CASI) in improving the validity of self-reported drug use (Lessler and O'Reilly, this volume). The setting also needs to be explored in validity research, because it logically makes sense that recent arrestees interviewed in jail prior to being sentenced may not provide especially valid reports of recent drug use. There is also very limited research on respondents' perceptions of risk of providing truthful information to sensitive drug questions (Willis et al., this volume). The scientific community needs to engage in a systematic research program, varying different aspects of the interview environment and methodology to improve understanding of factors that can improve the validity of self-reports. Further, other factors that contribute to validity in survey research must not be forgotten (Gfroerer et al., this volume). The generalizability of survey research is predicated on proper selection procedures from a known universe. Also, surveys must have a respectable response rate to ensure that bias is not introduced if members of the target population are consistently underrepresented in the sample.

DISCUSSION

Concern is often expressed about the validity of survey data on drug use. Survey-generated estimates of drug use are frequently criticized on the grounds that many survey respondents are not honest in reporting illicit drug use. At this point, it is not possible to judge how validly individuals report their drug use in surveys. There are inherent difficulties in trying to measure the validity of self-report based on the current available methods. The available chemical test(s) and methods now used to judge the validity of self-reported data on drug use surveys have limitations. Urine tests have a narrow window of detectability, which greatly reduces their usefulness. Studies based on record checks may be biased simply because the characteristics of people likely to have records may differ significantly from those in the general population.

Perhaps hair analysis holds the greatest promise in providing a standardized external validity criterion measure because it measures drug use over a longer period of time and hair samples can be obtained unobtrusively. However, more research is needed before this method can be deemed reliable and valid. Research has yet to answer unresolved problems with hair testing. The consensus of scientific opinion is that

hair analysis for the presence of drugs of abuse is unreliable and is not generally recognized by qualified experts as effective. In October 1994, the Society of Forensic Toxicologists held a meeting sponsored by NIDA to review the available research on hair testing. The participants concluded there were even more questions about hair testing for drugs of abuse than before (Society of Forensic Toxicologists 1994). Therefore, hair analysis cannot currently be described as useable with acceptable accuracy.

Thus far, the largest problem with most external criteria validation studies is that results are inherently not generalizable. Urinalysis studies, and now studies using hair analysis, are most often conducted on populations that have much higher rates of drug use than the general population. Respondents have not been randomly recruited from some definable population, so results can be generalized only to a discrete population group. Notably, validity rates for criminal justice populations cannot be generalized to the general population. Criminal justice populations may be less honest because they could be heavily penalized if their drug use were known to authorities. Consequently, those involved with the criminal justice system may make different decisions about candor in interviews than would the general population.

The research literature suggests that the validity of self-report varies by population subgroup. For example, arrestees are much less likely to provide honest reports of recent drug use than people in treatment. Self-report surveys of employees found higher prevalence rates based on self-report than on urine analysis or hair analysis. However, there is still a dearth of good studies that look at the validity of self-report in general populations. Most research conducted on validating self-report has focused on criminal justice and treatment populations, and is limited in its ability to determine how accurately respondents report drug use in general population surveys (such as household and student surveys).

Despite the concerns with the generalizability of the results of most validation studies, research does point to some general conclusions that may be drawn about the validity of self-reported drug use in a survey environment. Clues are provided in the consistency of results across several studies that show differences in self-reporting by drug type. The pattern of reporting is consistent with the social desirability hypothesis about more stigmatized drugs such as cocaine, the least validly reported. Also, as the use of drugs becomes more recent, it appears to be subject to increasing bias; respondents are most willing to report lifetime use and

least willing to report use that occurred in the very recent past. This has further implications for the usefulness of urinalysis to validate self-report information such as that derived from household or student surveys, since they are generally concerned with measures of lifetime, past-year, and past-month drug use, not use in the past 2 to 3 days (which urinalysis is able to measure). Another finding is that the use of self-administered questionnaires tends to produce higher prevalence rates (and ostensively, more valid data) than interviews in which the respondents must speak their responses aloud.

Some surveys undoubtedly obtain more valid information than others. Even within surveys, differences in interviewer styles and presentation influence validity. Probably what is most amazing is that individuals will admit to illicit drug use in surveys. There are definite limitations to survey research on drug use, but perhaps many of those can be overcome with research designed to further improve the validity of self-report. It is important to remember that most validity research, in fact, shows quite high congruence rates between self-report and assay results. Factoring in that some of the differences between self-report and urinalysis found in validation studies are also due to the interview process (i.e., question wording, interview expectations, setting) leads to the conclusion that even among at-risk populations, there is a high degree of congruence between self-report and urinalysis. Of course, the research also shows a lot of individual variation; many validation studies find only about half of those testing positive for an illicit drug report using that drug. But do not forget the limitations of the testing technology. The current state of the science suggests that the most appropriate presentation of results from hair testing, in particular, would be in the aggregate, and not at the individual level.

In conclusion, self-report information is always going to be necessary, because biological assays can only corroborate drug use. Assays cannot determine the age at which individuals initiated drug use, individuals' attitudes about the risk of harm, perceptions of drug availability, other factors that may co-vary with drug involvement (such as other deviant behavior), and even whether individuals have received treatment in the past. It will always be necessary to rely on self-report to collect some sensitive information, which suggests it is imperative to conduct research on those factors that can be manipulated within a survey environment to increase the validity of self-reporting of sensitive information. Since much of the research that scientists, policymakers, the media, and other interested individuals use to inform themselves about drugs is based on

self-report, it is important to engage in more systematic and rigorous scientific studies to improve the validity of self-report. To really determine how accurately self-report survey research methods measure drug use, it is essential to gather scientific data on what methodological or environmental circumstances can be manipulated to improve the validity of self-report for what types of population subgroups.

REFERENCES

American Medical Association, Council on Scientific Affairs. Scientific issues in drug testing. *JAMA* 257(22):3110-3114, 1987.

Baumgartner, W.A.; Hill, V.A.; and Blahd, W.H. Hair analysis for drugs of abuse. *J Forensic Sci* 34(76):1433-1453, 1989.

Clark, J.P., and Tifft, L.L. Polygraph and interview validation of self-reported deviant behavior. *Am Sociol Rev* 31:516-523, 1966.

Cone, E. "Drug Testing in Hair." Paper presented at the Society of Forensic Toxicologists (SOFT) Conference on Drug Testing in Hair, Tampa, FL, Oct. 29-30, 1994*a*.

Cone, E. "Potential for Ethnic Bias in Hair Testing." Paper presented at the SOFT Conference on Drug Testing in Hair, Tampa, FL, Oct. 29-30, 1994*b*.

Fendrich, M., and Xu, Y. The validity of drug use reports from juvenile arrestees. *Int J Addict* 29(8):971-985, 1994.

General Accounting Office (GAO). *Drug Use Measurement Strengths, Limitations, and Recommendations for Improvement.* Report GAO-PEMD 93-18. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1993.

Harkey, M.R., and Henderson, G.L. Hair analysis for drugs of abuse. In: Basalt, R., ed. *Advances in Analytical Toxicology*. Chicago: Year Book Medical Publishers, 1989.

Harrell, A.V. Validation of self-report: The research record. In: Rouse, B.; Kozel, N.; and Richards, L., eds. *Self-Report Methods of Estimating Drug Use*. National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985. pp. 12-21.

Harrison, L.D. "The Validity of Self-Reported Drug Use Among Arrestees." Paper presented at the 41st meeting of the American Society of Criminology, Reno, Nov. 8-12, 1989.

Harrison, L.D. The validity of self-reported data on drug use. *J Drug Issues* 25(1):91-111, 1995.

Harrison, L.D. Trends in illicit drug use in the USA; Conflicting results from national surveys. *Int J Addict* 27(7):817-847, 1992.

Henderson, G.L.; Harkey, M.R.; and Jones, R. "Hair Analysis for Drugs of Abuse." Final report of activities under National Institute of Justice/ National Institute on Drug Abuse grant no. NIJ90-NIJ-CX-0012, 1993.

Hindin, R.; McCusker, J.; Vickers-Lahti, M.; Bigelow, C.; Garfield, F.; and Lewis, L. Radioimmunoassay of hair for determination of cocaine, heroin, and marijuana exposure: Comparison with self-report. *Int J Addict* 29(6):771-789, 1994.

Johnston, L.D.; O'Malley, P.N.; and Bachman, J.G. *National Survey Results on Drug Use from the Monitoring the Future Study, 1975-1992.* Rockville, MD: National Institute on Drug Abuse, 1995.

Kidwell, D.A., and Blank, D.L. "Incorporation of Drugs of Abuse into and Their Removal from Hair." Paper presented at the SOFT Conference on Drug Testing in Hair, Tampa, FL, Oct. 29-30, 1994.

McElrath, K. A comparison of two methods for examining inmates' self-reported drug use. *Int J Addict* 29(4):517-524, 1994.

Mensch, B., and Kandel, D. Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects. *Pub Opin Q* 52:100-124, 1988.

Mieczkowski, T. The accuracy of self-reported drug use: An evaluation and analysis of new data, In: Weisheit. R., ed. *Drugs and Crime and the Criminal Justice System.* Cincinnati: Anderson Publishing Co., 1990.

Mieczkowski, T.; Barzelay, D.; Gropper, B.; and Wish, E. Concordance of three measures of cocaine use in an arrestee population: Hair, urine and self-report. *J Psychoactive Drugs* 23(3):241-249, 1991*a*.

Mieczkowski, T.; Landress, H.J.; Newel, R.; and Coletti, S.D. "The Concordance of Drug Use Indicators: Urine, Hair and Self-Report in an Arrestee Population." Report of activities under National Institute of Justice contract no. 90-IJ-CX-0023, 1991*b*.

Mieczkowski, T.; Mumm, R.; and Connick, H.F. The use of hair analysis in a pretrial diversion program in New Orleans. *Int J Offender Ther* 39(3):222-241, 1995.

National Institute of Justice. *DUF: 1988 Drug Use Forecasting Annual Report.* Washington, DC: National Institute of Justice, 1990.

O'Malley, P.M.; Bachman, J.G.; and Johnston, L.D. Reliability and consistency in self-reports of drug use. *Int J Addict* 18:805-824, 1984.

Randall, T. Infants, children test positive for cocaine after exposure to second-hand crack smoke. [Letter to the editor]. *JAMA* 267(8), 1992.

Schober, S.E.; Caces, M.F.; Pergamit, M.R.; and Branden, L. Effect of mode of administration on reporting of drug use in the National Longitudinal Survey. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 267-276.

Smith, F., and Liu, R. Detection of phenobarbital in bloodstains, semen, seminal stains, perspiration stains and hair. *J For Sci* 26:582-586, 1986.

Smith, F.P.; Kidwell, D.A.; and Cook, L.F. "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, Oct. 29-30, 1994.

Society of Forensic Toxicologists (SOFT). *Consensus Opinion Summarizing the Current Applicability of Hair Analysis to Testing for Drugs of Abuse.* SOFT, June 1990.

Stephens, R.C. The truthfulness of addict respondents in research projects. *Int J Addict* 7:549-558, 1972.

Stephenson, R.L. "The Potential for Collision: Drug Testing in the Workplace and in the Criminal Justice System." Paper presented at the National Institute of Justice/Bureau of Justice Assistance Third Annual Evaluating Drug Control Initiatives Conference, Washington, DC, July 27, 1992.

Substance Abuse and Mental Health Services Administration (SAMHSA). *National Household Survey on Drug Abuse: Main Findings 1992.* DHHS Pub. No. (SMA)94-3012. Rockville, MD: SAMHSA, 1995*a*.

Substance Abuse and Mental Health Services Administration (SAMHSA). *Preliminary Estimates from the 1994 National Household Survey on Drug Abuse.* Rockville, MD: SAMHSA, 1995*b*.

U.S. Senate. *Drug Use in America: Is the Epidemic Really Over?* Summary of findings of a majority staff study prepared for the Judiciary Committee, 101st Congress, Dec. 19, 1990.

Visher, C. A comparison of urinalysis technologies for drug testing in criminal justice. *National Institute of Justice Research Report.* Washington, DC: U.S. Department of Justice, National Institute of Justice, 1991.

Visher, C., and McFadden, K. A comparison of urinalysis technologies for drug testing in criminal justice. *National Institute of Justice Research in Action.* Washington, DC: U.S. Department of Justice, National Institute of Justice, 1991.

Weatherby, N.L.; Needle, R.; Cesari, H.; Booth, R.; McCoy, C.B.; Watters, J.K.; Williams, M.; and Chitwood, D.D. Validity of self-reported drug use among injection drug users and crack cocaine users recruited through street outreach. *Eval Prog Plan* 17(4):347-355, 1994.

Zanis, D.A.; McLellan, A.T.; and Randall, M. Can you trust patient self-reports of drug use during treatment. *Drug Alcohol Depend* 35:127-132, 1994.

AUTHOR

Lana Harrison, Ph.D.
Associate Director
Center for Drug and Alcohol Studies
University of Delaware
77 East Main Street
Newark, DE 19716

167341

The Validity of Self-Reported Drug Use Data: The Accuracy of Responses on Confidential Self-Administered Answered Sheets

Adele V. Harrell

ABSTRACT

Official records offer a relatively inexpensive, nonintrusive strategy for checking on the accuracy of self-reported drug use. Responses of a small sample (N = 67) of former drug treatment clients interviewed using procedures exactly modeled on the National Household Survey on Drug Abuse were compared to their clinic records. The accuracy of reports compared to clinic records varied by drug, with the percentage of known users reporting their use highest for marijuana, followed by cocaine and hallucinogens, and lowest for heroin. Almost half of this sample of former treatment clients denied ever receiving drug treatment.

INTRODUCTION

Self-reported data are a mainstay of social research. Almost any topic can be investigated by asking questions; indeed, this may be the only way to obtain information on some topics such as attitudes, motivations, beliefs, and behaviors known only to the respondent. Unfortunately, self-reported data on such topics may be seriously flawed by respondents' inability or unwillingness to provide the requested information. If respondents are asked to report facts they have forgotten (or perhaps never knew), they may guess or invent answers. If respondents are asked to report facts that are potentially embarrassing or damaging, they may deny or distort what they know to be true. For these reasons, the validity of self-reported data, particularly self-reported drug use, simply cannot be taken for granted.

The National Household Survey on Drug Abuse (NHSDA) is a major source of information on the prevalence and patterns of illegal drug use in this country and the validity of its estimates depend upon the accuracy of the self-reports of respondents. The NHSDA collects information about drug use on self-administered questionnaires during a face-to-face

interview in the respondent's home with the assurance that no one will ever know how they respond. This procedure is designed to reduce the chance that respondents will give answers designed to make a favorable impression on the interviewer or others in the home or to avoid the possibility of negative social or legal sanctions associated with illicit drug use. The assumption is that the assurance of confidentiality and anonymity will offset any potential tendency of the respondent to distort the accuracy of responses.

The study reported here is designed as a criterion validity test of the NHSDA procedures. In criterion validity studies, two different measures of the same trait or experience are available: a candidate measure and an external, independent criterion measure that is treated as an error-free measure of the construct. The use of official records to verify respondent reports has a lengthy history in social science dating back at least to Hyman's classic World War II study of whether the sellers of war bonds, when interviewed, accurately reported this apparently disloyal act (Hyman 1944).

In the current study, the underreporting of illegal drug use was investigated in a sample of 67 former drug treatment clients by comparing their survey responses to clinic records on drug problems at time of admission. The criterion measures are based on self-reported marijuana, cocaine, hallucinogens, and heroin use.

Drug treatment records were obtained from the files of publicly funded drug treatment programs in three States. The study followed the NHSDA interviewing and questionnaire procedures closely. To avoid bias from interviewer expectations and to protect the respondent's privacy, the sample of treatment clients was embedded in a larger sample of respondents. Interviewers were not told that the respondents had been treated for drug abuse. Special sample selection directions, tailored to match the target respondent's age and sex, were used to select the former drug treatment client within the household, simulating the random selection screening instrument used in the NHSDA.

This analysis compares reports of past-year and lifetime use of marijuana, cocaine, hallucinogens, and heroin by the former drug treatment clients to the drugs listed as problematic at time of admission to treatment. This analysis also examines factors that might influence the respondents' willingness or ability to respond accurately, such as the

level of privacy during the interview and the amount of time between admission to the program and the interview.

The limitations of this validity study should be emphasized. Major difficulties in locating respondents based on addresses provided by the drug clinics resulted in a very low response rate. Only 67 eligible respondents were located and interviewed, despite an exhaustive search for former treatment clients. The small sample means differences in the population may not be detected by the analysis. The low response rate means the sample may not be representative of the treatment population. Furthermore, the extent to which reporting by former drug treatment clients resembles reporting by members of the general population of household members is unknown. Treatment may reduce denial of drug use; alternatively, clients in public drug treatment may be more motivated to underreport their past drug use than casual drug users. As a result, the findings of this analysis must be regarded as preliminary and used primarily to illustrate the potential for using official records to assess the validity of drug reports.

USING OFFICIAL RECORDS AS CRITERION MEASURES

Official records have some distinct advantages in the context of validating self-reported illicit drug use. Records based on existing data are relatively inexpensive to obtain. The method is unobtrusive, requiring no additional effort on the part of a survey respondent. One special advantage is the opportunity to check a wide range of questions, such as those about drug-related consequences, that cannot be validated with biochemical tests.

Unfortunately, official records may not always be satisfactory, independent, error-free criterion measures. Records may be incomplete. Definitions and data-collection procedures may vary over time and across locations. Occasional or periodic lapses in data entry, as well as occasional entry errors, can reduce the reliability of record-based data. In many cases, official records are based on self-reports provided in a setting that can affect willingness to report an illegal behavior. For example, arrestees may provide less accurate information on their drug use to intake personnel at a jail than to a researcher on a self-administered questionnaire given with a guarantee of confidentiality. Reluctance to disclose drug use can occur even in an alternative setting such as a clinic (which is expected to be less conducive to underreporting) if the respondent is interested in conveying a positive image to the interviewer.

Records are also limited by the extent to which the individuals and events of interest are included by the mission or catchment area of the agency maintaining the records. Potential differences in drug use beliefs, behaviors, and attitudes between the individuals on whom records are maintained and the population to which the results will be generalized must always be considered. For example, clinic records reflect only the experiences of respondents who sought treatment and qualified for a treatment slot. Criminal justice records on arrests for driving under the influence are affected by factors such as speeding or breaking the law. In such cases, differences in risk-taking, poverty, or other characteristics of those arrested threaten the validity of comparisons to law-abiding citizens. Official records can rarely be used to assess overreporting due to omissions of some events. For example, hospital emergency room records will not include those drug overdose incidents for which respondents failed to seek medical help or went to a doctor's office or clinic. Similarly, only some drug-selling transactions will be reflected in official arrest records of drug dealers.

Despite these potential limitations, official records, used carefully, may be the best available criterion when circumstances or lack of resources prevent the use of a more objective measure of use. This type of criterion measure may be essential when the focus of investigation is the validity of self-reports of drug-related experiences not captured in biochemical tests.

THREATS TO VALIDITY OF SELF-REPORTED DRUG USE DATA: ISSUES AND FINDINGS

Response distortion to avoid social stigma is a serious risk in surveys of value-laden issues such as illicit drug use. Survey respondents may be unwilling to report drug use to avoid adverse reactions from others or to present themselves to the interviewer in a favorable way. Conversely, respondents with positive views of drug use may exaggerate their drug use to impress the interviewer or others, or to live up to a self-image that perceives drug use as positive. These hypotheses are consistent with social desirability theory (Edwards 1957), which suggests that distortion of self-reports, by underreporting or overreporting, occurs as a function of the perceived acceptability of the behavior in question.

Evidence from validity studies with highly reliable and valid external criteria (Cahalan 1968; Hyman 1944; Parry and Crossley 1950; Weiss 1968) indicates that many types of behavior viewed as socially desirable

are overreported, while those viewed as less desirable are underreported. Several studies also indicate that the tendency to underreport varies across social groups that hold differing norms and values regarding the desirability of the behaviors or traits under investigation (Harrell 1985; Hyman 1944; Hindelang et al. 1981; Parry et al. 1971; Philips and Clancy 1970). Thus, even when validity studies indicate a bias towards underreporting a stigmatized behavior, the bias cannot be assumed to be constant across all respondents.

Underreporting has been found to vary by drug, with serious levels of underreporting associated with heroin, the most highly stigmatized drug at the time of these studies. Estimates of heroin prevalence based on indirect methods such as the item count or randomized response techniques that conceal the respondent's answers from the interviewer were higher than those produced by items modeled on the NHSDA, suggesting that the survey respondents underreported on direct questions about heroin use (Miller 1983, 1984). In an earlier study, Cisin and Parry (1980) found that approximately two-thirds of respondents identified as heroin users in clinic records denied heroin use during a survey. In that study, net levels of underreporting appeared to be very low for other drugs such as marijuana and cocaine. While these studies may indicate that only the most undesirable or stigmatized drug behaviors are likely to be underreported, Cisin and Parry noted that the clinic data criterion used in that study was subject to error and that some patients may have inadvertently failed to mention softer drugs such as marijuana during the intake history—thus giving a false degree of net validity to survey reports on the softer drug.

Factors other than social desirability also threaten the accuracy of self-reported drug use data. Respondents may fear legal consequences to reporting drug use if they distrust survey assurances of confidentiality. They may be unable to report drug use accurately, particularly when questions involve detailed accounts of drug consumption at times in the past. They may not be able to remember the circumstances of use, when they used a drug, or even whether they ever used a particular drug. Heavier drug users are likely to find particular facts more difficult to recall and may experience memory impairment.

A number of studies conducted during the 1960s and 1970s compared addicts' reported drug use, arrest record, and demographic information to hospital records, law enforcement records, biochemical tests, and reports of significant others (Amsel et al. 1976; Ball 1967; Cottrell and O'Donnell

1967; Robins and Murphy 1967; Stephens 1972). For example, Ball compared the responses to a structured interview of 59 narcotic addicts to data from hospital records, Federal Bureau of Investigation (FBI) records, and tests conducted immediately after the interview to determine whether "deviant groups, especially those engaged in illegal behavior, are motivated to—and do—conceal or deny their proscribed behavior" (Ball 1967, p. 650). Five items were used for comparison: (1) the age of the subject, (2) age at onset of drug use, (3) type and place of first arrest, (4) total number of arrests, and (5) drug use at the time of interview. Responses to the items related to deviant behavior "indicate a rather surprising veracity on the part of former addicts" (Ball 1967, p. 653). However, recall of detailed information may have reduced the validity of some drug use items. Higher rates of distortion are reported on items that request exact information (e.g., age at first arrest and age of first drug use) (Ball 1967; Cottrell and O'Donnell 1967) than on easier questions such as, "Have you used marijuana?" Because the addicts appeared willing to provide authentic drug information, the implication is that faulty memory produced these inaccurate answers. In general, most research on former addicts concludes that addicts are willing to reveal the facts of their drug use and arrest record. A notable exception is a study by Amsel and colleagues (1976), which found relatively high denial rates for drug use.

Self-reported drug use data may also suffer if reports are inconsistent or incomplete. Analyses of self-reported drug use data collected by the NHSDA have found consistent patterns of self-reported friendship with users of specific drugs, opportunity to use these drugs, and actual use of these substances (Somerville and Miller 1980); and sequential patterns of the first use of various drugs that show a Guttman-like hierarchy of progressive statutes of involvement in drug use, so that for example, virtually all users of cocaine, hallucinogens, and/or heroin report that marijuana use preceded their first use of any these other illicit drugs (Harrell and Wirtz 1980). The convergence of birth cohort data derived from successive NHSDA surveys (see Cisin et al. 1978; Miller and Cisin 1983), as well as the consistency between trends reported by the NHSDA and those from the national surveys of high school seniors (c.f., Johnston et al. 1993; Substance Abuse and Mental Health Administration 1995), support the reliability of survey estimates over time. More recent analyses found that substantial proportions of those reporting some drug use provided at least one inconsistent response on the survey. Although the inconsistencies were most numerous for alcohol, the proportion giving inconsistent responses for illegal drugs was higher (Cox et al. 1992). The complexity of the cognitive task and the demands of recall

associated with questions about time periods may also contribute to measurement error on some NHSDA questions, including those about past-year drug use (Forsyth et al. 1992).

For purposes of the current study, the findings of these earlier studies suggest that former drug treatment clients generally appear willing to report past drug use, but that questions about past-year drug use may be subject to measurement error related to the cognitive complexity and demands on recall of these items.

METHODS

The sample consists of former clients of public treatment programs in Maryland (six programs), New Jersey (six programs), and Pennsylvania (three programs). Clients admitted to treatment between July and December of 1985 were selected for the sample. They were assumed to have used drugs listed as problems at the time of admission during the month before admission. Initially, 600 eligible cases were randomly selected, 50 from each of 12 strata defined by age (12 to 17, 18 to 25, and 26 and older), sex, and race (white and African American) to permit analysis by demographic characteristics. The two strata of young females 12 to 17 (African American and white) were dropped because so few cases were available and many of the available cases had male siblings in the sample. Due to problems in locating sample members from addresses provided in the clinic records, a second sample was selected midway through the study.

The study encountered substantial problems in locating respondents at the addresses provided in the clinic records. Although some problems may have resulted from data-entry errors at the clinic, a larger portion appeared to result from deliberate client misrepresentation. Address verification, undertaken by an independent tracing firm prior to interview assignment, found that 287 of 714 addresses listed in the clinic records did not exist, were out of State (which should have made the clients ineligible for program entry), or referred to vacant lots, office buildings, and other nonresidential structures. At another 243 addresses, the household did not contain anyone in the age/sex group of the former treatment client, despite the fact that residents in the majority of these cases said they had lived at that address for a year or longer. In a smaller number of cases, the household residents were new to the dwelling, so that it is possible that the target respondent had moved from the address. No household roster was obtained

for 52 households (39 refused and 13 could not be contacted), leaving 132 households with potential respondents.

Interviews were completed with respondents in 73 (55 percent) of the 132 households. The selected individual refused interview in 28 households and was never found at home in 31 households. Interviews with six respondents were discarded because they did not match clinic clients on sex and date of birth. The remaining 67 respondents, all of whom had entered treatment within the year of their interview, form the sample used in the analysis. Most had been treated in outpatient programs (96 percent) and in programs focused on drug abstinence (94 percent). Almost half (48 percent) were referred to treatment through a court order.

Respondents were interviewed between April and August of 1986. Only clients who entered treatment less than 1 year before the interview were considered eligible so that their drug use during the month before admission could be considered a measure of past-year drug use.

The survey was designed to duplicate the NHSDA procedures for in-person interviews with randomly selected members of the household population. Respondent selection forms, or screening instruments, are used in the NHSDA to randomly select a member of the household for the interview. For NHSDA, the screening instruments are constructed to disproportionately sample by age group. For the validity study, the forms were modified to select the age and sex of the clinic patient residing at each address, but retained the appearance of a random selection procedure for the respondent and interviewer. The specificity of the selection criteria made it unlikely that the wrong member of a household containing the former clinic patient would be selected, although the final matching criterion was the actual date of birth recorded in both the interview and the clinic record.

Most interviewers (13 of 16) had worked on previous NHSDA surveys. Their training for this survey was similar to that used in earlier surveys. To reduce the chance that interviewers would become aware that respondents had been preselected on the basis of their known drug use, nontreatment households were included in the sample. These households were chosen during address verification on the basis of similarity in location and appearance to households of the former clinic patients and interspersed in the lists of addresses provided to interviewers. In these households, the respondent-selection procedures

resulted in relatively few eligible respondents on the household listing and the results were not used in the analysis.

Interviewing followed the NHSDA procedures for minimizing denial of drug use. Respondents were assured that their answers would be kept private and confidential, never seen by the interviewer or anyone else in the household. Questions about illicit drug use were presented on self-administered answer sheets, sealed in an envelope at the end of the interview, and mailed immediately. Respondents were invited to accompany the interviewer to the mailbox to ensure that the envelope was not opened.

The test of whether respondents reported drug use on the survey was based on answers to the question, "When was the most recent time that you used (the drug)?" Answers were classified as: ever used versus never used; and used in past year versus never used, or used most recently more that 12 months ago. Four classes of illicit drugs—marijuana, cocaine, hallucinogens, and heroin—were examined. Nonmedical use of psychotherapeutic drugs including sedatives, tranquilizers, and stimulants was excluded because so few of the sample clients were admitted to treatment for the abuse of these substances.

Clinic records from the Client-Oriented Data Acquisition Process (CODAP) system maintained by the National Institute of Drug Abuse (NIDA) (1987) provided data on the drugs (up to three) considered problems at the time of admission to treatment. Problems at time of admission were determined at the clinics during intake interviews and were used as the criterion measure of past-year use of these drugs. CODAP forms provided a consistent format across clinics for identifying drugs abused at the time of admission and use of these drugs in the month before admission. Records with missing information on sex, age, race, address, or drug use were excluded. Checks of the reliability of the computerized drug use items contained in the CODAP records found no inconsistencies with hard-copy files maintained at the clinics. Although many respondents reported drugs on the survey that were not listed in the clinic records, this cannot be interpreted as overreporting because no effort was made to list all drugs ever used in the records.

The analysis also examined the validity of self-reported data on drug treatment as reported in the NHSDA. This question is of particular interest in the validity study because all members of the sample were known to have been in treatment so that the validity of the criterion measure is high, and because questions about treatment participation

cannot be validated using biochemical tests. Survey respondents were classified as: ever receiving treatment for the use of drugs other than alcohol versus never receiving treatment; and receiving treatment during the past year versus no treatment in the past year, based on two questions: "Have you ever gotten treatment for your other drug use, not counting cigarettes or alcohol?" and "Have you received treatment in the past 12 months for your drug use (not counting cigarettes or alcohol)?"

RESULTS

The analysis includes 49 respondents with a history of marijuana use, 25 known to have used cocaine, 20 hallucinogen users, 28 heroin users, and 7 psychotherapeutic drug (stimulants, sedatives, or tranquilizers) users. These numbers reflect the fact that most respondents (70 percent) had problems with more than one drug: 48 percent listed two drugs as problems at the time of admission to treatment and 22 percent listed three.

Table 1 describes the sample of abusers by drug category. Eighty-five percent of the respondents were 18 or older, and 72 percent were men. Demographic characteristics varied by type of drug abused. Marijuana and hallucinogen abusers were younger than the samples of heroin and cocaine abusers and included a greater proportion of white respondents.

For most analyses, all 67 cases were used. However, the analysis of self-reported past-year drug use was limited to respondents who used the drug(s) of abuse within the month before admission to the drug treatment program. Cases were limited to respondents interviewed within 11 months of clinic admission (333 days) to ensure that their drug use occurred within the past year at the time of interview. The sample of past-year users consisted of 28 marijuana users, 13 cocaine users, 6 hallucinogen users, and 17 heroin users.

Drug Use Underreporting by Drug

The number and percentage of known users who reported their use on the survey is shown in table 2. Almost all respondents admitted to treatment for marijuana use reported some previous marijuana use (96 percent). Clients treated for cocaine abuse were somewhat less likely to report any past use (84 percent of the users). Reporting accuracy was lower for heroin and hallucinogen use, with 68 to 70 percent of the users reporting any use of these substances. The accuracy

TABLE 1. *Demographic characteristics of respondents by drug abused at time of clinic admission.*

	Drug abused at admission/# affirmative respondents*				
	Marijuana	Cocaine	Hallucinogens	Heroin	Any drug
Total users	49	25	20	28	67
Age group					
12-17	9	2	7	0	10
18-25	19	13	6	12	27
26+	21	10	7	16	30
Sex					
Male	37	20	16	18	48
Female	12	5	4	10	19
Race					
White	26	10	16	7	39
African American	23	15	4	21	28

KEY: * = Drug use categories are not mutually exclusive.

of reporting of past-year use was generally lower. Most of those who had used marijuana in the month before clinic admission (N = 28) admitted past-year use on the survey (86 percent). Fewer of those who used cocaine in the month before treatment reported past-year use (69 percent), and even smaller portions of those whose records indicated use of heroin and hallucinogens in the month before admission reported past-year use. The results are consistent with the thesis that underreporting results from the risk of social stigma associated with revealing use of these drugs. Social desirability theory predicts that response accuracy will decline as the level of stigma increases, so that respondents are expected to be more willing to report the use of widely used drugs such as marijuana than use of those less prevalent and more deviant drugs such as heroin, with cocaine and hallucinogens in the middle.

TABLE 2. *Self-reported drug use by drug type and time period.*

Drug type and time period	Number admitted for abuse	Percent reporting use
Marijuana		
Ever use	49	96
Past-year use	28	86
Cocaine		
Ever use	25	84
Past-year use	13	69
Hallucinogens		
Ever use	20	70
Past-year use	6	33
Heroin		
Ever use	28	68
Past-year use	17	59

An analysis of reporting accuracy of users of more than one drug category was conducted to see whether reporting accuracy within individuals was related to the level of stigma associated with the drug category. This analysis controls for differences in the sample composition of the various drug user categories—differences that could affect reporting accuracy unrelated to drug stigma.

Clinic records indicated that 47 respondents abused two or more drugs at the time of admission. Almost three-quarters of these multiple-drug users (72 percent) reported all drugs used, while 9 percent denied use of all drug categories (table 3). The remaining 19 percent reported some drugs and failed to report others. In every case, the drugs not reported were more stigmatized than those reported. None of the multiple-drug users denied use of a lower stigma drug while reporting use of high-stigma drug. Thus, even respondents with considerable involvement in illicit drug use and exposure to the social norms of drug users are likely to underreport highly stigmatized drugs.

TABLE 3. *Patterns of underreporting drug use among abusers of more than one drug category.*

Reporting pattern	Number	Percent
Denied use of all abused drug categories	4	9
Denied use of all higher stigma drug categories, reported use of lower stigma drug category	9	19
Reported use of a higher stigma drug category, denied use of a lower stigma drug category	0	0
Reported use of all abused drug categories	34	72
Total	47	100

Other Potential Correlates of Drug Use Underreporting

The willingness of respondents to report deviant or socially undesirable behavior may well be influenced by the level of self-disclosure required. It would seem plausible from a social desirability perspective to expect underreporting to be more prevalent in the less private interviews, as reported elsewhere (Bradburn and Sudman 1979; Turner et al. 1992). Although the NHSDA questionnaire procedures are designed to maximize the amount of privacy afforded to respondents, it is sometimes necessary to hold an interview in the presence or hearing of others in the household. According to interviewer rating, 75 percent of the interviews in this survey were conducted under conditions of complete privacy, while 25 percent were conducted in less than totally private circumstances.

The analysis of response accuracy is shown in table 4. Response accuracy appears better under less than total privacy for the higher stigma drugs—cocaine, hallucinogens, and heroin. However, these differences are not statistically significant, possibly because of the small sample size. Respondents may be more willing to report drug use in nonprivate interviews in order to appear truthful to other members of the household who know of their drug involvement. However, the accuracy of reported marijuana use appeared higher under conditions of total privacy, although again this difference was not significant at the 0.05 level.

The willingness to report drug use may also be a function of the length of time since some drug use was initiated, with longer periods of use

TABLE 4. *Self-reported drug use by privacy during interview.*

Reporting pattern	Number admitted for abuse*	Percent reporting use of drug(s) on clinic record
All drugs on clinic record		
Total privacy	47	72
Less than total privacy	16	88
Marijuana		
Total privacy	36	100
Less than total privacy	11	82
Cocaine		
Total privacy	15	80
Less than total privacy	9	89
Hallucinogens		
Total privacy	12	58
Less than total privacy	8	88
Heroin		
Total privacy	18	61
Less than total privacy	6	82

KEY: * = Excludes 4 cases with missing data on privacy.

associated with increased self-acceptance of the drug user identity and label. For the same reason, respondents admitted to drug treatment more than once (43 percent of the sample) might be more willing to report drug use than respondents whose first admission had occurred in the preceding year. A logistic regression analysis was used to test the hypothesis that failure to report one or more drugs (all drug reports accurate versus at least one drug not reported) was a function of the number of years since the first use of an illicit drug. The results found that the length of time the respondent had used drugs was not related to whether all known drug use was reported. Similar analyses found no significant differences in reporting all known drug use between respondents admitted for the first time to drug treatment and those previously admitted.

Clients referred to treatment by the courts might be more likely to underreport drugs on a survey than other clients. They may fear legal

consequences of admitting drug use, have less trust of others, feel greater hostility towards persons questioning them, or resist acknowledging drug involvement. There were significant differences in the percentage reporting all known drug use: 66 percent of court-ordered clients reported use of all clinic-listed drugs compared to 86 percent of those who entered treatment voluntarily ($p < 0.05$).

Although willingness to report use of an illicit drug varies by demographic characteristics that define groups with differences in the level of stigma attached to drug use and perceptions of risk attached to reporting illegal behavior, this hypothesis could not be tested. The only drug use category with enough respondents to permit analysis of sociodemographic correlates of reporting accuracy was use of marijuana, but only 2 of 49 marijuana users failed to report their past use.

Drug Treatment Underreporting by Drug

Although the records indicated that all respondents had entered drug treatment during the year before the survey for use of an illicit drug, many failed to report their drug treatment on the survey. As table 5 indicates, 56 percent of all respondents reported ever receiving any treatment for use of a drug other than cigarettes and alcohol. However, the reluctance to report drug treatment did not increase with the level of stigma associated with the primary drug problem at the time of admission to treatment. Indeed, heroin abusers were slightly, but not significantly, more likely to give accurate reports on drug treatment experiences than were those whose primary drug of abuse was less stigmatized: 71 percent of those whose primary drug was heroin reported ever receiving drug treatment compared to about 50 percent of those whose primary drug was one of the other drugs. Similarly, only 38 percent of this group of clients treated within the past year reported receiving drug treatment during the past year. Again, failure to report past-year drug treatment did not increase with the stigma of the abused substance, and former heroin patients were more likely than abusers of other drugs to report past-year treatment.

Drug Treatment Underreporting by Previous Treatment Episodes

Willingness to report drug treatment was related to the number of treatment episodes. Compared to respondents with more than one

TABLE 5. *Reporting of drug treatment experience by primary drug at admission and time period.*

Primary drug and time period	Total number*	Percent reporting treatment
Marijuana		
Treatment ever	17	47
Treatment past year	17	29
Heroin		
Treatment ever	21	71
Treatment past year	21	52
Other drugs		
Treatment ever	28	50
Treatment past year	28	32
Total sample		
Treatment ever	66	56
Treatment past year	66	38

KEY: * = Excludes one case with missing data on treatment experience.

treatment experience, respondents who were admitted to treatment for drug abuse for the first time during the preceding year were less likely to report ever receiving drug treatment (43 percent compared to 72 percent, chi square = 5.6, $p < 0.05$) and less likely to report receiving drug treatment in the past year (36 percent compared to 68 percent, chi square = 6.6, $p < 0.05$). More than one treatment episode was more prevalent among the heroin abusers in the sample than among others: 76 percent of the heroin admissions had previously received drug treatment compared to 6 percent of the marijuana admissions and 41 percent of those admitted for other primary drugs. Thus, the slightly better accuracy of the drug treatment data of those admitted with heroin as the primary drug problem may be associated with multiple treatment episodes.

Other Potential Correlates of Drug Use Underreporting

Willingness to report drug treatment showed no significant relationship to other potential correlates of underreporting, including the privacy of the interview (completely private compared to less than completely

private), the source of referral to treatment (court ordered compared to voluntary), and the time between entering treatment and the interview. Similarly, differences by age, sex, and race in the percentage reporting drug treatment in the past year or at any time in the past (table 6) were not statistically significant.

TABLE 6. *Reporting of receiving drug treatment by demographic characteristics.*

Demographic group	Total number*	Percent reporting treatment during past year	Percent reporting ever receiving treatment
Age			
12-17	9	33	67
18-25		30	44
26+		47	63
Sex			
Male	47	36	57
Female	19	42	53
Race			
White	32	44	56
African American	34	32	56
Total sample			
Treatment ever	66	56	56
Treatment past year	66	38	38

KEY: * = Excludes one case with missing data on treatment experience.

These results suggest that drug treatment is an experience that respondents from diverse social groups are reluctant to report in an interview, even under conditions designed to protect their anonymity and confidentiality.

DISCUSSION

The results indicate that underreporting of drug use increased as the social stigma associated with the drug increased. Most (more than

80 percent) of the former drug treatment clients interviewed using the NHSDA procedures reported ever using marijuana and cocaine when these drugs were listed as problems at the time respondents were admitted to drug treatment. A smaller portion, but still over two-thirds, of those whose clinic records indicated problems with hallucinogens and heroin reported ever using these drugs. More than 80 percent of the known marijuana users reported their past-year use, more than two-thirds of the known cocaine users reported their past-year use, and less than two-thirds of the hallucinogen and heroin users reported past-year use. The within-user analysis shows that in every instance, mixed reporting accuracy errs on the side of failing to report: Known cocaine users reported their past-year use and fewer than two-thirds of the hallucinogen and heroin users reported past-year use. The within-user analysis shows that in every instance, mixed reporting accuracy errs on the side of failing to report more stigmatized drugs. The lower rates of past-year use may result from a combination of failure to recall the time of most recent use accurately and a reluctance to admit more recent drug use.

Reporting accuracy did not vary significantly by the privacy of the interview, the number of years of drug use, or whether the respondent had one or more drug treatment episodes. These findings should be interpreted with caution due to the small sample size. However, those former clients who entered drug treatment under court order were less likely to report their drug use accurately.

Past-year drug treatment should have been reported by 100 percent of the sample, but was reported by less than 40 percent of those known to have been admitted to treatment in the past year for drugs other than heroin. Failure to report treatment was not correlated with the level of stigma attached to the drugs listed as problems at the time of admission, with heroin abusers more likely to report past-year treatment than those admitted to treatment for the abuse of other drugs. Clients who had received drug treatment more than once were more likely to report any previous drug treatment and drug treatment in the past year than clients who had been in treatment only once. Since heroin users were more likely to have multiple treatment episodes, the tendency to underreport events associated with this stigmatized drug may have been offset by reduced denial associated with multiple treatment episodes. The privacy of the interview, court referral, and demographic characteristics were not related to the reporting accuracy. There was no significant relationship between reporting accuracy and the time elapsed between clinic admission

and the interview, suggesting that memory failures did not play a substantive role in the underreporting.

The difficulty in locating respondents and the resulting low response indicate caution in generalizing these results to the population of former drug treatment clients. However, the effect of the bias introduced by the survey nonresponse may be to reduce the level of observed underreporting, if it can be assumed that those former clients who provided incorrect addresses to the clinic would also be less truthful in reporting their drug use and drug treatment on an interview. The small sample size also limits the power of the analysis to detect significant differences, suggesting that future study is indicated of factors found to be unrelated to underreporting.

The study illustrates both strengths and weaknesses of records-based validity studies of self-reported drug data. The validity test could be conducted in a natural setting that avoided bias introduced by interviewer expectations or the realization on the part of respondents that underreporting would be detected. The research method did not require any special effort on the part of respondents, nor the expense of special tests. The criterion in this case was found to be reliable, since checks against clinic records showed that clinic files contained information on the use of the drugs that were not reported.

The disadvantages to record-based validation are also clearly demonstrated. The clinic records on address location were very inaccurate, resulting in a poor rate of locating sample members. As a result, the extent to which interviewed former treatment clients are representative of the population of former drug treatment clients is unknown. Further caution is required in generalizing the results to the household population included in the NHSDA. Underreporting may be less prevalent among those who have received treatment because their drug use is not a secret and they have had to discuss it in interviews with clinic personnel, while those who have not previously discussed their use may be more motivated to conceal their drug use. However, the opposite may true. Denial is known to be a problem among serious abusers, while casual users may be less likely to regard their use as a problem to be denied.

One of the two criterion measures, drug problems reported at the time of admission, must be viewed with some caution because it is based on self-reported data, albeit self-reports collected in a setting likely to produce

accurate information. The level of underreporting on this criterion is unknown. Some of the discrepancies in past-year drug use may have resulted from errors in reporting past-month use at the time of clinic admission or, as noted above, from errors in recalling the recency of use. In contrast, the clinic records provided a very robust basis for a test of the validity of reports on drug treatment because all respondents were known to have received treatment, a population of treatment clients (limited to public treatment facilities in three States) served as the sampling frame, and alternative procedures are not available for verifying the validity of responses on this consequence of drug use. Future records-based validity tests must be undertaken with careful assessment of these issues as they relate to the specific self-report data and set of records to be compared.

REFERENCES

Amsel, Z.; Mandell, D.; and Matthias, C. Reliability and validity of self-reported illegal activities and drug use collected from narcotic addicts. *Int J Addict* 11(2):325-336, 1976.

Ball, J. The reliability and validity of interview data from 59 narcotic drug addicts. *Am J Sociol* 72:650-654, 1967.

Bradburn, M.N., and Sudman, S. *Improving Interview Methods and Questionnaire Design.* Washington, DC: Jossey-Bass, 1979.

Cahalan, D. Correlates of respondent accuracy in the Denver validity survey. *Public Opin Q* 32:607-621, 1968.

Cisin, I., and Parry, H.J. Sensitivity of survey techniques in measuring illicit drug use. In: Rittenhouse, J.D., ed. *Developmental Papers: Attempts to Improve the Measurement of Heroin Use in the National Survey.* Rockville, MD: National Institute on Drug Abuse, 1980.

Cisin, I.H.; Miller, J.D.; and Harrell, A.V. *Highlights from the National Survey on Drug Abuse: 1977.* Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1978.

Cottrell, E.S., and O'Donnell, J.A. "Reliability of Admission Data." Report prepared for the National Institute of Mental Health Clinical Research Center, Lexington, KY, 1967.

Cox, B.G.; Witt, M.B.; Traccarella, M.A.; and Perez-Michael, A.M. Inconsistent reporting of drug use in 1988. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* National Institute on Drug Abuse Pub. No. 271-88-8310. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 109-153.

Edwards, A.L. *The Social Desirability Variable in Personality Assessment and Research.* New York: Dryden, 1957.

Forsyth, B.; Lessler, J.; and Hubbard, M. Cognitive evaluation of the questionnaire. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* National Institute on Drug Abuse Pub. No. 271-88-8310. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 13-52.

Harrell, A.V. Validation of self-report: The research record in self-report methods of estimating drug use. In: Rouse, B.E.; Kozel, N.J.; and Richards, L.G., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity.* National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985.

Harrell, A.V., and Wirtz, P. "Developmental Sequences of Illicit Drug Use." Paper prepared for National Institute on Drug Abuse, 1980.

Hindelang, M.J.; Hirschi, T.; and Weiss, J.G. *Measuring Delinquency.* Beverly Hills, CA: Sage Publications, 1981.

Hyman, H. Do they tell the truth? *Public Opin Q* 8:557-559, 1944.

Johnston, L.D.; O'Malley, P.M.; and Bachman, J.G. *National Survey Results on Drug Use from the Monitoring the Future Study, 1975-1992.* Rockville, MD: National Institute on Drug Abuse, 1993.

Miller, J.D. "A New Survey Technique for Studying Deviant Behavior." Ph.D. diss., George Washington University, 1984.

Miller, J.D. "The Nominative Technique: Method and Heroin Estimates from the 1982 National Survey on Drug Abuse." Paper prepared for the National Institute on Drug Abuse, 1983.

Miller, J.D., and Cisin, I.H. *Highlights of the National Survey on Drug Abuse.* Washington, DC: Supt. of Docs., U.S. Govt. Print Off., 1983.

National Institute on Drug Abuse. *Demographic Characteristics and Patterns of Drug Use of Clients Admitted to Drug Abuse Treatment Programs in Selected States: Annual Data 1983.* Washington, DC: U.S. Department of Health and Human Services, 1987.

Parry, H.J., and Crossley, H.M., Validity of responses to survey questions. *Public Opin Q* 14:61-80, 1950.

Parry, H.J.; Balter, M.B.; and Cisin, I.H. Primary levels of under-reporting psychotherapeutic drug use. *Public Opin Q* 34:582-592, 1970-1971.

Philips, D.L., and Clancy, K.J. Response bias in field studies of mental illness. *Am Sociol Rev* 35:503-515, 1970.

Robins, L.N., and Murphy, D.E. Drug use in normal population of young Negro men. *Am J Public Health* 57:1580, 1967.

Somerville, S.N., and Miller, J.D. "Opportunity and Deviance: An Analysis of Drug Use Entry." Paper prepared for the National Institute on Drug Abuse, 1980.

Stephens, R. The truthfulness of addict respondents in research projects. *Int J Addict* 7(3):549-588, 1972.

Substance Abuse and Mental Health Services Administration. *National Household Survey on Drug Abuse: Main Findings 1992.* DHHS Pub. No. (SMA) 94-3012. Rockville, MD: Substance Abuse and Mental Health Services Administration, 1995.

Turner, C.; Lessler, J.; and Devore, J. Effects of mode of administration and wording on reporting of drug use. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. National Institute on Drug Abuse Pub. No. 271-88-8310. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 177-220.

Weiss, C. Validity of welfare mothers' interview responses. *Public Opin Q* 32:622-633, 1968.

AUTHOR

Adele V. Harrell, Ph.D.
Director
Program on Law and Behavior
Urban Institute
2100 M Street, NW
Washington, DC 20037

147342

The Recanting of Earlier Reported Drug Use by Young Adults

Lloyd D. Johnston and Patrick M. O'Malley

ABSTRACT

One approach to determining the validity of self-reported drug use measures is to examine the extent of logically inconsistent responses over time. Because lifetime use logically should never decline, the rate of subsequent recanting of earlier reported lifetime use provides relevant evidence on validity. In this chapter, recanting rates are examined in nationally representative samples of high school seniors (18-year-olds) surveyed in the Monitoring the Future study as they are followed up on seven occasions through age 32. For the illegal drugs examined (marijuana, cocaine, and lysergic acid diethylamide (LSD)), recanting rates prove to be quite modest, but for the psychotherapeutic drugs, they were more substantial, possibly because of their greater definitional ambiguity. In general, there were no large individual differences in recanting rates as a function of sex, household composition, community size, or education level. Consistent with previous work, minorities (particularly African Americans) had somewhat higher rates of recanting on the illegal drugs. So did respondents in certain occupations, namely, the military and police/firefighting. In general, however, the evidence is quite good for validity of self-reported (by mail) lifetime use of the illegal drugs in young adulthood.

INTRODUCTION

This chapter addresses research issues of concern to those collecting or interpreting self-report data on illicit drug use: the extent to which young adults recant earlier reported drug use in subsequent followup surveys, the extent to which such recanting varies by type of drug and type of respondent, and the extent to which the findings have implications for interpreting prevalence rates for cross-sectional studies of adults.

Even when recanting—the denial of earlier reported use—occurs, the issue of interpretation remains. Recanters may knowingly or unknowingly change their answers. Perhaps it should not be assumed

that the earlier answers are the more valid; later ones may reflect improved respondent understanding and actually may be the more accurate. For example, uncertainty in accurately characterizing some substances that may have been taken at earlier ages could prove to be a factor in recanting.

BACKGROUND

Like a great many studies in the drug field, the Monitoring the Future study relies on self-report measures of drug use to make prevalence and trend estimates on large segments of the population (Johnston et al. 1991, 1995). A strong case has been made for the reliability and cross-time stability of a number of the measures (O'Malley et al. 1983), and for their validity in the context of cross-sectional school-based surveys (Johnston et al. 1991, 1995; Johnston and O'Malley 1985; Wallace and Bachman 1993). On the other hand, some intriguing findings showing some degree of recanting at later points in time have been reported, based on panel studies of respondents initially questioned in high school (Johnston et al. 1995). So far, these data have been presented only in a cross-sectional format based on several contiguous cohorts who received questionnaires at the same point in time. Because important secular trends in drug use have been occurring as these various class cohorts have passed through high school, it is important to distinguish whether age differences in recanting rates (older cohorts have higher rates) are a function of cohort or of aging. If, indeed, there is an age effect in recanting rates, the question remains of how seriously it biases the prevalence estimates for different age groups, and of whether national cross-sectional surveys of drug use in the general population, such as the National Household Survey on Drug Abuse (Substance Abuse and Mental Health Services Administration (SAMHSA) 1995) might have serious underreporting biases.

The availability of panel data initially collected in adolescence provides an opportunity to address these questions. Some questions will not be answered definitively, because the panel data available were not gathered using an experimental design that would be needed to distinguish among some possible explanations. For example, recanting demonstrated in the Monitoring the Future followup panels might be explained by either a change in situation or a change in age. All respondents shift from an in-school self-administered survey given in a classroom to a self-administered survey sent by mail that usually is

completed in the home. At the same time, respondents are undergoing important developmental transitions as they mature, including the attainment of higher education, full-time employment, marriage, and parenthood.

Only a few other researchers, using interview methods, have investigated the recanting phenomenon in longitudinal studies of illicit drug use. Fendrich and Vaughn (1994) investigated recanting rates between the 1984 and 1988 waves of the National Longitudinal Survey of Youth. The most consistent demographic correlates of recanting of marijuana and cocaine use were race/ethnicity and educational status. Minority respondents (particularly African Americans) and dropouts were more likely to recant, even after controlling for other demographic variables. Earlier, Mensch and Kandel (1988) also reported that minority respondents in the National Longitudinal Survey of Youth were more likely than nonminorities to recant their previous reports of lifetime marijuana use (based on the 1980 and 1984 waves of data), even after controlling for educational status.

METHODS

Samples

All of the data presented in this chapter derive from the Monitoring the Future study, which included among its various design features large, nationally representative cross-sectional surveys of high school seniors each year beginning in 1975 (N = 16,000 to 18,000 annually). Representative subsamples of 2,400 target respondents have been selected from each graduating class cohort to comprise followup panels. These panels each receive seven followup surveys at 2-year intervals, with a random split-half sample receiving questionnaires in odd-numbered years and the other split-half receiving questionnaires in even-numbered years. For the current analyses, the split-halves were combined.

The analyses presenting cross-sectional statistics for respondents in the age range 18 to 32 in 1993 are based on approximately 16,300 respondent cases at age 18 (the seniors surveyed in school that year) and 8,900 weighted cases in the age range 19 to 32, all of whom were surveyed by mail in 1993. In the panel analyses of the several adjacent cohorts for whom data through age 32 are available (i.e., the classes of 1977 to 1979), the findings are based on approximately 5,300 weighted cases,

which falls to 4,500 in the first followup (reflecting an 85 percent retention rate at ages 19 to 20) and then gradually to 3,500 by the seventh followup (a 66 percent retention rate at ages 31 to 32).

Field Procedures

The data from all senior classes were gathered using a 40-minute self-administered questionnaire distributed to classrooms of students by University of Michigan Survey Research Center interviewers. Identifying information for followup was gathered on tear-off cards that could not be connected with the respondent's questionnaires except through the use of randomly matched identifying numbers; these numbers in turn can be connected only through the use of a special computer file maintained under security at the University of Michigan. Similar random-number identifiers were used on the followup questionnaires, which contained no other individual identification when they were mailed back to the Survey Research Center.

The followup questionnaires were sent by certified mail to all panel respondents except those previously declining to be in the panel, accompanied by a check for $5 payable to the respondent ($10, beginning with the class of 1992). Respondents also received a self-addressed, postage-paid return envelope and a description of the confidentiality protection procedures. Roughly 6 to 9 months after each followup survey, respondents received a newsletter from the study, which also thanked them for their continued participation.

Measures

The variables used in the analyses presented here were measured on all respondents. That is, they were common to all questionnaire forms, even though five or six different questionnaire forms are used with these age groups to permit the inclusion of a great many more variables in the study than could be contained in a single form. This chapter discusses the prevalence of use of five different drugs: marijuana, cocaine, LSD, tranquilizers, and barbiturates. A common measurement format was used for all of these substances by asking respondents: "On how many occasions (if any) have you used [drug]..." followed by three time periods: "...in your lifetime," "...in the past 12 months," and "...in the past 30 days." The respondent was given a 7-point frequency scale to answer separately for each of the three time periods. Nearly all analyses

reported here focus on prevalence, rather than frequency, rates—that is, whether the respondent used the drug at all in each of the time periods.

It is only with respect to lifetime use that it is possible to determine that a respondent has recanted; accordingly, most of the discussion here deals with lifetime prevalence. In a way this is unfortunate, because the annual and 30-day prevalence rates are probably more important for policy purposes. However, because annual and 30-day prevalences actually can decline over time, recanting cannot be determined for these rates (using the available measures), whereas lifetime prevalence logically cannot decline. (It would be possible, of course, to utilize retrospective reports in the followup surveys of use in earlier periods—for example, during senior year of high school—to allow for some estimate of recanting, but the study does not include such questions.)

FINDINGS

Results relevant to recanting rates are presented for five drugs in figures 1 through 5; within each of these figures, three graphics (A, B, C) are provided. The top panel (A) in each figure displays the lifetime and annual prevalence rates for different age groups surveyed cross-sectionally in 1993 (using the single measurement taken in 1993). Adjusted lifetime prevalence rates are also provided in this graphic; they include respondents who had twice previously reported use of the drug, but who did not in the 1993 survey (the figures are presented in volume II of the study's annual monographs).

The middle graphic (B) in each figure displays lifetime and annual prevalence rates at different ages for a panel of several adjacent class cohorts surveyed eight times between ages 18 to 32 (the graduating classes of 1977, 1978, and 1979). Panel B gives a truer measure of change as a function of age—at least for these cohorts—than does the simultaneous cross-section of different age groups. Panel B also contains an adjusted lifetime prevalence rate, again correcting for respondents who had twice previously reported use of the drug but who did not do so when surveyed at the age indicated.

The bottom graphic (C) for each drug is based on the same panel of respondents as in B; it differs only in the criterion used to show recanting

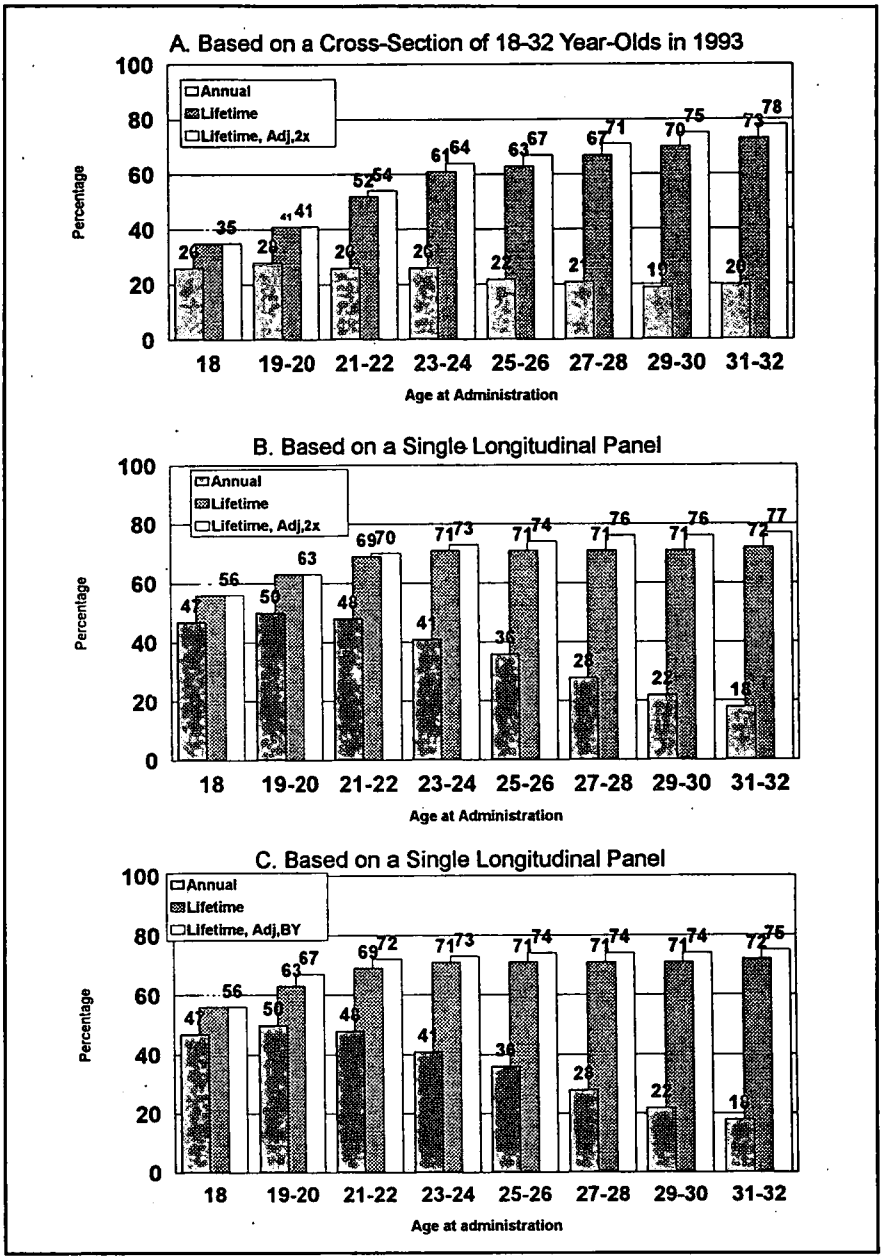


FIGURE 1. *Marijuana use.*

KEY: The adjusted bars in the two top graphs are based on those who had twice previously reported use; the bottom graph is based on only one previous report of use, specifically in the 12th grade.

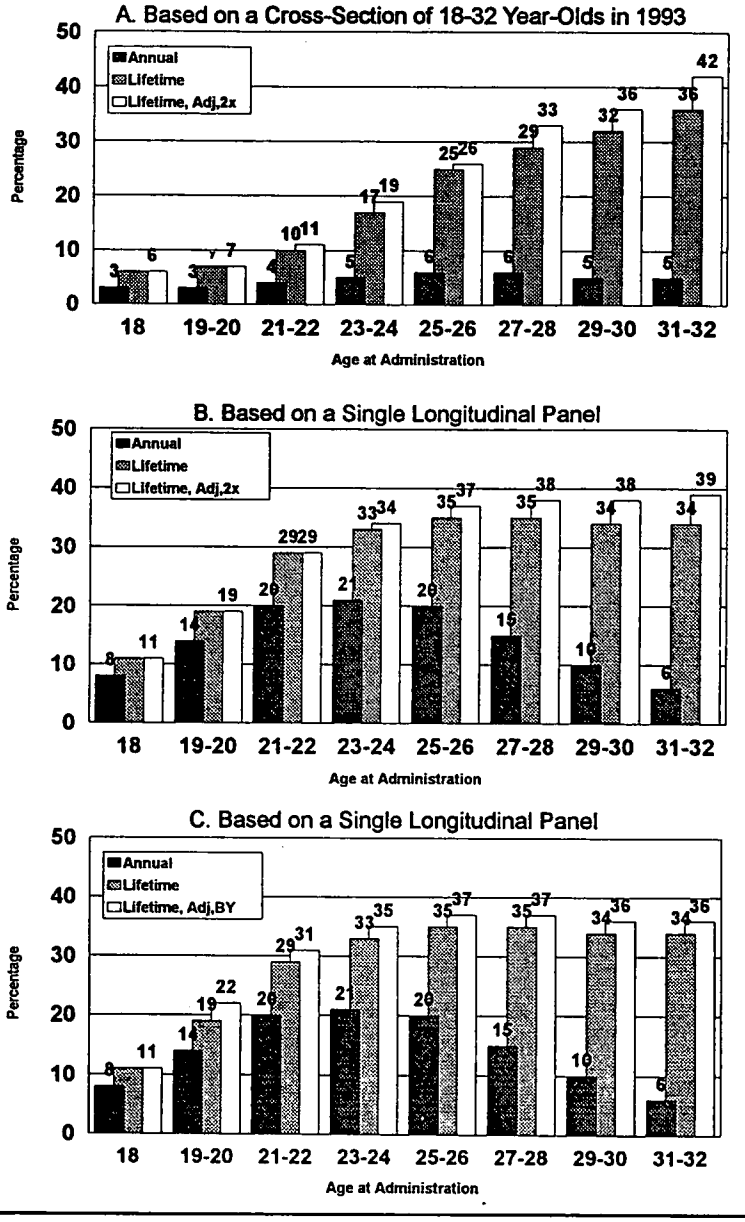


FIGURE 2. *Cocaine use.*

KEY: The adjusted bars in the two top graphs are based on those who had twice previously reported use; those in the bottom graphs are based on only one previous report of use, specifically in the 12th grade.

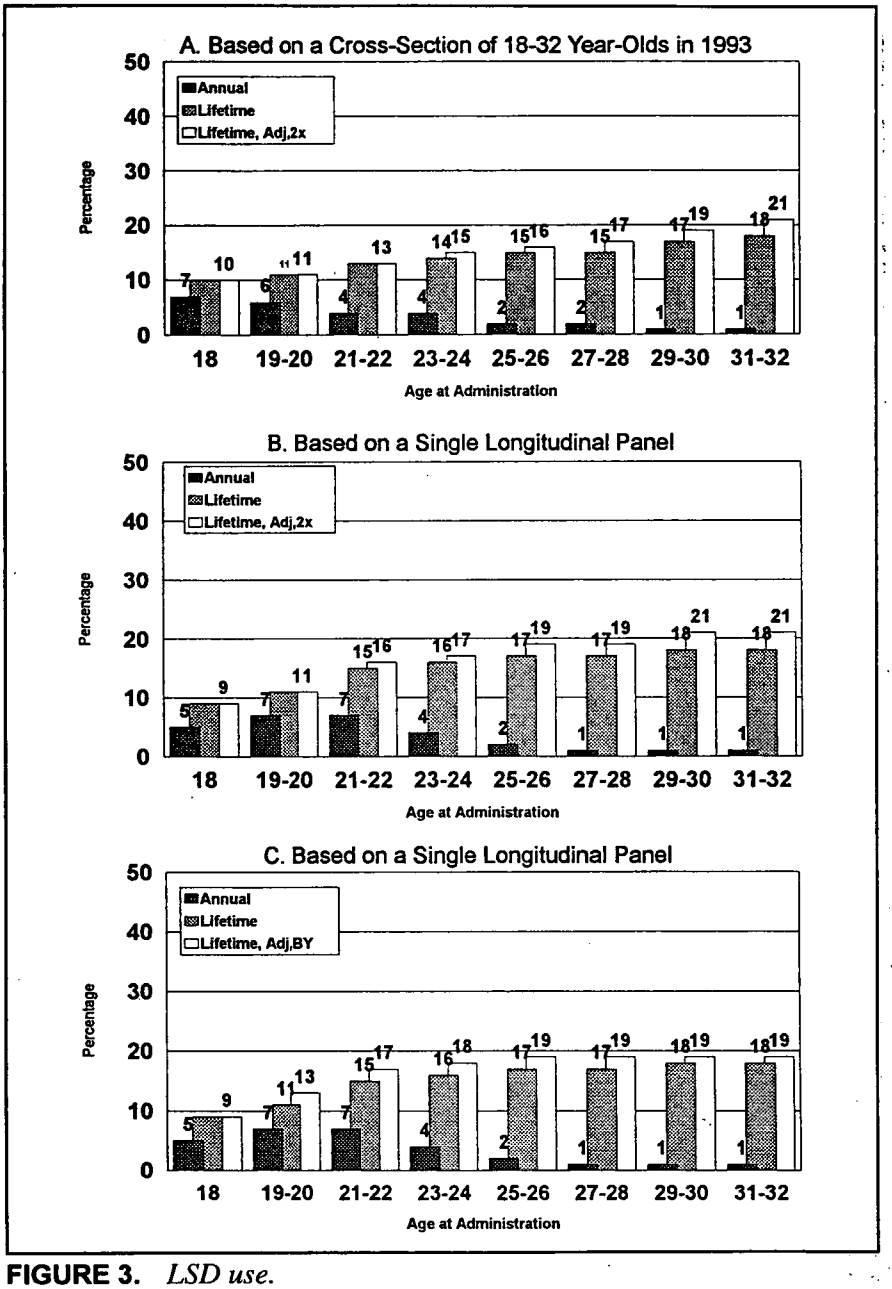


FIGURE 3. *LSD use.*

KEY: The adjusted bars in the two top graphs are based on those who had twice previously reported use; those in the bottom graphs are based on only one previous report of use, specifically in the 12th grade.

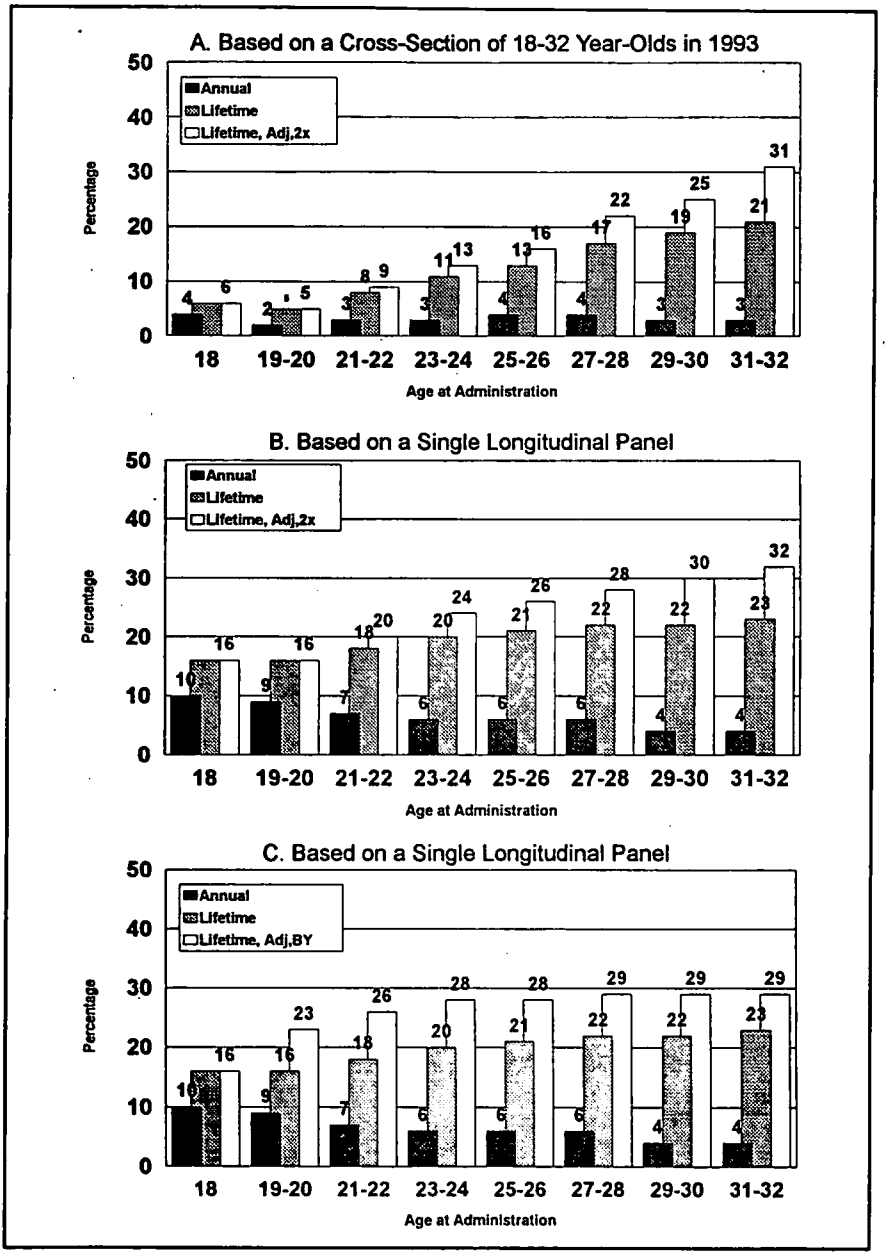


FIGURE 4. *Tranquilizer use.*

KEY: The adjusted bars in the two top graphs are based on those who had twice previously reported use; those in the bottom graphs are based on only one previous report of use, specifically in the 12th grade.

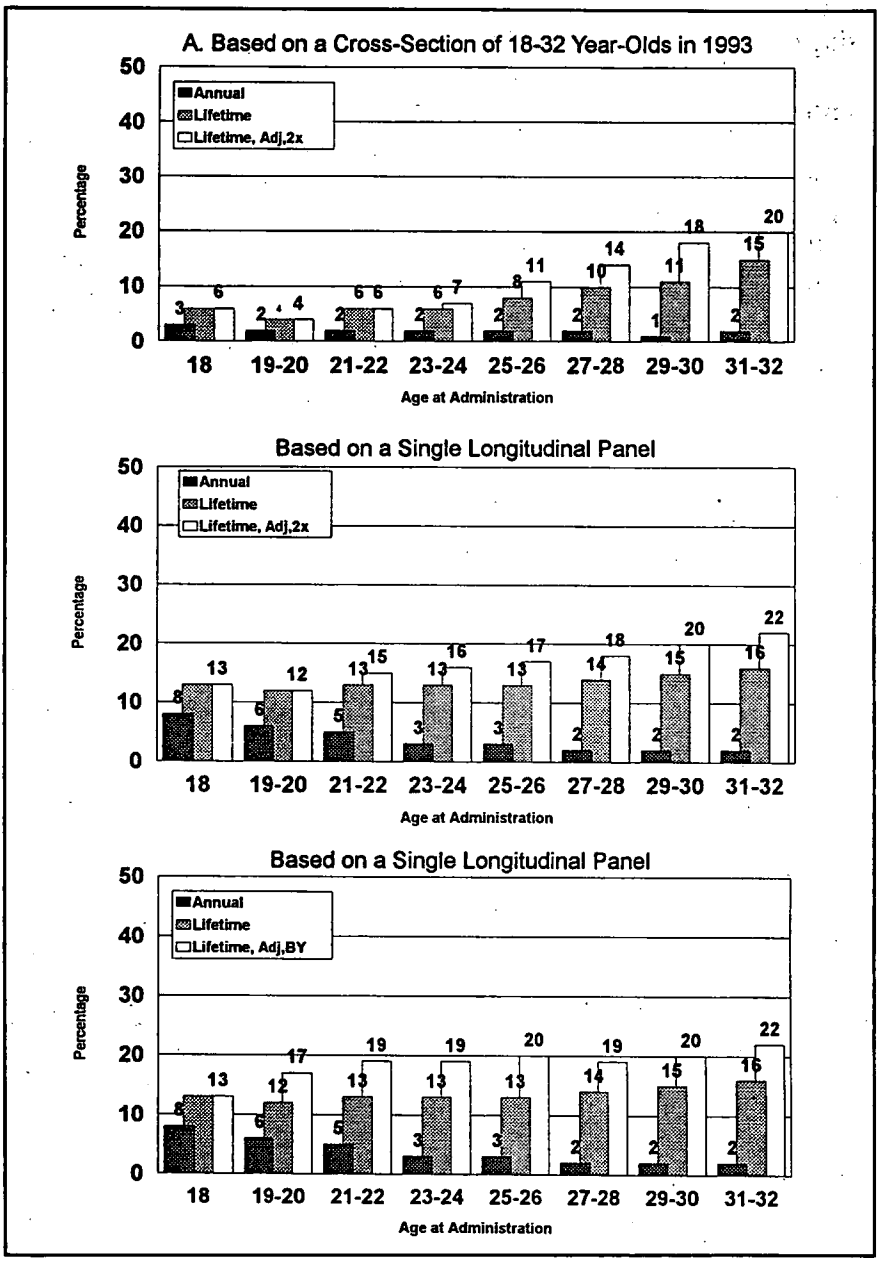


FIGURE 5. *Barbiturates use.*

KEY: The adjusted bars in the two top graphs are based on those who had twice previously reported use; those in the bottom graphs are based on only one previous report of use, specifically in the 12th grade.

at each followup. In this case, recanting is defined as denying use after a single previous mention, specifically in high school.

The Illegal Drugs

Figure 1 presents these three data sets for marijuana, figure 2 for cocaine, and figure 3 for LSD. In all three cases, it may be seen that while lifetime prevalence rises with age, whether age is examined cross-sectionally or longitudinally, annual prevalence generally tends to decline with age. The increase in lifetime prevalence is particularly sharp for cocaine until respondents reach their midtwenties.

For both marijuana and cocaine, the cross-time age profile for annual prevalence is quite different when the panel data are used than when cross-sectional data are examined. The panels show a much sharper decline with age in past-year use because important downward secular trends were occurring in the use of both of these drugs in the 1980s. That is, 18-year-olds in the late 1970s (the cohorts used in the panel analyses) had much higher rates of marijuana and cocaine use than did 18-year-olds in 1993 (figures 1 and 2). Thus the declines with age in annual prevalence observed in these data exaggerate the true age effect because downward secular trends were also occurring as the panels aged and contributed to the steep decline with age. (See O'Malley et al. 1988 for an empirical estimation of the magnitude of these two effects.)

Because both the middle (B) and bottom (C) graphics are based on the same panel, their annual and unadjusted lifetime prevalence rates are identical. However, the adjusted lifetime prevalence rates differ somewhat because of different methods of adjustment: the middle graphic (B) counts a respondent at a given age as a user if that respondent twice previously reported use of the drug, even if the respondent did not report any lifetime use when surveyed at the given age. By way of contrast, the bottom graphic (C) counts a respondent as a user if that person reported use on just one previous occasion (specifically, in the senior year of high school), and shows slightly different adjusted prevalences. Corrections for recanting obtained with either correction method, however, are not very large.

For marijuana, adding the recanters of two previous mentions of use contributes only 5 percentage points to the panel estimates at age 27 to 28 (76 percent versus 71 percent), and the differential appears to remain unchanged thereafter. Correcting specifically for recanting of a mention

of use in 12th grade adds 3 to 4 percentage points to the lifetime estimate, beginning with the first followup after high school and continuing thereafter. Because the unadjusted prevalence rates are so high, these adjustments to the estimates of lifetime prevalence are extremely modest.

Corrections of a similar magnitude occur for the lifetime prevalence estimates for cocaine. Figure 2 shows that the size of the adjustment grows in the late twenties and early thirties, perhaps because the level of lifetime use rises considerably when respondents pass through their early twenties and, therefore, considerably more self-admitted users are able to recant. The adjustment rises slowly, from 1 percentage point at ages 23 to 24, to 5 percentage points by age 31 to 32 (raising the lifetime prevalence estimate from 34 to 39 percent). As figure 2 illustrates, the recanting of cocaine use first reported in the senior year of high school does not increase with age. The 2 percentage point adjustment from age 21 to 22 remains the same through age to 31 to 32.

A similar picture emerges for LSD. The panel B data in figure 3 show a gradually increasing adjustment with age, although a small one, growing from no adjustment at age 19 to 20 and rising to 3 percentage points by age 31 to 32 (raising the lifetime prevalence estimate from 18 to 21 percent). Lifetime experience with LSD rose considerably when those respondents were in their early twenties, as with cocaine, making more people available to recant reported use. Figure 3 shows that the recanting of LSD use first reported in high school did not increase with age, as it did with cocaine. The corrections at all ages in figure 3 amount to only 1 or 2 percentage points.

The Psychotherapeutic Drugs

A number of prescription-controlled classes of drugs were included in the study, but, in the interest of brevity, the authors chose tranquilizers and barbiturates for illustrative purposes. The same three graphics are presented for each of these drugs as for the illegal drugs (see figures 4 and 5). It is immediately apparent in these figures that the adjustments in the lifetime prevalence rates for these drugs are larger than for the illegal drugs, in both absolute and proportional terms. For tranquilizers, by age 31 to 32 the adjustment for recanting twice previously mentioned use adds 9 percentage points to the lifetime prevalence rate, and increases the unadjusted estimate of 23 percent to an unadjusted level of 32 percent. This correction increases with age, but figure 4 indicates that this is

primarily due to the recanting of use originally reported in high school. As early as the first followup at age 19 to 20, 7 percent of the cohort recanted tranquilizer use that they had reported in the senior year. This correction is relatively unchanged thereafter.

The picture is much the same for barbiturates. Figure 5 shows a growth with age in the recanting correction, reaching 6 percentage points by age 31 to 32, when the unadjusted lifetime prevalence rate was 16 percent, and the adjusted prevalence 22 percent. Figure 5 shows that most of this is again due to recanting the use first reported in high school. The correction also reaches 6 percentage points by age 31 to 32.

Subgroup Differences

Based on the literature cited above, the authors expected to find that race/ethnicity would be related to likelihood of recanting, as would the level of education attained (although the absence of dropouts in the current study would reduce the strength of the relationship). It was reasoned that respondents with higher levels of education should have higher comprehension levels and be more likely to complete their questionnaires carefully, resulting in less recanting. The authors further expected that occupational status level would be related positively to recanting, because respondents in higher status occupations would have more to lose if their drug use were exposed, and thus greater motivation to conceal it. In addition, the authors considered it likely that job setting and household composition might relate to rate of recanting. Specifically, it was expected that those in workplace settings that were least tolerant of drug use or more likely to invoke severe sanctions if drug use were revealed, would recant more often because of intentional concealment. The military and police were expected to meet this criterion. Finally, the authors expected that those living in households containing people (for example, the respondent's children and spouse) from whom the respondent might wish to conceal use would have higher recanting rates. Because these respondents might be concerned about these people seeing their answers, they would be more likely to conceal their use.

The following specific hypotheses were tested: (1) African-American respondents would have higher rates of recanting than white respondents; (2) those living with children or spouses would have higher rates of recanting than those who did not; (3) the more educated would have lower rates of recanting than those less educated; (4) those in high-status occupations would have higher rates of recanting; and (5) those in

workplace settings least tolerant of drug use would have higher rates of recanting.

Other demographic variables were chosen for these analyses without any particular hypotheses about the likely outcome—specifically, gender and community size.

Table 1 presents the recanting rates for different subgroups on reported use of marijuana, cocaine, LSD, tranquilizers, and barbiturates. Recanting rates are based on the data from the fourth followup, at ages 25 to 26, of the graduating classes of 1977 to 1987. The data from the fourth followup were used instead of those from longer term followups to increase the number of cases in the subgroups. Even so, the number of cases are somewhat limited in some of the subgroups discussed, and the reader is cautioned to note the weighted number of cases provided in table 1. Recanting rates are calculated as the difference between the adjusted and unadjusted lifetime prevalence rates, divided by the adjusted lifetime prevalence rates; in other words, the proportion recanting stated as a proportion of all of those who reported use at the time in question plus any who twice previously reported use but did not report use at the time in question.

The differences in recanting rates among categories on each variable were tested for significance using a chi-square analysis, which ignores ordinality in the relationship. The level of significance is shown under the recanting rate values for each set of answer categories in table 1.

Table 1 illustrates that some degree of recanting is found across virtually all of the subgroups, and in general the degree of recanting does not vary greatly from subgroup to subgroup. In sum, this appears to be a phenomenon which does not evidence great subgroup differences. The recanting differences between genders on marijuana and cocaine use are nonsignificant. They are fairly small, but statistically significant, on LSD, tranquilizers, and barbiturates.

Consistent with the first hypothesis, race/ethnicity shows some relatively large differences in recanting rates, with African Americans recanting more than white respondents for both marijuana and cocaine use.

Consequently, the prevalence rates for whites and African Americans for these drugs are more divergent on the unadjusted lifetime prevalence than after adjustment for recanting has occurred. Race/ethnicity

TABLE 1. *Lifetime prevalence of five drugs at the fourth followup by subgroups, classes of 1977-1987.*

				Marijuana			Cocaine		
		N(Wtd)	N(Unwtd)	Life-time, adjusted	Life-time	*R.R.*	Life-time, adjusted	Life-time	*R.R.*
Total:		13237	17276	70.4	66.9	*+0.051*	32.3	29.9	*+0.074*
Sex:									
Males		5894	7785	72.6	68.8	*+0.052*	37.4	34.8	*+0.070*
Females		7343	9491	68.7	65.3	*+0.049*	28.2	26.0	*+0.078*
Race:									
White		11070	14655	71.4	68.3	*+0.043*	33.6	31.4	*+0.065*
Black		1125	1280	66.2	58.8	*+0.112*	22.2	19.0	*+0.144*
Other		1077	1388	65.0	61.1	*+0.060* **	30.1	27.0	*+0.103* **
Household composition:									
Live with spouse:	yes	5811	7495	67.3	63.4	*+0.058*	26.2	23.6	*+0.099*
	no	7405	9749	72.8	69.6	*+0.044* **	37.1	34.9	*+0.059* **
Live with partner:	yes	1342	1923	83.3	79.8	*+0.042*	48.6	45.9	*+0.056*
	no	11874	15321	68.9	65.4	*+0.051*	30.4	28.1	*+0.076* *
Live with children:	yes	3635	4836	69.6	65.2	*+0.063*	27.3	24.8	*+0.092*
	no	9581	12408	70.7	67.5	*+0.045* **	34.2	31.9	*+0.067* **
Live with parents:	yes	2739	3524	67.6	64.2	*+0.050*	31.2	28.9	*+0.074*
	no	10477	13720	71.1	67.5	*+0.051*	32.6	30.2	*+0.074*
Community size:									
Farm/country		1716	2204	62.8	58.2	*+0.073*	22.4	20.5	*+0.085*
Small-medium city		6498	8508	70.0	66.3	*+0.053*	31.8	29.5	*+0.072*
Large city		2820	3693	72.8	69.7	*+0.043*	34.5	31.7	*+0.081*
Very large city		2045	2658	74.3	71.8	*+0.034* **	38.6	36.5	*+0.054*
Education level:									
High school		7110	9815	73.1	69.0	*+0.056*	35.4	32.8	*+0.073*
Associate degree		1575	2044	69.8	66.4	*+0.049*	31.0	28.1	*+0.094*
Bachelor degree		4082	4863	67.0	64.3	*+0.040*	28.4	26.6	*+0.063*
Master or Ph.D.		442	520	62.1	59.7	*+0.039* *	24.5	23.6	*+0.037*
Occupation:									
Semi-skilled		2157	2992	75.3	72.0	*+0.044*	37.4	35.1	*+0.062*
Clerical		2988	3863	67.4	63.9	*+0.052*	28.1	26.2	*+0.068*
Police/Fire		217	266	68.3	62.7	*+0.082*	26.8	22.4	*+0.164*
Military		391	486	66.7	60.2	*+0.097*	28.0	24.3	*+0.132*
Skilled		980	1455	77.5	73.9	*+0.046*	45.0	42.0	*+0.067*
Mngr/Prfessnl/Ph.D.		4668	5811	69.4	66.7	*+0.039* **	31.1	29.1	*+0.064*

Recanting rates are calculated as the percentage difference between lifetime users adjusted and lifetime users, divided by the lifetime users adjusted percentage. Lifetime users adjusted includes recanters; see text for full definition. Chi square analyses were performed distinguishing the recanters versus all users who did admit use at the fourth followup. Significant differences among subgroup categories are indicated by asterisks below the subgroups.

KEY: * = significant at the 0.05 level, ** = significant at the 0.01 level, R.R. = Recanting Rate.

TABLE 1. *Lifetime prevalence of five drugs at the fourth followup by subgroups, classes of 1977-1987 (continued).*

	LSD			Tranquillizers			Barbituates		
	Life-time, adjusted	Life-time	*R.R.*	Life-time, adjusted	Life-time	*R.R.*	Life-time, adjusted	Life-time	*R.R.*
Total:	16.6	15.0	*+0.093*	20.9	16.7	*+0.201*	13.4	10.3	*+0.237*
Sex:									
Males	21.6	19.9	*+0.079*	21.1	17.4	*+0.175*	15.5	12.2	*+0.213*
Females	12.5	11.1	*+0.112* *	20.7	16.1	*+0.222* **	11.8	8.7	*+0.263* *
Race:									
White	17.9	16.3	*+0.089*	22.2	17.9	*+0.194*	14.2	10.9	*+0.232*
Black	4.5	4.0	*+0.111*	9.3	7.5	*+0.194*	6.3	4.7	*+0.254*
Other	15.5	13.7	*+0.116*	19.0	14.5	*+0.237*	13.5	10.3	*+0.237*
Hshld comp; live w/:									
Spouse: yes	13.7	12.1	*+0.117*	19.8	15.2	*+0.232*	12.6	9.3	*+0.262*
no	18.9	17.4	*+0.079* **	21.6	17.9	*+0.171* **	14.1	11.0	*+0.220* *
Partner: yes	26.5	24.8	*+0.064*	28.5	23.8	*+0.165*	20.4	16.4	*+0.196*
no	15.5	14.0	*+0.097*	20.0	15.9	*+0.205*	12.6	9.6	*+0.238*
Children: yes	14.9	13.1	*+0.121*	20.8	15.6	*+0.250*	14.2	10.8	*+0.239*
no	17.2	15.8	*+0.081* **	20.8	17.1	*+0.178* **	13.1	10.1	*+0.229*
Parents: yes	14.5	13.0	*+0.103*	18.5	14.6	*+0.211*	12.1	9.5	*+0.215*
no	17.1	15.6	*+0.088*	21.4	17.3	*+0.192*	13.8	10.5	*+0.239*
Community size:									
Farm/country	14.0	12.4	*+0.114*	19.3	15.4	*+0.202*	13.4	10.4	*+0.224*
Small-medium city	16.5	15.0	*+0.091*	20.9	16.5	*+0.211*	14.1	10.7	*+0.241*
Large city	17.6	15.7	*+0.108*	20.9	16.8	*+0.196*	12.5	9.3	*+0.256*
Very large city	17.4	16.5	*+0.052* *	21.5	18.0	*+0.163*	12.4	10.1	*+0.185*
Education level:									
High school	20.1	18.1	*+0.100*	24.2	19.5	*+0.194*	17.2	13.3	*+0.227*
Associate degree	15.2	13.6	*+0.105*	19.7	14.9	*+0.244*	13.6	10.0	*+0.265*
Bachelor degree	11.9	11.0	*+0.076*	16.0	13.1	*+0.181*	7.8	5.9	*+0.244*
Master or Ph.D.	9.7	8.8	*+0.093*	16.4	12.7	*+0.226*	5.6	4.5	*+0.196*
Occupation:									
Semi-skilled	22.4	20.5	*+0.085*	25.8	21.3	*+0.174*	19.1	14.4	*+0.246*
Clerical	13.0	11.3	*+0.131*	20.0	15.5	*+0.225*	11.2	8.2	*+0.268*
Police/Fire	13.1	10.1	*+0.229*	16.6	12.8	*+0.229*	9.4	6.5	*+0.309*
Military	17.3	14.7	*+0.150*	15.9	12.4	*+0.220*	12.0	9.4	*+0.217*
Skilled	30.2	27.5	*+0.089*	28.1	23.1	*+0.178*	22.3	18.9	*+0.152*
Mngr/Prfessnl/Ph.D.	13.8	12.9	*+0.065* **	18.6	15.1	*+0.188*	10.6	8.1	*+0.236* *

differences for the other drugs are much smaller and not statistically significant.

The second hypothesis, concerning the effect of a spouse or children living in the household, also received empirical support. Those living with a spouse were more likely to recant use of all five drugs than those not living with a spouse; however, the differences are not very large. Similarly, differences were found in the recanting rates on four of the five drugs for those living with children versus those without children in the household. Those without children had lower recanting rates. Again, the differences were modest and of the same order of magnitude, suggesting that the effect of living with a spouse probably accounts for most of the apparent effect observed for living with children.

The third hypothesis was partially supported. Those who have the least schooling have the highest rate of recanting on marijuana use, but the relationship is very weak, with a recanting rate of 0.04 for those with only a high school education versus 0.06 for those who had attained a master's degree or more. For the other drugs, differences in recanting among categories are neither statistically significant nor ordinal. While there may be some tendency for the more educated to have lower recanting rates on the three illegal drugs (but not the psychotherapeutic drugs), the differences are hardly important. It should be noted in passing, however, that the absolute prevalence rates for all five drugs correlate negatively with the level of education attained, and that those differences are quite large in both absolute and proportional terms for LSD, tranquilizers, and barbiturates.

The fourth hypothesis, that those with high-status occupations would be more likely to recant because they have more to lose if their use were exposed, is not confirmed. Those in the top occupational category (defined as "managerial," "professional," or "requiring a Ph.D.") had among the lowest recanting rates.

Considerable support was found for the fifth hypothesis, that recanting would be highest among those working in the military or in police or firefighting (a combined category in the answer set) because the adverse consequences of possible exposure would be highest for them. The differences among the occupational groups were statistically significant for four of the five drugs examined (tranquilizers were the exception), and those in the military and police/firefighting professions had the highest rates of recanting on the three illegal drugs. This occupational

category also had the highest recanting rate for the psychotherapeutic drugs.

No hypothesis was offered about the effect of community size on recanting rates. Table 1 shows that for marijuana, recanting does decline very modestly with increasing community size. It also shows that those from the very large cities had the lowest rates of recanting across all five drugs; in general, however, the differences in recanting rates are not large across the community-size categories, nor are they very consistent.

DISCUSSION

For certain of the drugs—particularly marijuana and LSD, and to a lesser extent cocaine—the relatively low level of recanting of earlier reported use, even over a 14-year interval, is reassuring. It suggests that there is relatively little erosion in truth-telling with age, even as people are well along in their career paths and family formation. It also indicates that gathering data in the home setting by means of a mailed questionnaire is a reasonable approach for this age group. The data do suggest, however, that there may be some increase in concealment with age, and that age comparisons in cross-sectional studies of the general population will likely reflect this bias. (The estimates could conceivably be adjusted to correct for it, however.)

For the two psychotherapeutic drugs examined, the results were somewhat less reassuring. The recanting rates were higher in both absolute and proportional terms. Because the illicit use of these drugs is generally seen as no more deviant than the use of illegal drugs (as indicated by disapproval rates for various drugs; see Johnston et al. 1995), one would assume no greater motivation to conceal because of the threat of exposure. An alternative explanation, favored by the authors, is that the definitions of these substances are much less clear to respondents. There are, after all, nonprescription substances, such as over-the-counter sleep aids and diet aids, as well as mail order look-alike drugs, that are often given the same slang terms as the prescription drug about which the questions ask.[2] That fact raises the possibility that young respondents—particularly when still in high school—may be overinclusive in their earlier answers about drug classes such as tranquilizers and barbiturates. In subsequently recanting some of those overly inclusive answers, perhaps because of a better understanding of the intended distinctions, respondents may actually be providing more

accurate data in their later responses. (This interpretation is supported by the fact that for these two drugs, much of the recanting seems to involve use originally reported in high school.) In fact, the cohorts under study here might particularly show this effect because the distribution and use of look-alikes peaked around the late 1970s and early 1980s (Johnston et al. 1995). Further, the wording of the question was later revised to emphasize that only use of prescription substances should be included. To the extent that the rewording had this effect, subsequent cohorts may show less recanting. Still, the definitions for the psychotherapeutic drugs are difficult ones, and to the extent that the authors' hypothesis is true—namely, that young respondents tend to be overly inclusive in reporting their use—researchers surveying adolescents should be sensitive to the possibility that prevalence estimates may be high for these drugs. A further hypothesis is that the problem increases with even younger respondents, who may be less able to make some of the fine distinctions requested by the researchers. It is because of such concerns that the authors do not even report the prevalence rates for barbiturates and for narcotics other than heroin obtained from 8th and 10th grade respondents in the annual Monitoring the Future surveys.

Subgroup Differences

The finding that the rate of recanting is not much affected by the composition of the household setting in which the respondent receives the questionnaire is reassuring in many ways. In particular, it suggests that declines in use associated with getting married, or increases in use associated with leaving the parental home, are not methodological artifacts. Both such changes have been reported previously from the panel data from this study (Bachman et al. 1984).

It should also be reassuring to investigators in the field that recanting is not strongly associated with most of the other variables assessed here. That means that relationships between drug use and these variables, when examined in cross-sectional surveys of adults, are probably not biased by such a methodological artifact.

The two exceptions are, however, reason for some concern. If African-American respondents do tend to deny past use more than whites, then racial comparisons in cross-sectional surveys may need adjustments. The higher-than-average recanting rates for those in military and police/fire-fighting occupational settings also caution about how literally one takes survey data gathered on these populations. The data on these two groups

are only suggestive at this point, given the small subgroup sample sizes, but the findings are certainly worth further investigation. They are, however, consistent with the hypothesis that people in these settings have the most to lose if drug use were revealed and, therefore, may be more likely than average to conceal use.

CONCLUSIONS

Overall, recanting rates tend to be modest for the illegal drugs but less so for the illicit use of the psychotherapeutic drugs. This suggests that concealment effects are not strong, but that ambiguity in the definition of certain drug classes (clearly highest for the psychotherapeutic drugs) does lead to a modest amount of recanting. But, the "revised" answers may well be the more accurate ones, and the answers given at earlier ages for the psychotherapeutic classes of drugs may be inflated. In general, differences in recanting rates among subgroups are not large. The largest are the rates for African Americans who recant earlier reported marijuana and cocaine use (but not the use of the three other drugs), and those for young adults in the military or in police or firefighting occupations, who showed a tendency to recant more than other occupational groups. These findings raise some questions for those interpreting survey results based on these populations.

NOTES

1. The authors chose to base the criterion for recanting on the respondent's twice previously reporting use of the drug, rather than once, because they judged that it constituted an unambiguous statement by the respondent of having used. Simple reporting errors due to haste or misunderstanding should be reduced substantially using this method, and, in correcting lifetime prevalence estimates on the population, it seemed a reasonable procedure.

2. The study contains questions in a single questionnaire form administered to 12th graders on over-the-counter and look-alike stimulants, and reports significant levels of use of these drugs. Unfortunately, the study does not yet contain questions about sedating or tranquilizing agents sold over the counter or by mail order, although such products clearly do exist. When questions

about over-the-counter substances were last included in the study (1989), the 12-month prevalence rate for high school seniors was 16 percent for sleep aids and 5 percent for agents to "calm people down."

REFERENCES

Bachman, J.G.; O'Malley, P.M.; and Johnston, L.D. Drug use among young adults: The impacts of role status and social environments. *J Pers Soc Psychol* 47:629-645. 1984.

Fendrich, M., and Vaughn, C.M. Diminished lifetime substance use over time: An inquiry into differential underreporting. *Public Opin Q* 58:96-123, 1994.

Johnston, L.D., and O'Malley, P.M. Issues of validity and population coverage in student surveys of drug use. In: Rouse, B.A.; Kozel, N.J.; and Richards, L.G., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity.* National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985.

Johnston, L.D.; Driessen, F.M.H.M.; and Kokkevi, A. *Surveying Student Drug Misuse: A Six-Country Pilot Study.* Strasbourg, France: Council of Europe, 1991.

Johnston, L.D; O'Malley, P.M.; and Bachman, J.G. *National Survey Results on Drug Use from the Monitoring the Future Study, 1975-1994.* Vol. I: *Secondary School Students* and Vol. II: *College Students and Young Adults.* Rockville, MD: National Institute on Drug Abuse, 1995.

Mensch, B.S., and Kandel, D.B. Underreporting of substance use in a national youth cohort: Individual and interviewer effects. *Public Opin Q* 52:100-124, 1988.

O'Malley, P.M.; Bachman, J.G.; and Johnston, L.D. Period, age, and cohort effects on substance use among young Americans: A decade of change, 1976-1986. *Am J Public Health* 78:1315-1321, 1988.

O'Malley, P.M.; Bachman, J.G.; and Johnston, L.D. Reliability and consistency of self-reports of drug use. *Int J Addict* 18:805-824, 1983.

Substance Abuse and Mental Health Services Administration. *National Household Survey on Drug Abuse: Main Findings 1992.* DHHS Pub. No. (SMA)94-3012. Rockville, MD: Substance Abuse and Mental Health Services Administration, 1995.

Wallace, J.M., Jr., and Bachman, J.G. Validity of self-reports in student based studies on minority populations: Issues and concerns. In: De La Rosa, M., and Recio-Andrados, J.L., eds. *Drug Abuse among Minority Youth: Advances in Research and Methodology*. National Institute on Drug Abuse Research Monograph 130. NIH Pub. No. 93-3479. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1993.

ACKNOWLEDGMENTS

The authors wish to acknowledge the analytic contributions of Dawn Bare and the editorial contributions of Joyce Buchanan to the development of this chapter. The work presented here was funded under grant number R01 DA-01411 from the National Institute on Drug Abuse.

AUTHORS

Lloyd D. Johnston, Ph.D.

Patrick M. O'Malley, Ph.D.

Institute for Social Research
University of Michigan
Ann Arbor, MI 48106-1248

The Reliability and Consistency of Drug Reporting in Ethnographic Samples

Michael Fendrich, Mary Ellen Mackesy-Amiti, Joseph S. Wislar, and Paul Goldstein

ABSTRACT

Findings are addressed concerning the reliability of reporting on drug dealing and drug use. Reports provided in retrospective life history interviews are compared with reports gathered and summarized from eight prospective weekly interviews. Most subjects reporting involvement in drug dealing during the weekly interviews, also reported involvement in this behavior during the life history report. There was a tendency for subjects to deny current involvement in drug dealing during the life history reports, even though they reported involvement in drug dealing in the weekly interviews. Binary indicators derived from life history interviews about current drug use were consistent with reports provided prospectively. Subjects reported considerably higher use quantities and frequencies for substances in the life history reports than they did in the weekly interview reports. These results are examined in the context of other recent work examining the reliability of retrospective substance involvement reports. Implications for ethnographic research on drug use are discussed.

INTRODUCTION

A growing body of ethnographic research describes drug use practices in untreated samples drawn from subcultures where many forms of drug use are normative (Adler 1993; Goldstein et al. 1990; Johnson et al. 1985; Waldorf et al. 1991; Weibel 1988). In ethnographic studies, groups of users are followed for weeks, months, or years, in order to evaluate drug use patterns, correlates, and consequences. More recently, ethnographic research approaches have been incorporated into the planning and evaluation of human immunodeficiency virus (HIV) prevention programs with intravenous (IV) drug users (Stephens et al. 1991; Weibel 1988). Conclusions with respect to program effectiveness

as well as about the generalizability of previous findings from ethnographic studies of drug abusers require a clear understanding of the validity of the interview measures and procedures employed in these studies.

Ethnographic research places an emphasis on observational accounts of subject behavior (Goldstein et al. 1991). Thus, discussions of validity in ethnographic research on drug use have focused on observational verification of subject responses. For example, Biernecki and Waldorf (1981, p. 151) report that in their study of former opiate addicts, researchers would verify reports of nonuse by asking to "examine a respondent's arms in order to check for relatively fresh signs of needle injections." Johnson and colleagues (1985) discuss observations drawn from a visit to a heroin addict's apartment as validating lifestyle information provided in earlier interviews. Biernecki and Waldorf (1981), Johnson and colleagues (1985), and Goldstein and colleagues (1987, 1988) also report that information provided by a subject was sometimes validated by information provided by other informants enlisted in the research project (triangulation). Adler (1993) discusses the use of cross-checking to verify accounts provided by participants in a study of drug dealers. This procedure included corroboration of accounts with other sources and investigation of available hard facts (such as arrest records, visible evidence, and newspaper reports), as well as direct, critical observation of the drug scene around them.

These earlier approaches failed to address a more basic issue in the assessment of validity. For measures to be valid, they need to be reliable (Lord and Novick 1968). To the extent that informants provide consistent responses when they are asked to discuss the same behavior, their responses may be considered reliable. But the question is: To what extent are ethnographic accounts provided by individuals reliable? Johnson and colleagues (1985) and Goldstein and colleagues (1987) note that they examined internal consistencies and the correspondence between replicate measures of the same behaviors within their respective research summaries. Nevertheless, neither of these studies provided a formal statistical assessment of reliability. Fendrich and colleagues (1992) reanalyzed the data discussed in Goldstein and colleagues (1987, 1988) to statistically assess the consistency of drug use reports provided in prospective weekly interviews. They found that individuals were more consistent in their reports of drug use frequency (days of consumption) than they were in their reports of drug use amount (cost of drugs consumed) over an 8-week period. A particularly striking finding was a general tendency for respondents to report diminishing levels of drug use

(irrespective of measure) over the 8-week reporting period. The authors suggested three possible interpretations of this finding. It may have reflected real changes in behavior. It also may reflect the phenomenon of retest artifact (Jorm et al. 1989). Psychiatric research suggests that levels of symptomatology (and substance use) diminish when subjects are reinterviewed (Bromet et al. 1986; Rubio-Stipec et al. 1992). Finally, since subjects were aware that drug use was an important qualifying characteristic for study entry (and subsequent receipt of subject payment), higher initial reporting levels could have reflected perceived demand characteristics of the study; subjects may have overestimated their levels of drug use initially to appear as better qualified subjects.

In this chapter, the authors follow up on previous analyses of reliability in ethnographic research by examining the reliability of retrospectively provided life history information about drug use and drug dealing. Retrospective summary information about typical patterns of substance use and involvement in drug dealing is compared to prospectively gathered weekly reports about similar behavior. The aim is to address the following questions: How consistent is retrospectively provided information with information provided prospectively? Does consistency with respect to reporting on drug dealing differ from consistency with respect to reporting on drug use? Does consistency vary by type of substance or by type of substance use measure (i.e., frequency versus volume)? Are retrospective reports an overestimate or underestimate of behavior reported prospectively? Do trends in reporting consistency vary by respondent characteristics?

METHODS

Sample

Two different ethnographic studies were undertaken on the Lower East Side of New York City between 1984 and 1987. Interviews for a study examining the drugs/violence nexus among adult male drug users and distributors were carried out between November 1984 and April 1986 (Project DRIVE (Drug Related Involvement in Violent Episodes)) (Goldstein et al. 1987). Interviews for a similar study of female drug users and distributors were carried out between April 1986 and May 1987 (Project FEMDRIVE (Female Drug Related Involvement in Violent Episodes)) (Goldstein et al. 1988). Respondents from both studies were

adults over the age of 18 who were recruited from field contacts, through snowball sampling techniques, and from a local methadone maintenance treatment program. Interviewing took place in an ethnographic field station established solely for the purposes of these projects. Descriptive characteristics of this sample have been discussed in detail elsewhere (Fendrich et al. 1992). To briefly summarize, both samples were racially and ethnically heterogeneous, with African Americans representing the modal racial category. The majority of the men and women were high school graduates, and a substantial portion had attended college. The modal living situation for both men and women was in shelters for the homeless.

Study Design

Respondents in both studies were interviewed using a similar set of semistructured interview instruments. Upon recruitment to the study, all respondents were given a life history interview (DRIVE respondents completed this interview in an average of 2.5 sessions; FEMDRIVE respondents completed this interview in an average of 5 sessions). This interview focused on a wide range of issues, including drug use history, participation in treatment programs, involvement in drug sales and distribution, criminal history, and history of involvement in violence. After the final life history interview session, respondents were interviewed in detail about their activity over the previous 7 days. Detailed information was collected about drug use and drug dealing, criminal activity, violent perpetrations and victimizations, sources of income, and types of expenditures on each of the 7 days. Data covering 7 discrete days were collected for each respondent. Respondents were asked to return to the field station to complete additional indepth interviews about daily activity over the course of 7 weeks. The eight weekly interviews were not necessarily consecutive. Interviews about daily behavior pertaining to 8 distinct weeks were obtained for 152 males for the initial study and 133 females for the second study. All subjects included in these analyses completed all phases of the study.

Interview Format

Life History Interview. The life history survey was a semistructured, open-ended interview in which respondents were asked to describe patterns of substance use, exposure to violence, and criminal involvement; they were asked to recall whether they had ever tried a particular substance. Respondents who disclosed substance involvement were

asked about specific periods of involvement; for each period of involvement, participants were asked to specify their frequency and typical cost of substance use. Participants were also asked about their involvement in a number of specific criminal behaviors. Respondents who disclosed criminal involvement were asked about specific periods of involvement; for each period of involvement, they were asked to specify how often they were involved as well as whether any violence or injuries resulted from their involvement.

Weekly Interview. The weekly interviews were constructed in a more structured, diary format. For each weekly interview, the respondent was asked to retrospectively report on the estimated dollar amount of substances purchased and on the estimated dollar amount consumed. The substances covered in the weekly interviews paralleled those asked about in the life history report. Additionally, the respondent was asked about a range of economic and criminal activities engaged in on each day of the previous week. Specific daily criminal activities along with dollar amounts they generated were recorded by interviewers.

Measures

Life History Interview Measures. For these analyses, three life history measures of drug-dealing involvement were constructed—one general measure of lifetime drug dealing and two indices reflecting recent drug dealing. The latter two measures are based on reports provided by informants of the "age of last involvement" with this activity; included is an indicator of involvement in the past 2 years and involvement in the past year. Two measures of substance use were derived from the life history interviews for use in comparative analyses. One described the most recent typical use frequency for each substance; the other the most recent typical cost per use day for each substance.[1] Current use status was also coded in the life history interview. Based on an examination of the data,[2] use frequency categories were divided into four mutually exclusive groups (coded on a scale from 1 to 4, with 4 indicating a higher frequency): Infrequent users were those who characterized their use as monthly or less; moderate users were those admitting to use on weekends or on no more than 2 days during any particular week; regular users used at least 3 days per week but no more than 5 days per week; daily users used nearly every day (6 or 7 days per week). Most recent cost per day was derived from an actual dollar amount estimate of typical cost-per-day of use provided for each substance.

Weekly Interview Measures. A measure of weekly drug dealing was constructed by evaluating whether any drug dealing was reported over the course of the 8-week interview period. To obtain an estimate of drug use cost comparable to that used in the life history report, the authors constructed a use volume index. The total dollar amount consumed in the course of 8 weeks was divided by the total number of days in which use was indicated. Those who consumed no substances over the course of 8 weeks were assigned a "0" on this measure. For each substance, use volume on the weekly interviews was compared to the typical cost-per-day estimates provided in the life history reports. Frequency pattern variables for the weekly interviews were constructed as measures of the total number of days used per week of use. First, a numerator was constructed based on the total number of use days over the course of 8 weeks. Next, a denominator was constructed based on the total number of weeks during which use was recorded. Thus, for each substance, each individual had a ratio of days used per use week. All individuals with no use were coded as "0" on this ratio. As a final step, this measure was divided into four use categories (ranging from light use to daily use) that were roughly equivalent to the four categories coded for in the life history measure.[3]

The measures used for this study are described and summarized in table 1, which indicates the source of each measure (life history report or weekly interview) and any transformations made on each measure for the purposes of data analysis. This table also indicates the variables that were compared in quantitative analyses.

RESULTS

Drug-Dealing Activity

The first focus in the analysis compares drug-dealing activity reported in the life history section of the interview with that reported in the weekly interviews. Comparisons are described for the three life history indices of drug dealing in table 2.[4] Lifetime prevalence of drug-dealing activity exceeds the prevalence of this behavior during the weekly interviews. In DRIVE, 81 percent of the respondents disclosed in the life history interview that they had been involved in drug dealing at least once in their lifetimes; 66 percent of the respondents disclosed involvement in drug dealing during the weekly interviews. Similarly, 67 percent of the FEMDRIVE respondents disclosed involvement in drug dealing during

TABLE 1. *Measures of drug use and drug dealing.*

Measure	Definition	Source	Comparison measure
Any drug dealing	Subject has dealt drugs	Life history interview	Weekly drug dealing
Recent drug dealing: last 2 years	Age at last occurrence of drug dealing was no more than 2 years less than current age	Life history interview	Weekly drug dealing
Recent drug dealing: last year	Age at last occurrence of drug dealing was no more than 1 year less than current age	Life history interview	Weekly drug dealing
Weekly drug dealing	Any drug dealing reported in any of the 8 weeks	Weekly interview	Any drug dealing/recent drug dealing
Current use	Subject using substance at time of interview (explicitly stated, or last reported use in the current year), or quit using less than 1 month ago	Life history interview	Weekly use
Weekly use	Subject reported use of substance in any of the 8 weeks	Weekly interview	Current use
Cost per day	Typical cost per day of drug use for the most recent period of use	Life history interview	Use volume
Frequency pattern	Typical frequency of use for the most recent period of use, coded into four categories: infrequent (once a month or less), moderate (2 times/ month to 2 days/week), regular (3-4 days/week), and daily (5-7 days/week)	Life history interview	Average days per week
Use volume	Average dollar amount of drug used per day of drug use (total cost of drug used over 8 weeks/number of days used over 8 weeks)	Weekly interview	Cost per day
Average days per week	Average number of days used per week used (total number of days used over 8 weeks/number of weeks used); recoded into four categories: infrequent (< 0.5), moderate (0.5-2.5), regular (2.5-5.5), and daily (5.5 or more)	Weekly interview	Frequency pattern

TABLE 2. *Drug dealing: Life history and weekly interviews.*

Life history measure	Life history prevalence %	Weekly prevalence %	Sensitivity of LH report[1] %	(n1/n2)	Kappa	Conditional Kappa
DRIVE						
Any drug dealing (N = 146)	80.8	66.4	83.5	(81/97)	0.09	0.15
Dealing in past 2 years (N = 123)	36.6	64.2	39.2	(31/79)	0.06	0.04
Dealing in past year (N = 123)	28.5	64.2	32.9	(26/79)	0.1	0.06
FEMDRIVE						
Any drug dealing (N = 132)	67.4	43.2	78.9	(45/57)	0.19	0.35
Dealing in past 2 years (N = 128)	22.7	43.8	30.4	(17/56)	0.14	0.1
Dealing in past year (N = 128)	13.3	43.8	16.1	(9/56)	0.05	0.03

KEY: 1 = Sensitivity is defined as the percentage of weekly drug dealers (n2) who also identify themselves as drug dealers (either lifetime, in the past 2 years, or in the past year) in the life history report (n1).

the life history interview; 43 percent disclosed involvement with this activity during the weekly interviews (see table 2).

Since lifetime behavior encompasses a longer frame of reference than current behavior, one should expect current behavior to differ from past behavior. Nevertheless, three additional statistics suggest a certain degree of unexpected inconsistency with respect to lifetime and weekly interview reports. The sensitivity of life history reports was considerably less than unity for both DRIVE and FEMDRIVE. In both samples, close to 20 percent of those disclosing drug-dealing activity during the weekly interviews reported that they never were involved in drug dealing during the life history interviews. This may suggest underreporting of lifetime

drug dealing in the life history reports. This possible underreporting is also supported by relatively low conditional Kappa statistics. The Kappa statistic should approach at least a value of 0.40 to be considered "fair." Conditional Kappa statistics (Bishop et al. 1975) measure agreement with respect to drug-dealing behavior, conditional on that behavior's occurring during the weekly interviews.[5] Note that when lifetime drug dealing is the comparison measure, conditional Kappa statistics for neither sample reach a level considered to be acceptable. Although there is general inconsistency with respect to the reporting of drug-dealing behavior (Kappas of 0.09 and 0.19 were observed for lifetime drug-dealing comparisons in DRIVE and FEMDRIVE), the use of conditional statistics yield substantial improvements in the evaluation of chance-corrected agreement only for FEMDRIVE (the coefficient increases from 0.19 to 0.35 in FEMDRIVE and from 0.09 to 0.15 in DRIVE). Both of the conditional agreement statistics suggest poor levels of agreement conditional on drug-dealing reports in the weekly interviews.

As a second step, the agreement between recent drug-dealing activity in the life history reports and drug-dealing activity in the weekly interviews was examined. When reports provided in the weekly interviews were used as criteria, sensitivity rates sharply declined from their previous levels in both DRIVE and FEMDRIVE. In DRIVE, only 39 percent of those reporting involvement in drug dealing during the weekly interviews also reported life history involvement in this behavior during the past 2 years; 33 percent of those reporting involvement during the weekly interviews also reported life history involvement during the past year. In FEMDRIVE, the shift to the more narrowly defined dealing recency measure results in a dramatic decrement of sensitivity: Only 30 percent of those reporting involvement in drug dealing during the weekly interviews also reported life history involvement in the past 2 years, and only 16 percent of those reporting involvement during the weekly interviews also reported life history involvement in the past year. These findings are paralleled by relatively low coefficients for Kappa and conditional Kappa statistics for both measures of recent involvement in both samples.

Current Drug Use Reporting

Table 3 describes the overall rates of substance involvement across interviews and presents the agreement between binary measures of substance involvement for all subjects who had complete life history responses

TABLE 3. *Current substance use involvement in life history and weekly interviews: Prevalence and agreement statistics.*

Substance	Reported Any Involvement								
	Life history prevalence[1]		Weekly interview prevalence		Sensitivity [2]			Agreement coefficient Kappa	Conditional Kappa
	N	%	N	%	N	%	(n1/n2)		
DRIVE									
Heroin	151	55.0	83	50.3	76	91	(69/76)	0.72	0.80
Cocaine	150	80.7	121	81.3	122	92	(112/122)	0.59	0.58
Marijuana	148	79.1	117	77.0	114	90	(103/114)	0.51	0.54
Alcohol	146	74.7	109	83.6	122	84	(103/122)	0.49	0.38
FEMDRIVE									
Heroin	133	47.4	63	38.3	51	84	(43/51)	0.57	0.70
Cocaine	132	79.5	105	78.0	103	88	(91/103)	0.41	0.43
Marijuana	128	60.9	78	60.2	77	87	(66/77)	0.66	0.66
Alcohol	121	71.1	86	71.9	87	83	(72/87)	0.41	0.41

KEY: 1 = Respondents who were classified as current users based on the life history interview. 2 = Sensitivity is defined as the percentage of weekly drug users (n2) who also identify themselves as drug users in the life history report (n1).

available on the questionnaire. Use in the life history reports is limited to those who were counted as current[6] users at the time of the retrospective interview. In general, rates of reported use were consistently close across interview phases for most substances. In DRIVE, only alcohol use reports show a statistically significant shift across interviews; a significant number of respondents shifted from noncurrent use in the life history to current use in the weekly interviews (McNemar $\chi^2 = 6.76$; $p < 0.01$). In FEMDRIVE, only heroin use reports show a statistically significant change across interviews; a significant number of respondents shift from current use in the life history reports to nonuse in the weekly interviews (McNemar $\chi^2 = 5.14$; $p < 0.05$).

Kappa coefficients evaluating the overall level of agreement on the binary measure of use at each phase of interviewing are displayed in the last column of table 3. While agreement between interviews with respect to classification of current use was far from perfect, all coefficients fell within a range considered to be "fair to good." With one exception (marijuana use reports), levels of agreement were generally higher between interviews for DRIVE men than for FEMDRIVE women. In DRIVE, the largest coefficient measured agreement on heroin use; a Kappa of 0.72 suggested a relatively high level of agreement. The agreement coefficient for heroin was also relatively high in FEMDRIVE; a Kappa of 0.57 was second only to the coefficient of 0.66 generated for FEMDRIVE reports of current marijuana use. The coefficients for current cocaine use (0.41) and current alcohol use (0.41) in FEMDRIVE barely exceeded a level indicative of poor agreement.

The findings in table 3 stand in contrast to findings about reports of drug dealing suggested in table 2. New reports of previously unreported current drug use behavior during the weekly interviews were relatively infrequent. Assessment of conditional levels of agreement and sensitivity statistics in both samples underscores the relative consistency of use reports across interview phases. When respondents reported use in the weekly interviews, they almost always were classified as current users in the life history interviews. Sensitivity statistics all exceeded 80 percent in FEMDRIVE; three of four sensitivity statistics were at least 90 percent in DRIVE. Conditional Kappa values for heroin use classification status jumped to 0.80 in DRIVE and to 0.70 in FEMDRIVE.

It should also be noted that alcohol was the substance that was most underreported[7] during the life history reports. Nineteen DRIVE and 15 FEMDRIVE subjects who were not classified as current alcohol users in the life history reports disclosed alcohol use during the weekly interviews. This underreporting stands in considerable contrast with the relatively low levels of life history underreporting for heroin in both samples (only seven subjects in DRIVE and eight subjects in FEMDRIVE underreported heroin use in the life history reports). Inspection of case files suggested that subjects who were involved in a variety of harder substances may have minimized their involvement with alcohol during the life history interviews.

The data suggest that, at least with respect to heroin and cocaine use, the phenomenon of overreporting was more common than the phenomenon of underreporting. Fourteen DRIVE subjects and 20 FEMDRIVE

subjects who were classified as current heroin users from the life history data reported no use during weekly interviews. Nine DRIVE subjects and 14 FEMDRIVE subjects classified as current cocaine users in the life history reports reported no use during the weekly interviews. Followup analyses suggested that current life history users who failed to report any heroin use during the weekly reports ("stoppers") were significantly more likely to be enrolled in methadone maintenance programs for the entire 8-week prospective interview period (for DRIVE, χ^2, 2 d.f. = 16.88, $p < 0.001$; for FEMDRIVE, χ^2, 2 d.f. = 14.53, $p < 0.001$). Another variable differentiating stoppers from nonstoppers was most recent use quantity (cost per day) reported in the life history interview. Two differences in FEMDRIVE and one difference in DRIVE were nonsignificant but reflected an important trend in the data. For cocaine use in both studies and for heroin use in FEMDRIVE, subjects who stopped reporting use of the substance during the weekly interviews reported a lower most recent cost of use in the life history than did those who reported continued use. In FEMDRIVE, heroin and cocaine use was $23 and $34 less, respectively, for stoppers.

Levels of Drug Use

Table 4a describes summary statistics comparing levels of drug use over each phase of interviewing for DRIVE men; the analogous table for FEMDRIVE women is 4b.[8] Immediately apparent are the reduced sample sizes in the comparisons. For example, even though there were 83 current DRIVE heroin users in the life history report, volume comparisons are based on only 57 users. A great deal of information was missing from the life history data about use quantities. In a review of case files, the authors found numerous instances where exact dollar amounts pertaining to a subject's recent experience were not actually recorded. Some subjects were supplied with drugs for free so that their typical cost for substances was listed as $0. These subjects were excluded from comparisons. Problems with missing data and noncomparable cost values underscore the difficulties of using ethnographic data for examining issues of reliability in a systematic way. Most of the information contained in the more structured weekly interview format was complete.[9]

When the mean values across interviews are compared, retrospective reports appear to considerably overestimate weekly volume measures (cost per use day) for heroin and cocaine. Indeed, the estimated typical heroin cost per day in the life history report is more than twice the value reported in the weekly interviews. Similarly, the estimated value for

TABLE 4a. *Comparisons of life history reports of substance use to weekly reports by substance—DRIVE.*

	Measurement occasion							
	Life history interview		Weekly interview				Zero-order	Intraclass
Measure	Mean	(SD)	Mean	(SD)	N	Paired t	correlation	correlation
Heroin volume	66.0	(65.2)	31.3	(24.9)	57	4.25**	0.33	0.09
Cocaine volume	65.6	(70.6)	34.3	(39.4)	75	4.19**	0.42	0.27
Marijuana volume	04.5	(2.5)	3.2	(2.0)	58	3.73**	0.33	0.23
Alcohol volume	06.3	(6.2)	5.6	(4.7)	37	0.50	0.00	0.01
Heroin frequency	03.1	(1.2)	2.7	(1.1)	70	2.69*	0.41	0.37
Cocaine frequency	02.9	(1.2)	2.6	(0.8)	103	3.02*	0.47	0.41
Marijuana frequency	03.0	(1.2)	3.0	(1.0)	101	0.20	0.61	0.60
Alcohol frequency	02.8	(1.3)	2.9	(0.9)	84	-0.98	0.52	0.49

KEY: * = $p < 0.01$; ** = $p < 0.001$.

cocaine cost per day reported in the life history approaches twice the value reported in the weekly interviews. Estimates provided for marijuana and alcohol dollar costs correspond more closely between interviews. Nevertheless, statistical comparisons reflect significant decreases in mean levels for all substances except alcohol (table 4a). Including only decreases of greater than $5 per use day in the calculations, more than two-thirds of all heroin users and nearly two-thirds of all cocaine users show a decrease in volume between life history and weekly interviews (see table 4a).

TABLE 4b. *Comparisons of life history reports of substance use to weekly reports by substance—FEMDRIVE.*

	Measurement occasion							
	Life history interview		Weekly interview					
Measure	Mean	(SD)	Mean	(SD)	N	Paired t	Zero-order correlation	Intraclass correlation
Heroin volume	90.2	(75.2)	30.1	(28.9)	45	5.06**	0.04	-0.19
Cocaine volume	76.5	(97.3)	30.9	(27.1)	58	3.73**	0.29	0.05
Marijuana volume	6.7	(6.9)	3.1	(2.5)	38	3.05*	0.03	-0.10
Alcohol volume	8.2	(9.9)	3.2	(2.2)	29	2.98*	0.51	0.10
Heroin frequency	3.1	(1.3)	2.4	(1.1)	57	3.61*	0.41	0.32
Cocaine frequency	2.8	(1.1)	2.4	(0.8)	89	4.25**	0.46	0.19
Marijuana frequency	2.5	(1.2)	2.2	(0.7)	70	2.85*	0.49	0.40
Alcohol frequency	2.4	(1.2)	2.4	(0.9)	79	0.50	0.42	0.41

KEY: * = $p < 0.01$; ** = $p < 0.001$.

DRIVE frequency patterns gauged across the two types of interview segments show much closer correspondence than volume measures. Slight reductions in frequency patterns derived from the weekly interview compared with the life history reports were observed only for heroin and cocaine. There is no substance for which a majority drop more than one scale point in the weekly interviews compared to the life history report. Table 5 shows that although most subjects don't show increases in their use frequency reports, subjects were just as likely to report the same levels of heroin and cocaine use frequency as they were to report decreased use frequency for these substances.

TABLE 5. *Change in drug use frequency,[a] current users.*

		Frequency pattern change							
		Decrease		Same		Increase		Total	Weekly
Sample	Drug	%	N	%	N	%	N	N	nonusers
DRIVE	Heroin	41	29	40	28	19	13	70	12
	Cocaine	41	42	40	41	19	20	103	9
	Marijuana	27	27	49	49	25	25	101	9
	Alcohol	25	21	43	36	32	27	84	3
FEMDRIVE	Heroin	46	26	35	20	19	11	57	17
	Cocaine	46	41	37	33	17	15	89	12
	Marijuana	44	31	37	26	19	13	70	11
	Alcohol	30	24	41	32	29	23	79	11

KEY: a = Frequency change is calculated by subtracting weekly pattern based on average days per week used from life history most recent pattern.

Table 4b highlights volume and frequency comparisons over the two interview phases for FEMDRIVE women. Volume reductions for all substances are more pronounced for FEMDRIVE women than they were for DRIVE men. Mean heroin volume generated from the weekly interview reports is one-third of the mean volume generated from the life history report. Mean cocaine volume generated from the weekly reports is less than one-half of the volume generated from the life history reports.

All volume comparisons reflect significant decreases. Over three-quarters of all FEMDRIVE current heroin users reported reduced heroin volume use (see table 6). Slightly less than three-quarters of all current cocaine users reported reduced volume use in the weekly reports. Another striking contrast is the relatively low magnitude of the correlation coefficients generated for cocaine and heroin use volume reports (table 4b). In contrast to DRIVE reports, there seems to be little correspondence between use volume reports for heroin and cocaine use across interview phases for FEMDRIVE; in other words, those who appear as high volume users in the life history reports are not likely to appear as high volume users in the weekly reports. The only substance showing consistency with

TABLE 6. *Change in drug use volume,[a] current users.*

Sample	Drug	Mean	Median	N	% Decrease	% Same	% Increase	Weekly nonusers
DRIVE		-34.7	-18.6	57	67	12	21	9
	Cocaine	-31.3	-15.8	75	65	13	21	4
	Marijuana	-1.3	-1	58	55	26	19	3
	Alcohol	-0.6	0	37	35	22	43	1
FEMDRIVE		-60.1	-46.7	45	78	4	18	12
	Cocaine	-45.6	-23.3	58	74	7	19	5
	Marijuana	-3.6	-2	38	53	21	26	5
	Alcohol	-4.9	-1.9	29	66	28	7	3

KEY: a = Volume change is calculated by subtracting weekly volume from life history most recent cost per day. Thus, a positive score indicates an increase, and a negative score indicates a decrease from the life history interview to the weekly reports. NOTE: For heroin and cocaine, volume change of less than $5 was considered as no change; for marijuana and alcohol, change of less than $1 was considered as no change.

respect to volume level ranking was alcohol. Women who were large volume consumers of alcohol in the life history were also large volume consumers in the weekly reports. FEMDRIVE frequency comparisons across interview phases show levels of stability that are similar to those indicated in DRIVE (see table 4b). Frequencies for all substances reflect significantly diminished levels of use in the weekly reports compared with life history reports. Nevertheless, as in DRIVE, the absolute magnitude of the frequency differences is small. Additionally, both zero-order correlations and intraclass correlation coefficients suggest that compared with volume indicators, frequency magnitude estimates are relatively consistent across interview phases. As in DRIVE, frequency decreases were only slightly more common than level frequency reports; frequency increases were relatively uncommon for all substances (see table 5).

Correlates of Changes in Volume and Frequency Measures

Additional exploratory analyses attempted to identify correlates of diminished use frequency and use volume reports for heroin and cocaine, the two substances showing the largest declines in mean value across measures and samples. Building on earlier work in this area (Fendrich and Vaughn 1994), the authors looked at two sets of variables including a set of four demographic indicators (subject age at the life history interview, race/ethnicity, homeless shelter residence versus nonshelter, and education level) and two drug involvement indicators (life history drug use frequency and weekly involvement in drug dealing).

Frequency Change Comparisons. Frequency change variables were converted to dichotomous change indicators (reduction versus nonreduction); bivariate cross-tabulations examining the seven variables were examined, setting alpha to 0.01 in order to adjust for multiple comparisons (data not shown here). Six comparisons yielded significant results; four of the significant comparisons involved a single variable, use frequency. For both samples and for both drugs, those who were classified as daily users in the life history reports were significantly more likely than other users to report decreased use frequencies in the weekly interviews. Race/ethnic differences suggested that Hispanic women in FEMDRIVE had significantly elevated rates of heroin frequency reduction. FEMDRIVE women who were residents of homeless shelters were significantly less likely than others to report diminished heroin use frequency.

Volume Change Comparisons. Analysis of covariance (results not shown here) was used to assess the impact of the same seven indicators discussed above on change in volume level reports (continuous measures of change were the dependent variables; baseline volume reports were covariates in all models). Again, setting alpha to 0.01, only two variables reached significance in any of the analyses: age and race. In FEMDRIVE, older respondents showed significantly greater cocaine volume decrease compared with younger respondents. Race/ethnicity effects varied in FEMDRIVE: For heroin volume comparisons, African-American and Hispanic respondents showed a greater decrease in volume than white respondents. For cocaine volume comparisons, African-American respondents showed less of a volume decrease than white respondents. In DRIVE, a nonsignificant trend suggested that respondents 25 years old or younger showed lower heroin and cocaine volume decreases compared with older respondents.

DISCUSSION

Summary of the Findings

Drug Dealing. Most of the subjects who reported involvement in drug dealing during the weekly interviews also disclosed lifetime involvement in that behavior in the retrospective interview. Discrepancies with respect to dealing concerned the timing of dealing involvement. In general, respondents who reported involvement in drug dealing during the weekly interviews did not disclose recent involvement in this behavior in the life history report; if they admitted to drug dealing in the life history reports, they described this behavior as having last occurred in the more distant past (i.e., more than 2 years before the life history interview).

Drug Use. Current use reports of heroin and cocaine were relatively consistent across interview phases. Inconsistencies in drug use reports were mainly in the area of reported use quantities and frequencies. Subjects tended to report higher use volume and frequency for substances in the life history reports than they did in the weekly interview reports. Reductions in the weekly report compared with the life history report were particularly striking for heroin and cocaine. About two-thirds of all male weekly heroin and cocaine users and about three-quarters of all female heroin and cocaine users reported reduced volume use in the weekly interviews. In general, reports of use frequency were considerably more consistent across interview phases

than were reports of use volume. In both DRIVE and FEMDRIVE, cocaine was the substance that showed the highest rate of decrease in reported use frequency over the course of the two phases of interviewing. In both DRIVE and FEMDRIVE, heroin showed the highest rate of decrease in reported use volume over the course of the two phases of interviewing.

Correlates of Decrease in Reported Drug Use. Reduction in reported drug use, especially reduction in volume of heroin and cocaine, was very prevalent; consequently, no single variable consistently differentiated reducers from nonreducers. The phenomenon of use cessation was examined among life history current heroin users (weekly stoppers were compared with life history current users). Life history current heroin users enrolled in methadone treatment throughout the course of the weekly interviews were significantly more likely to cease using heroin over the course of the weekly interviews. For both DRIVE and FEMDRIVE heroin and cocaine users, those who were classified as daily users in the life history reports were significantly more likely than other users to report decreased use frequencies in the weekly reports.

Limitations

The analyses presented in this chapter are based on samples of drug users and distributors residing in New York City during the mid- to late 1980s. The findings may not be generalizable beyond this particular setting. Possible limitations with respect to generalizability beyond the specific time period are particularly important. The data were collected during a period in which cocaine use, crack in particular, was beginning to rise. In previous comparative analyses (Fendrich and Vaughn 1994), the authors have noted that historical shifts in attitudes about drug use may influence the willingness to disclose drug involvement. Magura and Kang (this volume) present findings from more recent data that stand in contrast to the current analyses; their results indicate that respondents were more willing to discuss drug dealing than drug use.

Implications

The authors treated the differences discussed above as if they reflected inconsistencies. However, the possibility exists that the differences reflect real changes in behavior. An examination of a range of behaviors characterizing the samples investigated here reinforces a sense of their instability: Many of the subjects included in the two samples were intermittently

involved in treatment during the course of the study. Subjects were in and out of jail during the course of the study; many resided in homeless shelters. These indicators of lifestyle instability may be accompanied by instability of actual drug use as well, resulting in unstable estimates of typical drug use.

Another variable that could affect shifts in behavior is the time between interviews. If shifts in behavior are occurring, one might expect more between interview phases if those interview phases are far apart; these shifts would lead to greater reporting discrepancies on use indices. The mean time between the first life history interview and the last weekly interview was 71 days for DRIVE and 97 days for FEMDRIVE; the total study period could have lasted as long as 317 days for DRIVE and 430 days for FEMDRIVE. The associations between interview timespan and changes in reported use levels for DRIVE and FEMDRIVE cocaine and heroin volume and frequency measures were investigated. The data suggest that for DRIVE subjects, larger decreases in levels of cocaine use volume may be associated with a longer study period ($r = 0.23$; $p < 0.05$).

Preliminary inspection of the case files suggests that some of the discrepancy in drug-dealing reports may be the result of discrepant definitions of drug dealing between interviewers and subjects. This seems especially applicable to low-level or sporadic dealers who reported they occasionally sold small quantities of substances during the weekly interviews. For example, some of the women on methadone maintenance who did not report involvement in drug dealing during the life history interviews reported in weekly interviews that they sold their methadone from the program. This raises the possibility that these subjects did not view this activity as drug dealing. In future analyses, the authors plan to examine the impact of other possible discrepancies in definitions of drug dealing on the consistency of drug-dealing reports.

Support for the hypothesis that life history reports are an exaggeration of current behavior derives partly from previous observations in ethnographic research. In a previous study of heroin addicts, Goldstein (1981, p. 82) noted:

> When addicts are asked how much heroin they have used during the course of a year, or longer, they may very well respond in terms of the "ideal" addict—the one they would like to be but, in fact, approximate only infrequently.

> They may forget about those days when they were not able to get over, and as a result, used little or no heroin.

These observations underscore some of the special difficulties involved with obtaining reliable reports about drug use from subjects for whom drug use consumes a major social role. Overestimation is not the only problem that has been observed in the literature. Impairment from drug use at the time of the interview or residual effects from drugs after intoxication can affect subject responses. Adler (1993, p. 22) notes that subjects who were high on marijuana were particularly difficult to interview since they became "confused, sleepy, or involved in eating." Possible cognitive effects of drug involvement need to be considered when interpreting the reports of active drug users.

In comparing the findings derived from the life history reports to those in the weekly interviews, differences in structure between the two assessments must be underscored. The life history reports were relatively open ended; the historical recollection of behavior was relatively unprompted, and respondents were forced to provide their own parameters for initiation and cessation of behaviors and for estimates of typical patterns and dollar cost of drug consumption. The lack of structure made it more difficult to code quantitative values for comparison in the analysis; the coding of an unstructured instrument is subject to greater error and discrepancy.

In the semistructured life history format, many subjects were unable to provide quantitative estimates for recent behavior. In the weekly interviews, information was collected in a diary format; respondents were prompted to recall specific quantities (dollar amounts) of substances used on specific days over the course of the previous week. Additionally, respondents were prompted to provide detailed information about dollar income and specific sources of that income. In the more structured format, inconsistencies can be handled more directly in the interview. Because respondents were asked about sources of income leading to purchase and consumption of drugs, it may have been more difficult for them to deny ongoing involvement in income-generating criminal behavior such as drug dealing. Thus, the more structured format may have elicited better information about ongoing involvement in illicit criminal behavior and drug use. In an unstructured retrospective format, those who were most drug involved may have been most prone to exaggeration and overestimation of their typical behavior. The structured format with daily behavior prompts may have allowed for more realistic estimates of

behavior. This conclusion is supported by research in other contexts; for example, cognitive studies have shown that dietary recall and recall of health services use are aided by the provision of memory cues and prompts about recent activity and experience (Jobe and Mingay 1991).

Findings with respect to subjects' unwillingness to disclose current involvement in drug dealing in the life history format parallel findings described by Hser and colleagues (1992). In contrast, the authors' findings obtained from a comparison of drug use volume and frequency measures reinforce the findings of Collins and colleagues (1985), Johnson and colleagues (1985), and Czarnecki and colleagues (1990): Retrospective reports of drug use may overestimate actual (current) use. The contrast between reporting for drug dealing and drug use underscores the point that even in ethnographic studies, certain kinds of information may be perceived as sensitive. Drug dealing may be a more sensitive topic than drug use; willingness to disclose such involvement may emerge as subjects become more comfortable with the field site and the data-gathering process established by the ethnographers.

Findings with respect to drug use consistency patterns parallel the authors' previous work in this area in two respects. Just as the authors observed a decline in reported use levels over the 8-week interview period (Fendrich et al. 1992), they also saw a decline in use quantities reported in the life history in comparison to the weekly reports. This supports the notion that continued interviews about quantities of drug use may in fact result in a retest artifact, which has previously been discussed in the literature. As in a previous study with this sample (Fendrich et al. 1992), the authors found that use frequency reports were considerably more consistent than reports related to dollar amount; in contrast to correlations between volume measures (based on dollar amount), correlations between frequency measures at each phase of interviewing were generally at an acceptable level. These findings continue to raise questions about the utility of dollar-based quantitative measures in ethnographic research.

These findings warn against static, retrospective assessments of lifetime patterns of drug use behavior. As in earlier methodological studies of substance abuse (Aiken 1986; Anglin et al. 1993; Collins et al. 1985; Czarnecki et al. 1990; Johnson et al. 1985), retrospective accounts in DRIVE and FEMDRIVE diverged in important and significant ways from accounts of ongoing behavior. The present analyses show that many of the same issues related to self-disclosure of sensitive behavior

relevant to responses in more structured surveys such as the National Longitudinal Survey of Youth (Fendrich and Mackesy-Amiti 1995; Fendrich and Vaughn 1994) are also relevant to responses in ethnographic research. The qualitative nature of ethnographic narrative accounts often prevents the quantitative examination of reliability issues. The authors were fortunate to have access to an ethnographic data set that facilitated the codification of qualitative responses about behavior; the data contained comparable quantitative information about drug use behavior over a clearly defined followup period. The inconsistencies that were uncovered in this process suggest that an informal and anecdotal assessment of reliability is insufficient in ethnographic research. Any organized effort to collect behavioral information about drug involvement will result in less than perfect reliability. Researchers need to systematically assess the scope and impact of reporting inconsistency as a prerequisite to further substantive analytic work.

NOTES

1. These measures were created specifically for this report. The authors returned to the original data files and coded interview responses for this information.

2. The categories for the most recent frequency variable followed those originally created by the researchers who first coded the interview data; because this was a secondary data analytic project, the authors followed the coding scheme suggested by the original investigators.

3. The coding was constructed so as to create categories that were roughly equivalent to life history codes. Those with a ratio value of less than 0.5 were coded as infrequent users. Those with a ratio value of at least 0.5 but less than 2.5 were classified into the moderate use category. Those with a ratio value of at least 2.5 but less than 5.5 were considered to be regular users. Finally, those with a ratio of 5.5 or greater were considered to be daily users. For purposes of data analysis, the frequency categories derived from both the life history and the weekly reports were ordered from 1 to 4, with higher scores indicating higher levels of use.

4. Sample sizes vary due to missing values on the "age of last drug deal" question on the life history interview. McNemar chi-square tests reflecting shifts in reporting across interviews are not shown

but were all highly significant. For both lifetime dealing measures, significant coefficients reflected the shift to less drug dealing during weekly interviews; for recent dealing measures, significant coefficients reflected the shift to more drug dealing during weekly interviews.

5. According to Fleiss (1981), Kappa values of less than 0.40 reflect "poor" agreement, values of between 0.40 and 0.74 reflect "fair to good" agreement, and values of greater than 0.75 reflect "excellent" agreement. The authors are not aware of standards specific to conditional Kappa statistics; these descriptive standards are applied to both types of Kappa coefficients.

6. A respondent was counted as a current user in the life history report if one of three conditions was met: the respondent described use of the substance during the same calendar year as the interview; the age of last use reported by the respondent corresponded to the respondent's current age; or the respondent explicitly stated that he or she was "now" using during the life history interview.

7. The authors realize that the use of the word "underreporting" assumes that behavior that was actually occurring was not being reported. Of course, differences in reports can reflect actual behavior changes and problems with coding and classification; these possibilities are discussed below.

8. The life history cost-per-day measures and weekly interview volume measures are considered comparable volume indices; both gauge the amount of use per substance use occasion. In tables 4 and 5, both measures are considered volume indicators and are labeled as such. Comparisons in this section were limited to those disclosing current use in the life history reports and were based on an assessment of the most recent use pattern expressed in the life history.

9. The authors investigated differences on other available indicators of drug involvement for those with missing values on heroin and cocaine life history volume and frequency indices. Bivariate comparisons (not shown here) suggested that those who were missing on volume indices tended to report lower life history use frequency levels; this suggests that light users are probably underrepresented in the analyses performed in this section.

REFERENCES

Adler, P.A. *Wheeling and Dealing: An Ethnography of an Upper-Level Drug Dealing and Smuggling Community*. New York: Columbia University Press, 1993.

Aiken, L.S. Retrospective self-reports by clients differ from original reports: Implications for the evaluation of drug treatment programs. *Int J Addict* 21:767-788, 1986.

Anglin, M.D.; Hser, Y.; and Chou, C. Reliability and validity of retrospective behavioral self-report by narcotics addicts. *Eval Rev* 17:91-108, 1993.

Biernacki, P., and Waldorf, D. Snowball sampling: Problems and techniques of chain referral sampling. *Sociol Meth Res* 10(2):141-163, 1981.

Bishop, Y.; Fienberg, S.; and Holland, P. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: The MIT Press, 1975.

Bromet, E.; Dunn, L.; Connell, M.; Dew, M.; and Schulbert, H. Long-term reliability of diagnosing lifetime major depression in a community sample. *Arch Gen Psychiatry* 43:435-440, 1986.

Collins, L.; Graham, J.; Hansen, W.; and Johnson, C. Agreement between retrospective accounts of substance use and earlier reported substance use. *Appl Psychol Measur* 9:301-309, 1985.

Czarnecki, D.M.; Russell, M.; Cooper, M.L.; and Salter, D. Five-year reliability of self-reported alcohol consumption. *J Stud Alcohol* 51:68-76, 1990.

Fendrich, M., and Mackesy-Amiti, M.E. Inconsistencies in lifetime cocaine and marijuana use reports: Impact on prevalence and incidence. *Addiction* 90:111-118, 1995.

Fendrich, M., and Vaughn, C. Diminished lifetime substance use over time. An inquiry into differential underreporting. *Public Opin Q* 58:96-123, 1994.

Fendrich, M.; Goldstein, P.; Tarshish, C.; and Bellucci, P. Longitudinal measurement of substance use in ethnographic samples. *J Community Psychol* 20:236-242, 1992.

Fleiss, J.L. *Statistical Methods for Rates and Proportions*. New York: John Wiley and Sons, 1981.

Goldstein, P. Getting over: Economic alternatives to predatory crime among street drug users. In: Inciardi, J.A., ed. *The Drugs Crime Connection*. Beverly Hills, CA: Sage, 1981. pp. 67-84.

Goldstein, P.; Bellucci, P.; Spunt, B.; and Miller, T. Volume of cocaine use and violence: A comparison between men and women. *J Drug Issues* 21:345-367, 1991.

Goldstein, P.; Bellucci, P.; Spunt, B.; Miller, T.; Cortez, N.; Khan, M.; Durrance, R.; and Vega, A. "Female Drug Related Involvement in Violent Episodes (FEMDRIVE)." Report prepared for the National Institute on Drug Abuse, 1988.

Goldstein, P.; Lipton, D.; Spunt, B.; Bellucci, P.; Miller, T.; Cortez, N.; Khan, M.; and Kale, A. "Drug Related Involvement in Violent Episodes (DRIVE)." Interim final report prepared for the National Institute on Drug Abuse under NIDA grant no. DA-03182, 1987.

Goldstein, P.; Spunt, B.; Miller, T.; and Bellucci, P. Ethnographic field stations. In: Lambert, E.Y., ed. *The Collection and Interpretation of Data from Hidden Populations*. National Institute on Drug Abuse Research Monograph 98. DHHS Pub. No. (ADM)90-1678. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Hser, Y.; Anglin, D.; and Chou, C. Reliability of retrospective self-report by narcotics addicts. *Psychol Assess* 4(2):207-213, 1992.

Jobe, J., and Mingay, D. Cognition and survey measurement: History and overview. *Appl Cog Psychol* 5:175-192, 1991.

Johnson, B.D.; Goldstein, P.; Preble, E.; Schmeidler, J.; Lipton, D.; Spunt, B.; and Miller, T. *Taking Care of Business: The Economics of Crime by Heroin Abusers*. Lexington, MA: DC Heath, 1985.

Jorm, A.F.; Duncan-Jones, P.; and Scott, R. An analysis of the retest artifact in longitudinal studies of psychiatric symptoms and personality. *Psychol Med* 19:487-493, 1989.

Lord, F.M., and Novick, M.R. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley Publishing Company, 1968.

Rubio-Stipec, M.; Freman, D.H.; Robins, L.; Shrout, P.; Canino, G.; and Bravo, M. Response error and the estimation of lifetime prevalence and incidence in alcoholism: Experience in a community survey. *Int J Meth Psychiatr Res* 2:501-508, 1992.

Stephens, R.; Feucht, T.; and Roman, S. Effects of an intervention program on AIDS-related drug and needle behavior among intravenous drug users. *Am J Public Health* 81:568-571, 1991.

Waldorf, D.; Reinarman, C.; and Murphy, S. *Cocaine Changes*. Philadelphia: Temple University Press, 1991.

Weibel, W.W. Combining ethnographic and epidemiologic methods in targeted AIDS interventions: The Chicago model. In: Battjes, R.J., and Pickens, R.W., eds. *Needle Sharing Among Intravenous Drug Abusers: National and International Perspectives*. National Institute on Drug Abuse Research Monograph 80. DHHS Pub. No. (ADM)88-1567. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1988.

ACKNOWLEDGMENT

This research was prepared with support from National Institute on Drug Abuse grant no. R29DA07995.

AUTHORS

Michael Fendrich, Ph.D.
Assistant Professor of Psychology in Psychiatry

Mary Ellen Mackesy-Amiti, Ph.D.
Research Specialist

Joseph S. Wislar
Research Scientist

Institute for Juvenile Research
Department of Psychiatry
University of Illinois at Chicago
Mail Code 747
907 South Wolcott Avenue
Chicago, IL 60612

Paul J. Goldstein, Ph.D.
Associate Professor
Department of Epidemiology and Biostatistics
School of Public Health
University of Illinois at Chicago
Mail Code 922
907 South Wolcott Avenue
Chicago, IL 60612

167344

New Developments in Biological Measures of Drug Prevalence

Edward J. Cone

ABSTRACT

Drug use among different populations such as household members, students, and arrestees vary substantially and the accuracy of their self-reports may be questionable. The accuracy of prevalence estimates based on self-report data can be monitored by chemical drug testing of biological specimens such as urine, saliva, sweat, and hair. Each biological specimen is unique and offers a somewhat different pattern of information regarding drug use over time. Also, each specimen has unique strengths and weaknesses regarding the type of information obtained from drug testing. The performance characteristics of the assay methodology may also be important. The validation of self-report data by drug testing must be performed with careful consideration of the limitations imposed by the testing methodology and the biological specimen.

INTRODUCTION

Illicit drug administration is often perceived by society to be risky or antisocial in nature. Such behavior can lead to many unfavorable outcomes for the individual and for society at large. The frequency of illicit drug use within various populations is a subject of much speculation and study. Drug policy decisions and intervention efforts aimed at reduction of illicit drug usage are often predicated on drug use measurements obtained through self-reports of drug use history. In the United States, drug prevalence estimates are obtained primarily from three sources: National Household Survey on Drug Abuse, Monitoring the Future survey, and Drug Abuse Warning Network (DAWN). Each involves data collection based partially or totally on self-reported drug use.

The validity of self-reported drug use data is subject to many influencing factors, such as the population examined, type of drugs used, environment, and methods used to elicit information (Magura et al. 1987). In addition, accurate recall of drug use by an individual can be affected directly by

their current mental and physical status. Underreporting of drug use is common in some populations, particularly those in which real or perceived punitive measures may result from admission of drug use. The problem of underreporting led the National Institute of Justice to establish the national Drug Use Forecasting (DUF) program in 1987 as a new source of drug prevalence estimates in which recent drug use trends in arrestees were measured by urinalysis. Early data from that program indicated that urinalysis revealed substantially higher rates of cocaine use than was indicated by self-report data (Wish 1990-1991).

The inclusion of more objective measures of drug use, such as urinalysis, complements self-report data and provides added assurance of the accuracy of prevalence estimates. The technology of urinalysis has progressed rapidly over the past decade because of widespread implementation of drug testing programs by the Federal Government, the military, and private industry. The need for reliable, inexpensive urine-based drug tests led to significant efforts in research and commercial development of such tests. At the same time, research has progressed on the evaluation of other biological fluids and tissues as useful matrices for drug detection.

Currently, there is growing interest in the use of alternate body fluids and tissues such as saliva, sweat, and hair in addition to urine for the diagnosis of drug use. The following discussion provides an overview of the validity of drug testing and the potential uses and limitations of urine, saliva, sweat, and hair testing for drugs of abuse as objective drug prevalence estimates in different populations.

CHARACTERISTICS OF DRUG-TESTING METHODS

The usefulness of a drug test resides in its ability to accurately detect the presence of parent drug or metabolite in a biological fluid or tissue following human drug administration. This ability has been referred to as the validity of the test system (Gorodetzky 1977). This definition reflects both chemical factors that influence test outcome such as sensitivity (the least amount of detectable drug), specificity (how selective the assay is for the drug), and accuracy, and pharmacologic considerations including dose, time of drug administration, and route of drug administration. Individual differences in rate of absorption, metabolism, and excretion are also pharmacologic variables that may influence test outcome. With the recent emphasis on forensic drug

testing, Cone and colleagues (1988) suggested that the definition of validity be extended to include confirmation of initial test results by a different chemical method (e.g., gas chromatography-mass spectrometry (GC/MS)). When there is a possibility of litigation, it is extremely important to use assay methods that are highly accurate, reliable, and specific for the analyte of interest.

A variety of commercial assays and published methodologies may be employed for urine drug testing. For the most part, these methods can be grouped into two categories: screening assays and confirmation assays. The performance characteristics of these assays are listed in table 1. The assays can also be adapted for measurement of drugs in other body fluids, but must be properly validated before use. Generally, screening assays (immunoassays and thin-layer chromatography (TLC)) are commercially based tests that are inexpensive and simple to perform. In contrast, confirmation assays (gas chromatography (GC) and GC/MS) are more expensive and more labor intensive, but sensitivity and specificity are usually higher than screening tests. Immunoassay-based screening tests may cross-react with a variety of similar chemical substances. For example, most commercial immunoassays for opiates give positive test results for specimens containing either morphine or codeine. In this case, a more specific methodology is needed if it is important to distinguish between these two drugs. Often, the less expensive screening tests are employed to eliminate specimens containing no drug or drug below the cutoff concentration. The more expensive, labor-intensive tests are employed for absolute drug identification and accurate quantitation.

For drug prevalence studies in which individuals are not identified, it becomes less important to employ expensive confirmation techniques unless there are known interferences within a particular assay. Indeed, some screening assays have shown exceptionally high correlations with GC/MS methods. For example, Cone and associates (1988) reported that urine test results from a specific assay for cocaine metabolite significantly correlated with results by GC/MS with no evidence of assay bias. Consequently, in many cases it may be more cost effective to use a highly selective immunoassay than to pay for the additional costs of confirmation. An added bonus often is realized when immunoassays are employed because of their rapid turnaround time. Results may be available immediately in some cases, and almost always are provided within 24 hours of receipt at the laboratory. It is also important to select an assay system with results that can be compared with those from other studies. Many comparisons between different assay systems are not valid

TABLE 1. *Performance characteristics of different types of assays for drugs of abuse.*

Assay	Sensitivity	Specificity	Accuracy	Turnaround time	Cost
Onsite	Moderate-high	Moderate	Qualitative*	Minutes	$4-25
EMIT, FPIA, RIA, KIMS	Moderate-high	Moderate	Low-high	1-4 hours	$1-5
TLC	Low-high	High	Qualitative*	1-4 hours	$1-4
GC	High	High	High	Days	$5-20
GC/MS	High	High	High	Days	$10-100

KEY: EMIT = enzyme-multiplied immunoassay technique; FPIA = fluorescent polarization immunoassay; RIA = radioimmunoassay; KIMS = kinetic interaction of microparticles in solution; TLC = thin layer chromatogra phy; GC = gas chromatography; GC/MS = GC/mass spectrometry. * = Results for onsite tests and TLC assays are generally expressed only in qualitative terms (i.e., positive/negative); consequently, accuracy may be difficult to assess.

simply because the immunoassay antibodies utilized in the assay were not targeted toward the same drug or metabolite. Further, even in situations in which the same assay system is employed, comparisons must be made on equal ground. A simple change in the cutoff concentration of an assay can substantially alter the detectability of a drug. Figure 1 illustrates the influence of detection time on cutoff concentrations. Obviously, it is important to select drug assays with equivalent performance characteristics if comparisons within and between studies are anticipated.

Urine

Urine is produced continuously by the kidney as an ultrafiltrate of blood. During urine production, essential substances are reabsorbed by the kidney, and excess water and waste products such as urea, organic substances, and inorganic substances are eliminated from the body. The daily amount and composition of urine varies widely depending upon

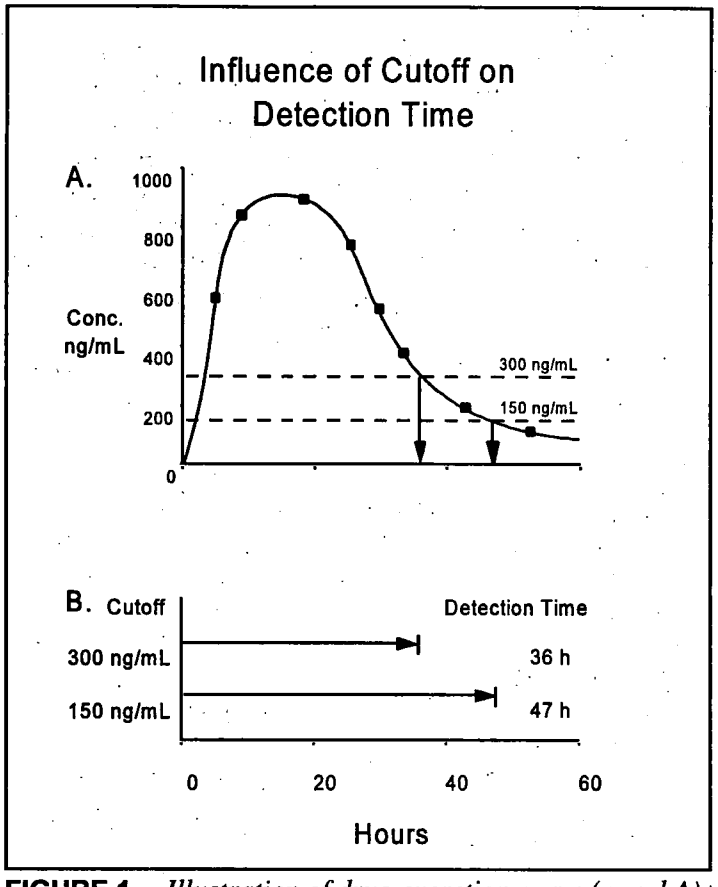


FIGURE 1. *Illustration of drug excretion curve (panel A) and relationship of detection time to cutoff concentration (panel B).*

many factors such as fluid intake, diet, health, drug effects, and environmental conditions. The volume of urine produced by a healthy adult in a 24-hour period ranges from 1 to 2 liters, but normal values outside these limits are frequently encountered. Creatinine, a protein byproduct of muscle metabolism, is present at a relatively constant concentration in blood and is excreted in urine. Consequently, the average 24-hour output of creatinine in urine also is constant. Comparison of creatinine concentration in urine and blood provides a means of assessing renal function. For most people, urine creatinine concentrations exceed 20 milligrams per deciliter (mg/dL) although concentrations lower than 20 mg/dL are occasionally encountered.

Urine specimens with creatinine concentrations below 20 mg/dL can be produced by excessive water intake. Drug users who are being urine tested sometimes attempt evasion by drinking large amounts of water or herbal teas in an attempt to dilute drug concentrations below cutoff concentrations. Consequently, many laboratories also test for creatinine and report specimens with creatinine concentrations below 20 mg/dL. Medical review officers who review results with abnormally low creatinine concentrations may request retesting the subject for drugs. Drug/creatinine ratios can be evaluated for evidence of attempted dilution of urine. A highly dilute specimen might test negative, but evaluation of the drug/creatinine ratio could provide convincing evidence that the sample would have been positive if normal water intake had occurred.

When a drug is administered by an intravenous or smoking route, absorption is nearly instantaneous and excretion in urine begins almost immediately. Absorption is slower when a drug is administered by the oral route and excretion in urine may be delayed for several hours. Normally, specimens voided within 6 hours after drug administration contain the highest concentration of parent drug and metabolites. Because drug excretion in urine normally occurs at an exponential rate, the majority of the drug dose of most illicit drugs is eliminated within 48 hours after administration. Detection times for drugs of abuse vary according to dose, frequency of administration, cutoff concentration, and numerous other factors. Despite wide variance, it is helpful to know average detection times when interpreting urine test data. Table 2 contains a list of average detection times and commonly used cutoff concentrations.

Most drugs of abuse have detection times of 2 to 4 days unless accumulation has occurred as a result of frequent, multiple dosing over an extended period of time. In drug prevalence studies, the relatively brief historical record of drug exposure provided by urinalysis must be considered when compared to retrospective self-report data. Urinalysis may cover only 2 to 4 days, but self-report data may encompass longer periods. Subjects who accurately report drug use within the past month could easily have negative urine results. In this case, the urine result does not support the self-report data. A better comparison can be made through the use of discrete multivariate analysis in which self-report data are compared with positive urine test results (Bishop et al. 1975; Magura et al. 1987). The reports of subjects with negative urine results are ignored, and subjects with a positive urine test who fail to report recent

TABLE 2. *Typical screening and confirmation cutoff concentrations and detection times for drugs of abuse.*

Drug	Screening cutoff concentrations ng/mL urine	Analyte tested in confirmation	Confirmation cutoff concentrations	Urine detection time
Amphetamine	1,000	Amphetamine	500	2-4 days
Barbiturates	200	Amobarbital, secobarbital, other barbiturates	200	2-4 days for short acting; up to 30 days for long acting
Benzodiazepines	200	Oxazepam, diazepam, other benzodiazepines	200	Up to 30 days
Cocaine	300	Benzoylecgonine	150	1-3 days
Codeine	300	Codeine, morphine	300; 300	1-3 days
Heroin	300	Morphine, 6-acetylmorphine	300; 10	1-3 days
Marijuana	100; 50; 20	Tetrahydrocannabinol	15	1-3 days for casual use; up to 30 days for chronic use
Methadone	300	Methadone	300	2-4 days
Methamphetamine	1000	Methamphetamine, amphetamine	500; 200	2-4 days
Phencyclidine	25	Phencyclidine	25	2-7 days for casual use; up to 30 days for chronic use

drug use are considered inaccurate reporters. Using this approach, Magura and associates (1987) found that self-reporting of methadone clients was least valid for opiates, while benzodiazepine and cocaine reporting were moderately and highly valid, respectively.

Self-reported drug use data can be compared with either qualitative (positive/negative) or quantitative urinalysis. Most comparisons that involve collection of a single urine specimen are made in the qualitative mode. In situations where multiple specimens are collected, particularly treatment and rehabilitation, quantitative urinalysis provides additional information that may be useful in determining whether drug use has decreased (Batki et al. 1993). Since cocaine metabolite is excreted for periods up to 4 days following use, several sequential samples collected within a short time period may be positive as a result of a single drug episode. Qualitative urinalysis provides multiple positive results from these episodes and overestimates the frequency of cocaine use. This problem is illustrated in figure 2, which shows results from a control subject in a cocaine treatment study who was urine tested 3 times a week. The study consisted of an initial 5-week period (baseline) during which all subjects reported to the outpatient treatment clinic and received counseling, followed by a 12-week period in which some subjects received contingency management therapy. The urine test data indicated that this subject was using cocaine sporadically throughout the baseline and treatment periods. Qualitative analysis indicated that the subject produced specimens positive for cocaine 73 percent of the time during the baseline period and 81 percent of the time during the treatment phase.

Clearly, some specimens were positive as a result of new cocaine use, while others simply represented carryover from earlier dosing. If one evaluates these data in a quantitative mode with rules that would identify instances of new drug use, an estimate of the frequency of drug use can be obtained. The rules for estimating instances of new cocaine use must be based on the known pharmacokinetic parameters of the cocaine metabolite, benzoylecgonine. Because this metabolite has an excretion half-life of approximately 7.5 hours (Ambre 1985), sequential urine specimens that are collected at intervals > 24 hours should have declined in concentration by more than 25 percent of the concentration of the previous positive specimen. For example, the third specimen in figure 2 showed a concentration of 48,810 nanograms (ng)/mL of benzoylecgonine equivalents and the fourth sample contained 11,540 ng/mL. Thus, the fourth sample likely was positive as a result of carryover. The fifth sample continued to decline in concentration (241 ng/mL), whereas the sixth sample represented a new occurrence of cocaine use (252,000 ng/mL). Application of new-use rules to figure 2 indicate that only 53 percent of the specimens represented instances of new use compared to positive rates of 73 percent and 81 percent by qualitative analysis.

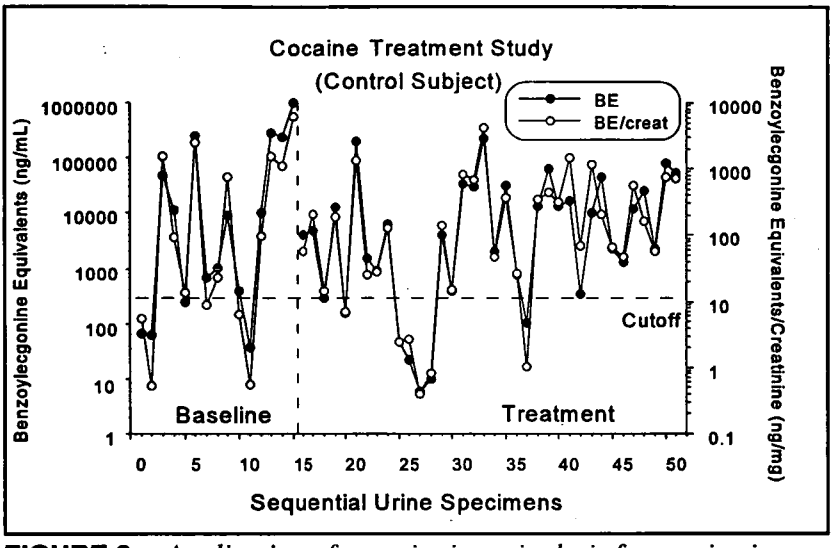


FIGURE 2. *Application of quantitative urinalysis for monitoring cocaine abuse in a treatment subject.*

KEY: Be = benzoylecgonine; creat = creatinine.

The value of quantitative urine test data is also evident when samples are tested for water dilution. Water loading can cause positive samples to test negative in the qualitative mode. However, the use of benzoylecgonine/creatinine ratios can provide evidence that a dilute sample would ordinarily have tested positive. In the example shown in figure 2, the benzoylecgonine/creatinine ratios closely parallel benzoylecgonine concentrations, indicating no evidence of attempted dilution by this individual.

Saliva

Saliva is secreted primarily by three glands: the parotid, submandibular, and sublingual glands. Secretions from serous and mucousal cells in these glands form saliva. Serous cells secrete watery fluid containing electrolytes and amylases and mucous cells produce mucins (mucoproteins and mucopolysaccharides). The flow of saliva is dependent upon neurotransmitter stimulation and can vary widely from zero flow to rates as high as 10 mL/minute. The pH of saliva generally is slightly acidic but increases with saliva flow rate from a low of pH 5.5 to pH 7.9. Saliva composition is also dependent upon flow, but generally consists of approximately 90 percent water with the remainder being electrolytes (e.g., sodium,

calcium, bicarbonates, magnesium), amylase, organics (glucose, urea, lipids), proteins (low concentrations), and hormones (cortisol, testosterone, estrogens, progesterone).

Drugs may enter saliva by passive diffusion from blood, ultrafiltration, and active secretion. Of these processes, passive diffusion represents the most important route of entry for most drugs with the possible exception of ethanol, a molecule small enough to enter by ultrafiltration. Several reports and reviews have appeared on the occurrence of drugs of abuse in saliva (Caddy 1984; Cone 1993; Schramm et al. 1992).

Saliva offers a number of advantages and some disadvantages in comparison to urine testing for drugs. The major advantages of saliva as a test medium include its ready accessibility for collection, less objectionable nature (compared to urine), presence of parent drug in higher abundance relationship between plasma morphine concentrations and saliva morphine following the injection of morphine by the intramuscular route (Cone 1993) is illustrated in figure 3. Saliva concentrations are reduced relative

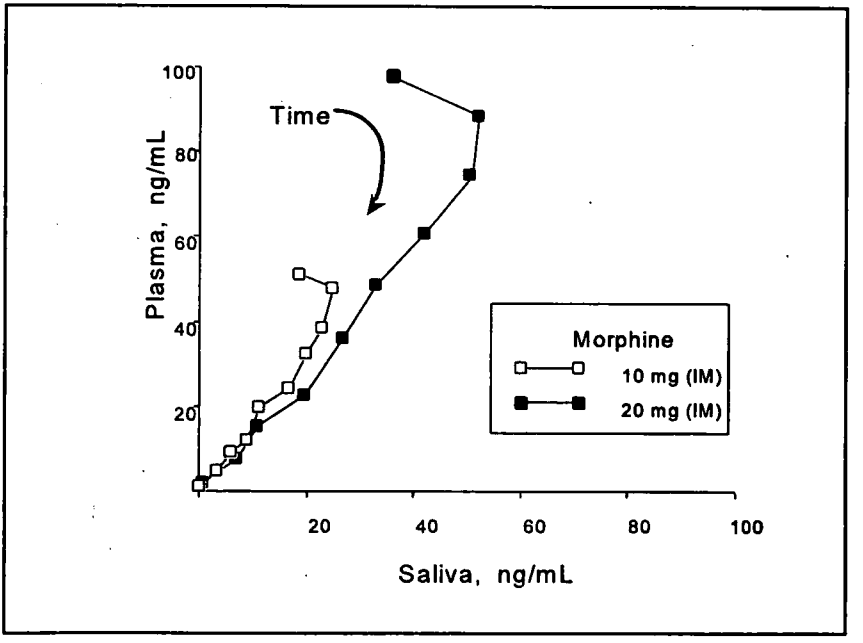


FIGURE 3. *Hysteresis plot of mean morphine concentrations in saliva and plasma of human subjects (N = 6) following intramuscular (IM) administration of morphine.*

SOURCE: Reproduced from Cone 1993.

to plasma by approximately one-third, equivalent to the amount of plasma protein binding for morphine. After a very short equilibration phase (15 to 30 minutes), saliva morphine declined in a manner parallel to plasma concentrations. Significant correlations of saliva drug concentrations with plasma have also been reported for cocaine. Cone and colleagues (1988) reported finding significant correlations of saliva than metabolites, and high correlation of saliva drug concentration that can be obtained with the free fraction of drug in blood (table 3). The cocaine concentrations with plasma and with responses on self-rating scales for drug sensation, psychotomimetic effects and feelings of rush, and heart rate. Figure 4 illustrates the temporal relationship between saliva cocaine concentrations, plasma, and heart rate changes following a 25 mg dose of cocaine hydrochloride (HCl) salt to a cocaine user by the intravenous route. It is clear from this illustration that saliva and plasma concentrations are similar for most of the time period and decline with heart rate in a parallel manner. The equivalent concentrations of cocaine in saliva and plasma are the result of pH influences and the lack of protein binding by cocaine in plasma.

Despite the numerous advantages of saliva, it does have some disadvantages. The use of saliva drug concentrations to predict blood concentrations is limited because of the possibility of contamination of saliva from drug use by the oral, smoked, and intranasal routes of drug administration. When drugs are administered by these routes, contamination of the oral cavity and saliva can greatly distort saliva/plasma ratios, thereby distorting useful pharmacokinetic relationships. Even with this obvious limitation, saliva measurements can be used as evidence of recent drug use even in situations in which oral contamination is likely to be involved (e.g., marijuana smoking). Cone (1993) reported that marijuana smoking produced contamination of the oral cavity by tetrahydrocannabinol (THC). Even though saliva concentrations of THC were derived from contamination, they were highly correlated with plasma concentrations.

The short time course for detectability of drugs in saliva prevents this biological fluid from being used to detect historical drug use. At the same time, this feature of saliva makes it useful for detection of very recent drug use. Most drugs disappear from saliva and blood within 12 to 24 hours after administration. There is often a temporal relationship between the disappearance of drugs in saliva and the duration of pharmacologic effects. Consequently, saliva is useful in the detection of recent drug use in automobile drivers, accident victims, and for testing employees before they engage in safety-sensitive activities.

TABLE 3. *Comparison of usefulness of urine, saliva, sweat, and hair as a biological matrix for drug detection.*

Biological matrix	Drug detection time	Major advantages	Major disadvantages	Primary use
Urine	2-4 days	Mature technology; onsite methods available; established cutoffs	Only detects recent use	Detection of recent drug use
Saliva	12-24 hours	Easily obtainable; samples "free" drug fraction; parent drug presence	Short detection time; oral drug contamination; collection methods influence pH and saliva/plasma ratios; only detects recent use; new technology	Linking positive drug test to behavior and performance impairment
Sweat	1-4 weeks	Cumulative measure of drug use	High potential for environmental contamination; new technology	Detection of recent drug use
Hair	months	Long-term measure of drug use; similar sample can be recollected	High potential for environmental contamination; new technology	Detection of drug use in recent past (1-6 months)

Sweat

Sweat is a watery fluid produced primarily by eccrine glands distributed widely across the skin surface of humans. The primary purpose of sweat production is heat regulation; consequently, the amount of sweat produced is highly dependent upon environmental conditions. Sweat consists mostly of water (99 percent) with the greatest concentrated solute being sodium chloride (Robinson and Robinson 1954). Routine sweat collection is difficult because of large variations in the rate of sweat production and the lack of devices suitable for collection of this type of biological fluid.

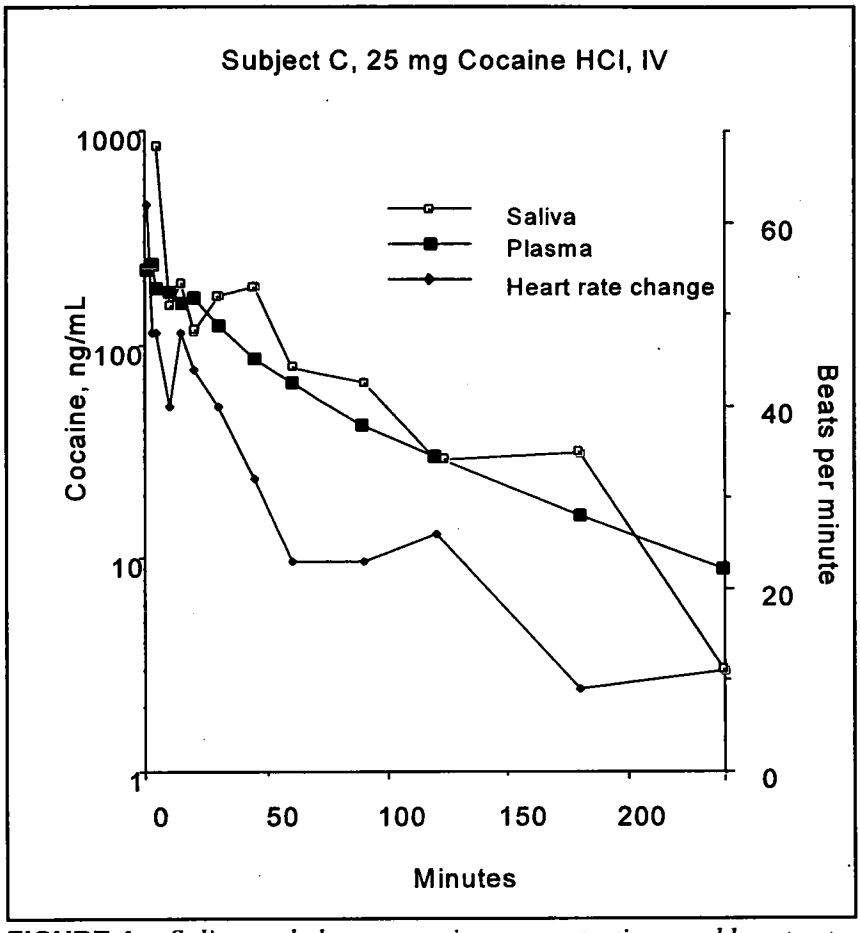


FIGURE 4. *Saliva and plasma cocaine concentrations and heart rate changes in a subject following administration of 25 mg of cocaine HCl by the intravenous route.*

A variety of drugs of abuse have been identified in sweat, including amphetamine, cocaine, ethanol, methadone, methamphetamine, morphine, nicotine, and phencyclidine. The mechanism for drug entry into sweat is unclear, but most likely occurs by passive diffusion from blood to the sweat gland. An alternate mechanism could involve drug diffusion through the stratum corneum to the skin surface where drug would be dissolved in sweat.

Research on sweat testing for drugs has been limited because of the difficulty in collecting sweat samples. Recently, a sweat-collection device was developed that appeared to offer promise for the collection of

sweat for drug monitoring. The device resembles an adhesive bandage that is applied to the skin and can be worn for a period of several days to several weeks. The "sweat patch" consists of an adhesive layer on a thin transparent film of surgical dressing to which a rectangular absorbent cellulose pad (14 square centimeters (cm^2)) is attached. Sweat is absorbed and concentrated on the cellulose pad. The transparent film allows oxygen, carbon dioxide, and water vapor to escape, but prevents loss of drug constituents excreted in sweat. Over a period of several days, sweat should saturate the pad and drug should slowly concentrate. The patch is then removed, and the absorbent pad is detached from the device and analyzed for drug content.

Sweat testing for cocaine was recently evaluated by Cone and associates (1994). Cocaine was administered in doses of 1 to 25 mg by the intravenous route to four cocaine-experienced, drug-free subjects. Sweat patches were worn for 24 to 48 hours following drug administration. Following removal, the patches were extracted and analyzed for cocaine and metabolites by GC/MS. The primary analyte excreted in sweat was parent cocaine, followed by ecgonine methyl ester and benzoylecgonine. Figure 5 illustrates the relationship between amount of cocaine collected by the sweat patch versus dose. Generally, there appeared to be a dose-concentration relationship; however, there was wide intersubject variability. Limited data were also collected in the same study on the excretion of heroin in sweat. Like cocaine, parent heroin was excreted in sweat along with metabolites that consisted of 6-acetylmorphine and morphine. Drug appeared in sweat as early as 1 hour following drug administration and peaked in concentration within 24 hours.

Apparent advantages of the sweat patch for drug monitoring include the following: high subject acceptability of wearing the patch, low incidence of allergic reactions to the patch adhesive, and ability to monitor drug intake for a period of several weeks with a single patch. In addition, the patch appears to be relatively tamper-proof (i.e., the patch adhesive is specially formulated so that the patch can only be applied once and cannot be removed and successfully reapplied to the skin surface).

Disadvantages of the sweat patch includes high intersubject variability, possibility of environmental contamination of the patch before application or after removal, and risk of accidental removal during a monitoring period. During patch application, extreme care must be taken to cleanse the skin surface prior to placement of the patch and also to avoid

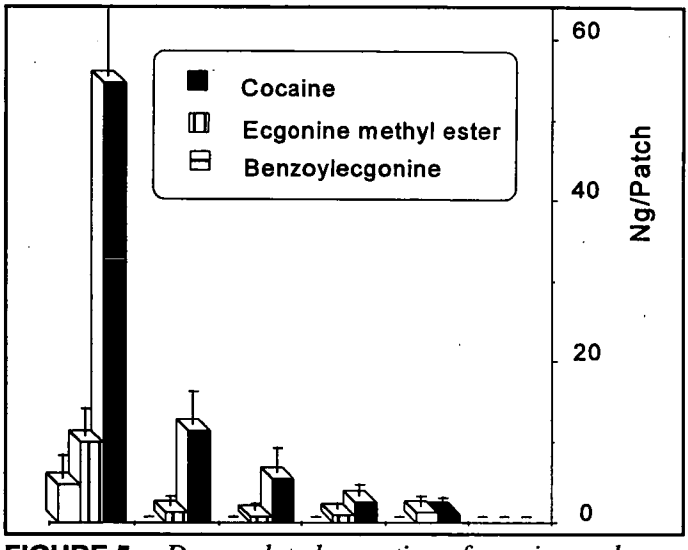


FIGURE 5. *Dose-related excretion of cocaine and metabolites in sweat. Sweat patches were worn for 24 hours following intravenous administration of a single dose of cocaine hydrochloride. Data represent the mean of four subjects with associated standard-error-of-the-mean error bars.*

SOURCE: Reproduced from Cone et al. 1994.

contamination of the cellulose pad during handling. Similar care must be taken when removing the patch and handling for analysis.

Hair

Hair is composed primarily of a fibrous network of keratin strands that are intertwined to form elongated strands. The strands are stabilized by interlinking disulfide and hydrogen bonds, which gives hair a semicrystalline structure. The inner structure of hair is protected by a layer of cuticle cells that restricts or retards entry of environmental pollutants. As hair ages, the cuticle deteriorates from exposure to ultraviolet radiation, chemicals, and mechanical stresses. Head hair grows at an average rate of 1.3 cm/month, although there is some variation according to sex, age, and ethnicity (Saitoh et al. 1969). Collection of hair for testing is most often performed by cutting locks of hair near the scalp surface at the vertex of the head. During collection, the root and tip of

the hair lock are identified for later use. Other types of hair, such as pubic, axillary, and arm hair, have also been used for drug testing.

Hair testing for drugs was first reported by Goldblum and associates (1954). Guinea pigs were administered varying doses of barbiturates and newly grown hair was found to be positive for parent drug. Baumgartner and colleagues (1982) reported the first evidence of drug in human hair by analyzing head hair of cocaine abusers by RIA for benzoylecgonine, the major metabolite of cocaine. Many other reports have subsequently appeared regarding the presence of drugs in hair. Drug representatives from virtually all classes of abused drugs have now been detected in hair. Currently, hair testing for drugs of abuse is performed in numerous laboratories, some of which offer commercial drug-testing services.

When hair is analyzed for drugs of abuse, the parent drug is often present in greater abundance than is found in urine. For example, the major analyte found in hair of cocaine users is parent cocaine. Benzoylecgonine, the primary urinary metabolite, is present in hair in amounts varying from trace concentrations to approximately one-third of parent cocaine (Cone et al. 1991). Heroin is found in hair in varying amounts together with 6-acetylmorphine and morphine (Goldberger et al. 1991). 6-Acetylmorphine is usually found in greatest abundance in hair, whereas conjugated morphine is the major metabolite in urine. Patterns of parent drug and metabolite distribution in hair and other biological matrices are listed in table 4.

Although the technology of hair assay has progressed rapidly over the last decade, several highly controversial aspects of hair testing remain unresolved. It remains unclear how drugs enter hair. The most likely entry routes involve: (a) diffusion from blood into the hair follicle and hair cells with subsequent binding to hair cell components; (b) excretion in sweat, which bathes hair follicles and hair strands; (c) excretion in oily secretions into the hair follicle and onto the skin surface; and (d) entry from the environment. The possibilities of drug entry from sweat and from the environment are particularly troubling, because this allows the possibility of the production of false positives if an individual's hair absorbs drugs from the environment or from another person's drug-laden sweat. Another controversial issue in hair analysis is the interpretation of dose and time relationships. Although it has been generally assumed that segmental analysis of hair provides a record of drug usage, studies with labeled cocaine have not supported this interpretation. At best, only limited dose and time relationships were found. Henderson and colleagues

TABLE 4. *Relative occurrence of parent drug and metabolite(s) in urine, saliva, sweat, and hair.*

Drug	Urine	Saliva	Sweat	Hair
Amphetamine	Amphetamine	Amphetamine	Amphetamine	Amphetamine
Cocaine	BE > EME > cocaine	Cocaine > BE ≈ EME	Cocaine > EME > BE	Cocaine > BE > EME
Marijuana	Carboxy-metabolite	THC	THC	?
Heroin	MO-glucuronide > MO	Heroin ≈ 6-AM > MO	Heroin ≈ 6-AM > MO	6-AM > heroin ≈ MO
CO	CO-glucuronide > CO > norcodeine	CO	CO	CO > MO
Metham	Metham > amphetamine	Metham	Metham	Metham > amphetamine
Phencyclidine	Phencyclidine	Phencyclidine	Phencyclidine	Phencyclidine
MO	MO-glucuronide > MO	MO	MO	MO

KEY: Metham = methamphetamine; MO = morphine; CO = codeine; BE = benzoylecgonine; EME = ecgonine methyl ester; THC = tetrahydrocannabinol; 6-AM = 6-acetylmorphine.

(1993, p. 2) concluded that "...there is not, at present, the necessary scientific foundation for hair analysis to be used to determine either the time or amount of cocaine use." Other controversial issues that remain unresolved are the possibility of ethnic bias in hair testing, appropriate means of differentiating drug users' hair from environmentally contaminated hair, appropriate applications of hair testing, and the feasibility of hair testing for marijuana usage.

Despite the controversial nature of some aspects of hair testing, this technique is being used on an increasingly broad scale in a variety of circumstances. One of the most promising applications of hair testing appears to be its use in prevalence studies. The time record of drug use available from hair is considerably longer than any other biological specimen currently employed for drug testing (figure 6). Self-reported drug use over a period of several months can be compared to hair test results from a hair strand (about 3.9 cm length) representative of the same time period. It is expected that this type of comparison would be more effective than urine testing because urine provides a historical record of only 2 to 4 days under most circumstances. Indeed, Mieczkowski and associates (1991) compared self-reported cocaine use

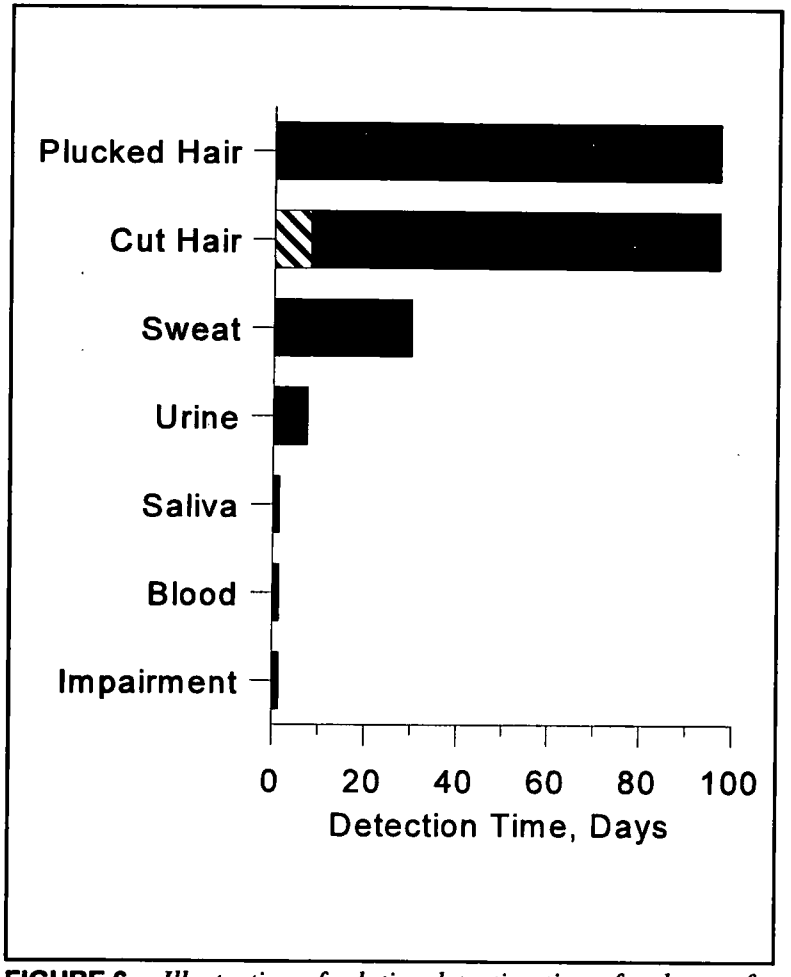


FIGURE 6. *Illustration of relative detection times for drugs of abuse in different biological fluids and tissues compared to duration of pharmacologic effects. Cross-hatched area of cut hair represents time required for hair to grow to skin surface.*

with hair and urine analysis in a group of 256 arrestees and found that hair analysis detected more drugs than either urinalysis or self-report. Of the 256 interviewed, 8.5 percent of the arrestees reported cocaine use within the past 30 days and 21.8 percent had positive urine tests, whereas 55 percent had positive hair tests. In a similar study involving 88 juvenile

arrestees, Feucht and colleagues (1994) found that only 3 individuals (3.4 percent) admitted cocaine use in the past month and only 7 subjects (8 percent) were positive by urinalysis, whereas 50 individuals (56.8 percent) were positive by hair analysis.

Other populations have shown somewhat higher concordance between hair assay and urinalysis or self-report. Magura and associates (1992) studied heroin addicts (N = 134) in which hair test results for opiates and cocaine were compared to confidential urinalysis and self-reporting. Hair test results were equivalent to urinalysis and/or self-report in 87 percent and 84 percent of the cases for cocaine and heroin, respectively. These data suggest the reliability of self-report data is highly dependent upon the population and the circumstances under which the data are collected.

Generally, hair analysis provides a longer estimate of drug use than either self-report measures or urinalysis. The wider window of detection is an advantage of hair testing as a prevalence measure for drug use. Other advantages include ease of obtaining, storing, and shipping specimens; ability to obtain a second sample for reanalysis; low potential for evasion or manipulation of test results; and low risk of disease transmission in the handling of samples. A potential disadvantage of hair analysis would be its inability to detect recent drug usage because of slow growth rate; however, this has not been investigated. Mounting evidence points to the likelihood that drug excretion in sweat is an important route of drug entry into hair. This allows the possibility of drug appearing in hair within hours of drug administration. Also, plucked hair should not have this limitation because hair below the scalp is removed (figure 6). Another consideration regarding the use of hair analysis is the limited number of laboratories offering commercial hair-testing services. Clearly, as demand for hair-testing services grows, commercial development also will proceed in simultaneous fashion. In addition, as more attention is focused on this new area of drug testing, many of the early controversies may be resolved by additional research.

SUMMARY

Drug use among different populations such as household members, students, and arrestees vary substantially and the accuracy of their self-reports may be questionable. Accurate assessment of drug prevalence in different populations helps policymakers identify

vulnerable groups and geographical areas with high rates of drug use. The accuracy of prevalence estimates based on self-report data can be monitored by drug testing biological specimens such as urine, saliva, sweat, and hair. Qualitative urinalysis (positive/negative drug use) is the most widely used technique and provides an objective measure of determining whether recent drug use has occurred over the past 2 to 4 days. Recently, interest has grown in the use of quantitative urine testing (concentration-based testing). Quantitative urine testing may further improve the usefulness of urinalysis by allowing intra- and intergroup comparisons of frequency and extent of drug use. Saliva testing, in comparison to urinalysis, offers different information regarding recency of drug use. The detection times for drugs in saliva are similar to those for blood (4 to 24 hours). Consequently, saliva testing may offer the possibility of revealing current drug use that affects an individual's performance in such complex psychomotor tasks as driving and operating heavy equipment.

Sweat testing has recently become feasible through the development of a new sweat patch device designed to collect nonvolatile drugs of abuse from human skin. The device is applied to the skin like an adhesive bandage. Substances with volatility equal to or greater than water leave the device through a membrane barrier. Less volatile substances (such as drugs) are concentrated on an absorption pad inside the patch. Subjects can wear the patch for periods up to several weeks, followed by removal, storage, and analysis of the contents of the absorption pad. Preliminary studies with the sweat patch indicate that it may be useful for detection of single and multiple drug use over a period of 1 to 4 weeks. Currently, its usefulness as a quantitative measure of drug use is being evaluated.

Hair testing appears to offer the possibility of monitoring drug use over an extended period of time that is dependent upon the length of an individual's hair. Drugs are sequestered in hair and remain bound for an extensive period of time. Because hair grows at an average rate of 1.0 to 1.5 cm per month, analysis of segments of hair for drug content can reveal historical drug use dating back months to years. Recent prevalence studies have indicated that substantially higher drug use rates are generally revealed by hair analysis than by urinalysis or self-report.

How each of the new drug-detection technologies will be used in the future for measuring drug prevalence is uncertain; however, it is clear that even greater reliance will be placed on chemical testing as a means of validating self-report. The technological base and general

understanding of the usefulness of urine, saliva, sweat, and hair as specimens for drug detection are certain to evolve at an even greater rate. The use of different biological specimens offers uniquely different information regarding the extent, frequency, and impact of drug use in selected populations.

REFERENCES

Ambre, J. The urinary excretion of cocaine and metabolites in humans: A kinetic analysis of published data. *J Forensic Sci* 9:241-245, 1985.

Batki, S.L.; Manfredi, L.B.; Jacob, P.; and Jones, R.T. Fluoxetine for cocaine dependence in methadone maintenance: Quantitative plasma and urine cocaine/benzoylecgonine concentrations. *J Clin Psychopharmacol* 12:243-250, 1993.

Baumgartner, W.A.; Black, C.T.; Jones, P.F.; and Blahd, W.H. Radioimmunoassay of cocaine in hair: Concise communication. *J Nucl Med* 23:790-792, 1982.

Bishop, Y.M.M.; Fienberg, S.E.; and Holland, P.W. *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, MA: MIT Press, 1975.

Caddy, B. Saliva as a specimen for drug analysis. In: Baselt, R.C., ed. *Advances in Analytical Toxicology.* Vol. 1. Foster City, CA: Biomedical Publications, 1984. pp. 198-254.

Cone, E.J. Saliva testing for drugs of abuse. In: Malamud, D., and Tabak, L., eds. *Saliva as a Diagnostic Fluid.* Vol. 694. New York: The New York Academy of Sciences, 1993. pp. 91-127.

Cone, E.J.; Yousefnejad, D.; Darwin, W.D.; and Maguire, T. Testing human hair for drugs of abuse. II. Identification of unique cocaine metabolites in hair of drug abusers and evaluation of decontamination procedures. *J Anal Toxicol* 15:250-255, 1991.

Cone, E.J.; Hillsgrove, M.J.; Jenkins, A.J.; Keenan, R.M.; and Darwin, W.D. Sweat testing for heroin, cocaine and metabolites. *J Anal Toxicol* 18:298-305, 1994.

Cone, E.J.; Menchen, S.L.; and Mitchell, J. Validity testing of the TDx® cocaine metabolite assay with human specimens obtained after intravenous cocaine administration. *Forensic Sci Int* 37:265-275, 1988.

Feucht, T.E.; Stephens, R.C.; and Walker, M.L. Drug use among juvenile arrestees: A comparison of self-report, urinalysis and hair assay. *J Drug Issues* 24(1):99-116, 1994.

Goldberger, B.A.; Caplan, Y.H.; Maguire, T.; and Cone, E.J. Testing human hair for drugs of abuse III. Identification of heroin and 6-acetylmorphine as indicators of heroin use. *J Anal Toxicol* 15:226-231, 1991.

Goldblum, R.W.; Goldbaum, L.R.; and Piper, W.N. Barbiturate concentrations in the skin and hair of guinea pigs. *J Invest Dermatol* 22:121-128, 1954.

Gorodetzky, C.W. Detection of drugs of abuse in biological fluids. In: Born, G.V.R.; Eichler, O.; Farah, A.; Herken, H.; and Welch, A.D., eds. *Handbook of Experimental Pharmacology.* Berlin: Springer-Verlag, 1977. pp. 319-323.

Henderson, G.L.; Harkey, M.R.; and Jones, R. "Hair Analysis for Drugs of Abuse." Final report of project activities under National Institute on Justice/National Institute on Drug Abuse grant no. NIJ 90-NIJ-CX-0012, 1993.

Magura, S.; Freeman, R.C.; Siddiqi, Q; and Lipton, D.S. The validity of hair analysis for detecting cocaine and heroin use among addicts. *Int J Addict* 27(1):51-69, 1992.

Magura, S.; Goldsmith, D.; Casriel, C.; Goldstein P.J.; and Lipton, D.S. The validity of methadone clients' self-reported drug use. *Int J Addict* 22(8):727-749, 1987.

Mieczkowski, T.; Barzelay, D.; Gropper, B.; and Wish, E. Concordance of three measures of cocaine use in an arrestee population: Hair, urine, and self-report. *J Psychoact Drugs* 23(3):241-249, 1991.

Robinson, S., and Robinson, A.H. Chemical composition of sweat. *Ann Rev Physiol* 34:202-20, 1954.

Saitoh, M.; Uzuka, M.; and Sakamoto, M. Rate of hair growth. In: Montagna, W., and Dobson, R.L., eds. *Advances in Biology of Skin.* Vol. IX. Oxford: Pergamon Press, 1969. pp. 183-201.

Schramm, W.; Smith, R.H.; Craig, P.A.; and Kidwell, D.A. Drugs of abuse in saliva: A review. *J Forensic Sci* 16:1-9, 1992.

Wish, R.D. U.S. Drug policy in the 1990's: Insights from new data from arrestees. *Int J Addict* 2(3A):377-409, 1990-1991.

AUTHOR

Edward J. Cone, Ph. D.
Chief
Chemistry and Drug Metabolism Section
Addiction Research Center
National Institute on Drug Abuse
P.O. Box 5180
Baltimore, MD 21224

167345

Comparison of Self-Reported Drug Use With Quantitative and Qualitative Urinalysis for Assessment of Drug Use in Treatment Studies

Kenzie L. Preston, Kenneth Silverman, Charles R. Schuster, and Edward J. Cone

ABSTRACT

The effectiveness of substance abuse treatment programs can be monitored by self-reported drug use and objectively measured by qualitative and quantitative urinalysis. The advantages and disadvantages of each of these three methods of assessing drug use are reviewed. Data collected in a clinical trial of a behavioral treatment for cocaine abuse are used to evaluate the relationships among qualitative and quantitative urinalysis for cocaine metabolite and self-reported cocaine use. Qualitative and quantitative urine testing showed greater rates of drug use than that shown by self-report, though there were significant correlations between self-reported use and urine toxicology results. Benzoylecgonine concentrations in urine specimens supported the suggestions that rates of drug use as determined by qualitative urinalysis are artifically high due to carryover and were informative about subjects' patterns of use.

INTRODUCTION

In clinical trials evaluating new treatments for abuse of drugs such as cocaine, an important outcome measure is the amount and frequency of illicit drug use. Unfortunately, the incidence and frequency of drug use are difficult to measure accurately. Drug use has been monitored by self-report in interviews and objectively by urinalysis. Although some clinical trials (Gawin and Kleber 1984) have relied principally on self-reported drug use and/or craving to assess outcome, most trials have used a combination of self-report and urine toxicology to monitor drug use (Weddington et al. 1991). Urine specimens usually are analyzed by

immunoassay or thin layer chromatography, and the result is reported in the qualitative mode (positive/negative). More recently, interest has grown in using quantitative testing to assess treatment outcome (Batki et al. 1993). Quantitative urinalysis has the potential to provide information regarding the amount and frequency of use (such as is gathered with self-report) while retaining the objectivity of drug testing.

The purpose of the chapter is to discuss the advantages and disadvantages of self-reported drug use and qualitative and quantitative urinalysis. Data from an ongoing clinical trial are used to evaluate the relationships among these three measures of drug use.

SELF-REPORT AS AN OUTCOME MEASURE—ADVANTAGES AND DISADVANTAGES

Self-reported drug use is usually reported in amount of drug (for example, in grams) or in amount of money spent on drugs. This information can be collected easily and nonintrusively and can cover a wide range of time periods (for example, the past 24 hours, the past week, or the past month). A significant drawback to relying upon self-reported drug use as an outcome measure in clinical trials is that the validity of the reports is questionable (Skog 1992). Self-reported drug use may not accurately reflect drug use for a number of reasons. Responses to questionnaires can be inaccurate because subjects do not know how much drug they have used, cannot remember, or are intentionally untruthful. Information about amounts of drug used (such as grams) is problematic because subjects may be poor judges of weights. In addition, the purity of drug purchased on the street is unknown and changes frequently. Collecting data in the form of dollar value has similar pitfalls, and, in addition, drug prices change over time and differ among localities.

Another frequently encountered problem is that drugs are often obtained as gifts or in exchange for goods or services, and subjects may not include drugs obtained in these ways in their reports. Recollection of amounts of drug used may be impaired by the duration of time since the use occurred (for example, when subjects are asked to estimate use in the past month) and by concurrent use of other psychoactive drugs (such as benzodiazepines) that have effects on memory. Subjects may intentionally inflate or underreport drug use, particularly if there is a real or perceived consequence to what is reported (Magura et al. 1987; Sherman and Bigelow 1992). Therefore, interviews or questionnaires must be

carefully worded, and the circumstances of their collection must be considered in order to get reports that are as accurate as possible.

QUALITATIVE AND QUANTITATIVE URINALYSIS AS OUTCOME MEASURES—ADVANTAGES AND DISADVANTAGES

Urinalysis has grown in importance as an outcome variable in substance abuse treatment research. Urinalysis is an objective measure that is independent of problems of subject memory or veracity. Typically, urine specimens are collected on a scheduled or random basis (usually one to three times per week) and analyzed in a qualitative mode for the presence of drug or metabolite at or above designated cutoff concentrations. Test results are usually expressed as positive or negative. The cutoff concentrations can vary from test to test, but standard values have been set by the Department of Health and Human Services (DHHS) for workplace testing (DHHS 1994). The following DHHS screening cutoff values are commonly used in clinical trials: cocaine/cocaine metabolite, 300 nanograms/milliliter (ng/mL); opiates, 300 ng/mL; amphetamines, 1000 ng/mL; marijuana, 50 ng/mL; and phencyclidine, 25 ng/mL. Such standardization is extremely useful when results from separate studies are compared or when data from multiple small studies are combined to increase statistical power in meta-analyses (Levin and Lehman 1991).

While having the advantage of objectivity, urinalysis also has some limitations. Unlike self-reported drug use, a drug must be present in the body in order for it to be detected; therefore, there is a relatively narrow window of time during which drug use can be detected by urinalysis. The duration of this time window is dependent on a number of factors, including the drug itself (e.g., biological half-life), dose, time of administration, amount of fluid consumed, individual differences in metabolism and excretion, and characteristics of the assay (for review see Cone and Dickerson 1992). Infrequent specimen collection can result in underrepresentation of drug use regardless of the analytic method used, though lowering cutoff concentrations can lengthen detection time. In contrast, frequent specimen collection can result in an overrepresentation of drug use. The drug or its metabolite may be detected in more than one urine specimen if the second specimen is collected before all drug or metabolite has been excreted. These multiple positives from a single use (referred to as carryover positives) artificially inflate the apparent rate of drug use by patients. Rates of carryover vary,

depending upon the same factors that affect the window of detection listed above. The impact of sample collection frequency has been reviewed elsewhere (Jain 1992).

Clinical evidence suggests that qualitative urine tests may have the significant disadvantage of being insensitive to moderate changes in drug use. For example, some clinical trials of cocaine treatments (Covi et al. 1994; Kolar et al. 1992) have found significant decreases in self-reported cocaine use without concomitant significant decreases in rates of cocaine-positive urine samples. Discrepancies between self-report and qualitative urinalysis can be partially explained by numerous factors. Moderate decreases in frequency of use may not be detected if urine tests remain positive between uses due to carryover. Decreases in amount of drug per use without changes in frequency of use may similarly not be detected by qualitative tests if the amount of drug use is high enough to produce urine concentrations above the cutoff. Although the clinical significance of decreases in drug use without complete abstinence is not clear, the identification of treatments that diminish cocaine use is important, particularly because no effective treatment agent is currently known.

As noted, there is a growing interest in the use of quantitative urine testing in clinical trials. Changes in the pattern, frequency, and amount of use that are not apparent from qualitative urinalysis might be discernible from quantitative urinalysis. On the other hand, quantitative urine testing is also somewhat more expensive than qualitative testing, and urine drug/metabolite concentration can be affected by such variables as the time between drug use and urine collection, fluid intake, and interindividual metabolic differences. For example, a urine specimen collected several days after self-administration of a large amount of drug could have the same drug/metabolite concentration as a specimen collected just after self-administration of a small amount of drug. Thus, the time of specimen collection could have greater impact on concentration than the total amount of drug used. Fluid intake can also affect urine drug/metabolite concentration, though corrections can be made using a biological indicator such as creatinine to adjust for water consumption.

To date only a few clinical trials have been conducted with quantitative testing. At least one study suggests that quantitative testing may be more sensitive to decreases in drug use than qualitative tests. Batki and colleagues (1993) found that fluoxetine significantly decreased cocaine use in a group of methadone maintenance patients as determined by

self-report and by quantitative analysis of urine cocaine and cocaine metabolite concentrations corrected by creatinine concentration; however, no significant effect of fluoxetine was shown when qualitative urine toxicology data were analyzed. McCarthy (1994) has also reported on the utility of quantitative urine drug testing in the context of substance abuse treatment. At this time, however, it is unclear whether the added cost of quantitative testing in clinical trials is justified; further comparison of the uses of quantitative and qualitative urine drug monitoring is needed.

COMPARISON OF SELF-REPORTED DRUG USE AND QUALITATIVE URINE TESTING IN A CLINICAL TRIAL

To evaluate the relationship between self-reported drug use and qualitative and quantitative urine testing, relevant data from a clinical trial of a behavioral treatment for cocaine abuse in methadone maintenance patients were analyzed. The study consisted of a randomized controlled trial comparing a voucher-based reinforcement contingency for cocaine abstinence to noncontingent voucher presentation in the context of an otherwise standard methadone maintenance program (Silverman et al. 1995). Under the reinforcement contingency, subjects received a voucher for each cocaine-free urine; the vouchers had monetary values that increased with the number of consecutive cocaine-free urines. In contrast, subjects in the control condition received vouchers in the same value, frequency, and pattern of presentation as the experimental group, but independent of their urine screen results. The vouchers could be exchanged for goods and services that were consistent with a drug-free lifestyle and patients' treatment goals.

The study was 17 weeks long, with a 5-week baseline phase in which subjects' drug use was monitored and a 12-week voucher phase in which the treatment intervention was in place. Participants were 37 patients who used cocaine consistently during the first 5 weeks of methadone maintenance treatment. Subjects visited the clinic 7 days per week to receive methadone (50 mg orally) for up to 17 weeks. Three days per week (Monday, Wednesday, and Friday) they also answered self-report questionnaires and submitted urine samples. Three days per week subjects were asked whether they had used any cocaine, and, if so, how much (in grams) in the last 24 hours. If the subject reported the use in dollars spent, the information was converted to grams using a conversion

factor of $10 per 0.1 gram of cocaine. This information was entered into a database as a dichotomous variable (yes/no) and as amount (grams).

All urine collections were observed by trained laboratory technicians. At the time of collection, a portion of each specimen was frozen for later quantitative analysis. The rest of the sample was refrigerated and sent to a commercial laboratory for qualitative testing on the day of collection. Testing was conducted with an enzyme multiplied immunoassay technique (EMIT) system that gave qualitative results for the presence of cocaine metabolite (cutoff concentration 300 ng/mL, benzoylecgonine equivalents). The EMIT assay primarily detects benzoylecgonine, the principal metabolite of cocaine. Results of the qualitative urine toxicology screens were available to the subjects and to the counselors for use in their treatment plans and counseling sessions with subjects. Primary outcome measures for the original study were cocaine abstinence in each study week and the longest duration of sustained cocaine abstinence as determined by qualitative urinalysis.

Quantitative testing of cocaine metabolite was conducted with an analyzer and cocaine metabolite reagents according to the manufacturer's recommended procedure. Results are expressed as benzoylecgonine equivalents (ng/mL). The sensitivity of the assay for benzoylecgonine as reported by the manufacturer was 30 ng/mL. The assay has been shown to be highly specific and accurate for the measurement of benzoylecgonine in urine. Cone and colleagues (1988) showed that results from the assay were highly correlated with benzoylecgonine concentrations determined by gas chromatography/mass spectrometry (GC/MS) for urine specimens collected from subjects who had received cocaine in a laboratory study.

Mean self-reported drug use in the past 24 hours (yes/no), grams of cocaine used in the past 24 hours, cocaine-positive urine specimens (qualitative assay), and benzoylecgonine equivalents concentrations were calculated across time for the 37 subjects participating in the 17-week trial. Means and standard deviations across subjects are listed in table 1. On average, subjects reported use of cocaine on 29 percent of occasions but tested positive for cocaine (qualitatively) on 68.2 percent of occasions. The concentration of benzoylecgonine equivalents varied widely, both across and within subjects, ranging from less than 30 ng/mL to more than 900,000 ng/mL. Overall, the mean benzoylecgonine were equivalent was 32,368 ± 29,254 ng/mL. Within-subject correlations between self-reported use (percent of reports positive for use) and urinalysis data (percent positive in qualitative tests or mean metabolite

TABLE 1. *Means, standard deviations, and correlation coefficients for three measures of cocaine use.*

Variable	Mean	Standard deviation	Correlation coefficient to self-reported cocaine use
Self-reported use (% yes)	29.04	25.39	--
% cocaine positive*	68.20	28.40	0.6934
Benzoylecgonine equivalents (ng/mL)	32,368	29,254	0.7975

KEY: * = Specimens were tested by EMIT for cocaine metabolite with a 300 ng/mL cutoff concentration for positive results.

concentration) were in the high range: R = 0.693 for qualitative results and R = 0.798 for benzoylecgonine equivalents. These data suggest that there was general correspondence between self-report and urinalysis results within subjects, such that subjects who reported more cocaine use also tested positive for cocaine more frequently and had higher benzoylecgonine concentrations.

To evaluate the correspondence between the cocaine use measures at individual data-collection points, data from the 37 study participants were combined, and 1,678 sets of concomitantly collected urine specimens and self-reports were examined (table 2). Overall, 1,124 (67 percent) of the specimens tested positive (yes/no) for cocaine, and 470 (28 percent) of the self-reports were positive for cocaine use. Chi-square analysis comparing cocaine-positive urine specimens and self-reports of cocaine use was highly significant ($p < 0.001$). When self-report was positive for cocaine use, correspondence between self-report and positive results by urinalysis was quite high: Of 470 occasions of self-reported use, 463 (98.5 percent) were also positive by qualitative urinalysis. In contrast, there was a lack of correspondence when qualitative urinalysis results were positive: Subjects reported using cocaine on only 41 percent of the 1,124 occasions that urine tested positive for cocaine. There was agreement between urinalysis and self-report (both positive or both negative for cocaine use) on 60.19 percent of occasions. A Kappa value of 0.307, in the moderate range, was computed from these data. Kappa (Cohen 1960) assesses the degree

TABLE 2. *Relationship between qualitative urinalysis and self-reported drug use in data analyzed as individual occasions.*

	Self-reported cocaine use		
Urinalysis*	No	Yes	Total
Negative	547	7	554 (33%)
Positive	661	463	1,124 (67%)
Total	1,208 (72%)	470 (28%)	1,678 (100%)

KEY: * = Specimens were tested by EMIT for cocaine metabolite with a 300 ng/mL cutoff concentration for positive results.

of validity between the self-reports of drug use and urinalysis beyond that expected by chance alone. Thus, self-report of cocaine use predicted a positive result on qualitative urinalysis, but positive urinalysis was not predictive of self-report because subjects reported using cocaine on only about half of these occasions.

CAN QUANTITATIVE URINALYSIS RESOLVE THE DISCREPANCY BETWEEN SELF-REPORT AND QUALITATIVE URINALYSIS?

Close inspection of individual data suggests that benzoylecgonine concentration (as determined by quantitative urinalysis) does provide a basis for examining the relationship between self-reported drug use and qualitative urinalysis. Data for the three measures of cocaine use (self-report, quantitative urinalysis results, and benzoylecgonine concentrations) of two representative subjects are shown in figures 1 and 2. Benzoylecgonine concentrations are indicated by open circles graphed on a log scale. Urine specimens were collected and analyzed three times per week over a period of 17 weeks for a total of 51 occasions; sequential urine specimens numbers 1 through 15 were collected during baseline; urine specimens numbers 16 through 51 were collected during the experimental treatment phase. The cutoff for the quantitative testing (300 ng/mL) is indicated by the horizontal dashed line. The subject in figure 1 showed a cyclical pattern of drug use (based on benzoylecgonine

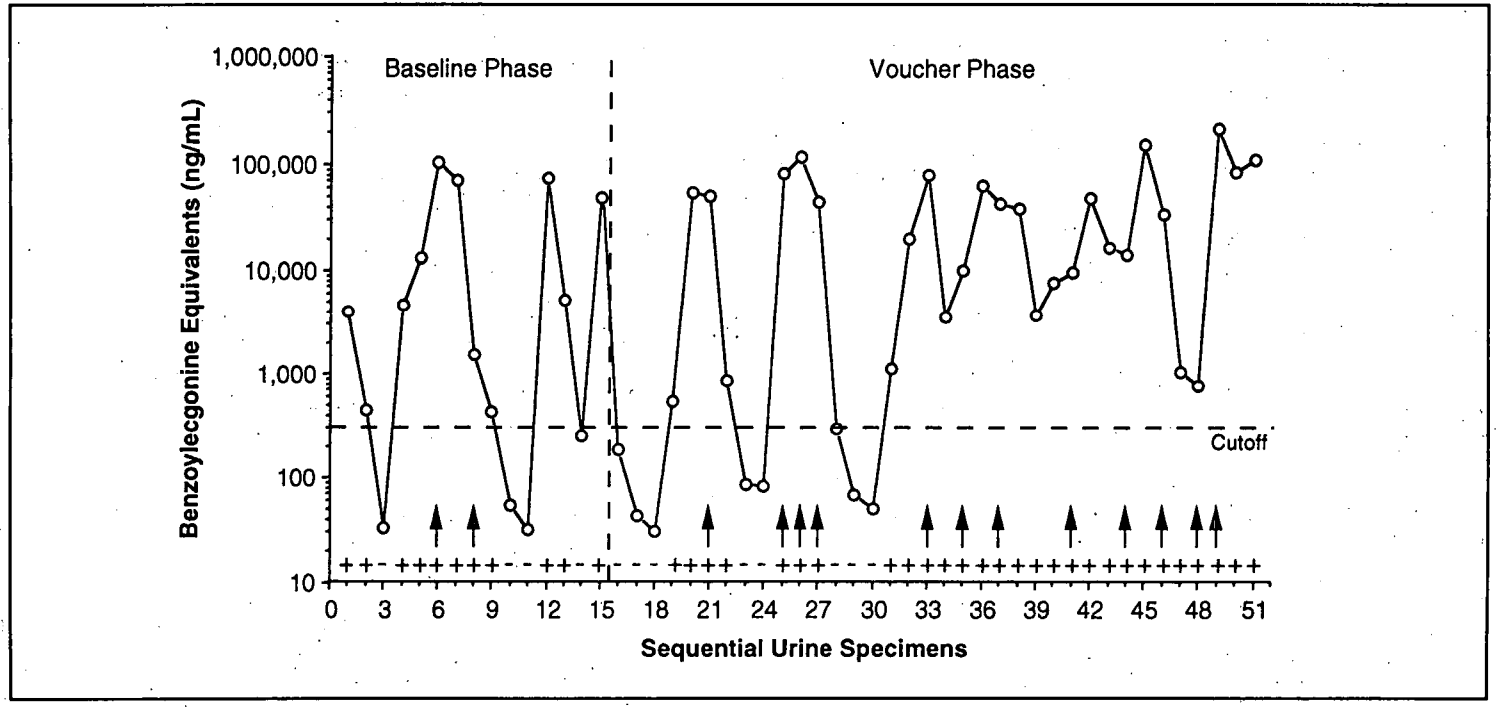


FIGURE 1. *Urinalysis results and self-reported cocaine use across time from a subject in the control group. Plus signs (+) above the X axis indicate urine samples testing positive for cocaine metabolites at concentrations of 300 n/mL or greater, and minus signs (-) indicate negative urine samples. Arrows indicate days on which the subject reported using cocaine within the previous 24 hours. Horizontal dashed line indicates 300 ng/mL; vertical dashed line indicates end of baseline phase.*

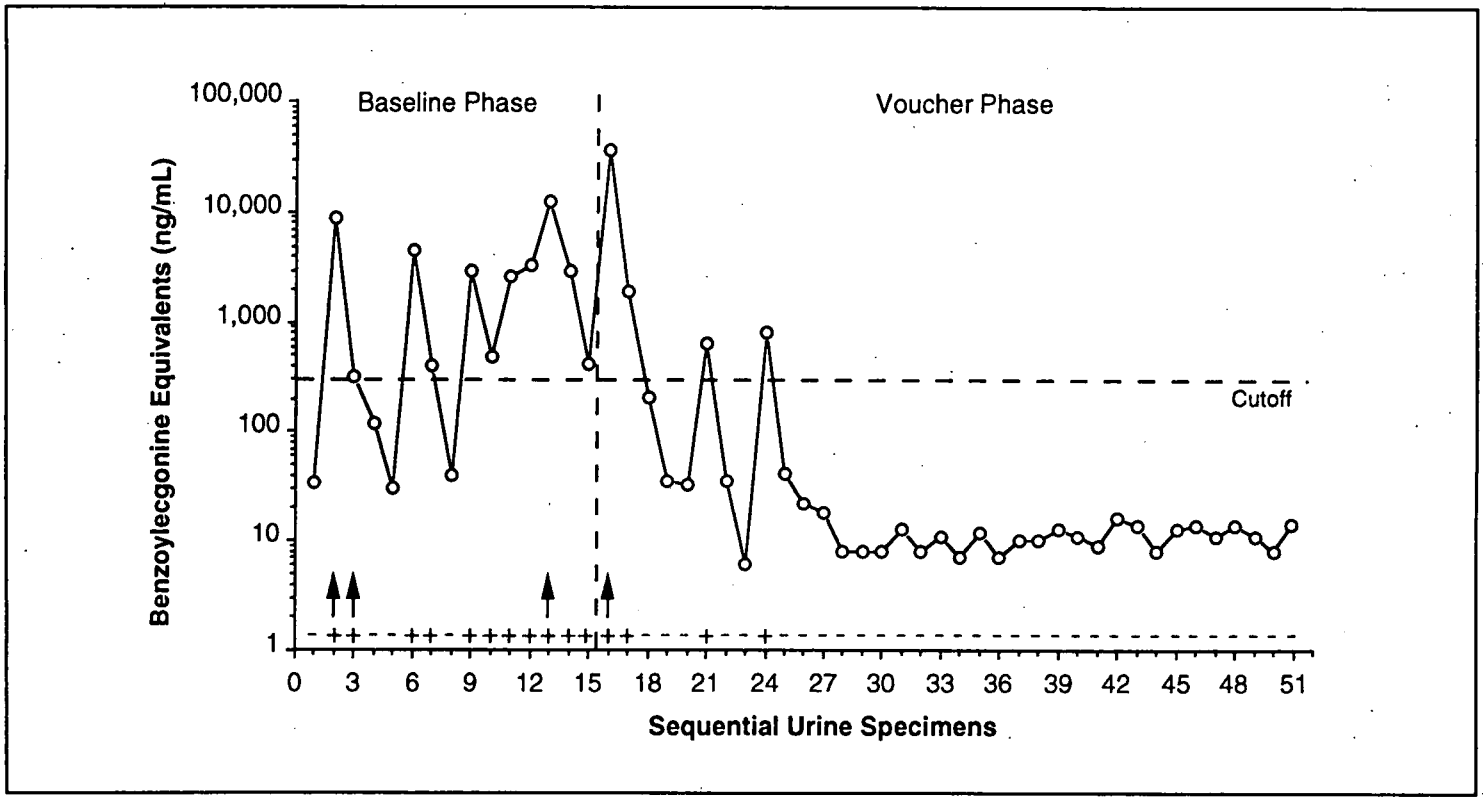


FIGURE 2. *Urinalysis results and self-reported cocaine use across time from a subject in the voucher-based reinforcement group. Details are the same as described for figure 1.*

concentrations), with episodes of varying length separated by periods of no use. The results for this subject are typical of other subjects in the control group, which showed no significant decrease in cocaine use during the voucher phase of the study. Benzoylecgonine equivalent concentrations varied over a wide range from 31 to 216,500 ng/mL.

In contrast, the participant whose results are illustrated in figure 2 had a cyclical pattern of drug use early in treatment, followed by sustained cocaine abstinence. During the first 5 weeks of methadone maintenance (baseline), benzoylecgonine equivalent concentrations varied from approximately 30 to 36,000 ng/mL. This subject decreased cocaine use with the initiation of the experimental treatment in the sixth week of treatment, and after two short relapses, stopped using cocaine completely; benzoylecgonine equivalent concentrations decreased to less than 30 ng/mL, the limit of detection of the assay.

Qualitative urinalysis results and self-reported cocaine use are also indicated in figures 1 and 2. Results of qualitative urinalysis are shown as plus signs (+) indicating urine samples testing positive for cocaine metabolites at concentrations of 300 ng/mL or greater; minus signs (-) indicate negative urine screens. Arrows indicate days on which the subject reported using cocaine within the previous 24 hours. Clearly, there is a lack of concordance between self-reported uses and cocaine-positive urine specimens for both subjects. In figure 1, 39 of 51 samples (76.5 percent) were above the 300 ng/mL cutoff, while the subject reported cocaine use within the previous 24 hours on only 14 occasions. Self-reports of use tended to coincide with the longer periods of cocaine-positive urine specimen, and multiple self-reported uses were associated with longer periods during which consecutive urine specimens were above the 300 ng/mL positive/negative cutoff. Early in treatment, the subject infrequently reported using cocaine in the previous 24 hours and had numerous negative qualitative urinalysis results. Beginning with the 33rd sequential urine specimen, the subject began reporting use more frequently, and qualitative urinalysis were continuously positive. Quantitative urinalysis, however, suggests a continuing cyclical pattern of use, even though the urine benzoylecgonine concentration never decreased to below the 300 ng/mL cutoff.

For the subject whose data are illustrated in figure 2, 15 (29.4 percent) out of 51 urine specimens tested above the 300 ng/mL cutoff; all of the positive urine specimens occurred during the first half of treatment. The subject reported using cocaine in the past 24 hours on four occasions; on

each occasion the subject also tested positive on the qualitative urinalysis. As with the subject described in figure 1, quantitative urinalysis suggests a continuing cyclical pattern of use, even during the period of sustained cocaine-positive urinalysis results from sequential urine numbers 9 through 17. Thus, quantitative urinalysis results provided additional information on patterns of drug use and documented the subject's response to treatment.

As described above, one of the potential reasons for discrepancies between self-reported drug use and qualitative urinalysis in clinical trials with frequent urine specimen collections is from carryover positives. Benzoylecgonine can usually be detected in the urine for about 48 hours after cocaine administration (Saxon et al. 1988), though even longer detection times are possible depending on the amount of cocaine taken and individual rates of excretion. Benzoylecgonine concentration data may provide a basis for evaluating the discrepancy between self-report and qualitative urinalysis and the impact of carryover. As noted above, self-reported cocaine use occurred at a much lower rate than cocaine-positive urine specimens for the study as a whole: 470 (28 percent) versus 1,124 (67 percent) out of 1,678 occasions (table 2). A similar pattern was seen in the data of the individual subjects illustrated in figures 1 and 2. Examination of the quantitative data supports the suggestion that at least part of the differential rates of self-report and qualitative cocaine-positive urine specimens was due to carryover. In figure 1, for example, benzoylecgonine concentration dropped substantially between sequential urine specimens numbers 21 and 22, but remained above 300 ng/mL. Possible carryover positives are also seen in figure 2 for sequential urine specimens numbers 7, 15, and 17. Further research may lead to a more systematic approach to estimating rates of cocaine use from urine benzoylecgonine concentrations.

SUMMARY

The effectiveness of substance abuse treatment programs can be monitored by self-reported drug use and objectively measured by urinalysis. Self-reported drug use is usually reported as amount of drug (for example, in grams) or amount of money spent on drugs. While this information can be collected easily and nonintrusively, the validity of the self-reported drug use is often questionable, particularly if there is a real or perceived consequence to what is reported. Therefore, urinalysis is a critical variable in treatment research. Typically, urine specimens are collected on a

scheduled or random basis and analyzed in a qualitative mode for the presence of drug or metabolite at or above a designated cutoff concentration, with testing results usually expressed as positive or negative. Qualitative urine testing may be insensitive to decreases in drug use because of carryover positives (more than one drug-positive test from a single use). Rates of carryover vary depending upon a number of factors including dose, time of drug administration, individual factors such as rates of metabolism and excretion, water consumption, and characteristics of the assay. Urine samples can also be tested with quantitative measures to determine urine drug/metabolite concentrations. Quantitative testing may be more sensitive to decreases in drug use, though many of the factors affecting qualitative tests also affect quantitative testing.

The relationships among qualitative and quantitative urinalysis for cocaine metabolite and self-reported drug use were assessed with data collected in a clinical trial of a voucher-based reinforcement contingency treatment intervention. There was significant correlation between self-reported use and urine toxicology results, although qualitative and quantitative urine testing showed greater rates of drug use than that shown by self-report. Benzoylecgonine concentrations in urine specimens were informative about subjects' patterns of use and the relationship between patterns of self-report and qualitative urinalysis. Benzoylecgonine concentrations also supported the suggestion that rates of drug use as determined by qualitative urinalysis are artificially high due to carryover. Quantitative urinalysis may be a useful measure of drug use in clinical trials of cocaine abuse treatments.

The value of quantitative testing in the context of community substance abuse treatment is unclear. In general, community treatment programs conduct relatively infrequent urine testing. Concentrations of drugs in urine specimens collected at intervals that are too long cannot give information about patterns of use. They may also not be particularly useful indicators of amount of drug use because urine concentrations can fluctuate dramatically even over relatively short periods of time (e.g., 48 hours, as in the current study). The problem of carryover positives is much less likely under current treatment practices when specimens are collected at wide intervals. In addition, the costs of testing may be prohibitive. However, in those settings where urine testing is frequent (for example, some programs associated with the justice system), quantitative testing could decrease the number of occasions when negative consequences are applied to individuals who test positive more than once because of carryover. If future research

demonstrates that rates and patterns of drug use are helpful for predicting treatment outcome or for identifying appropriate treatments for individual patients, increased funding and changes in standards of care that would permit frequent quantitative urinalysis might be justified.

REFERENCES

Batki, S.L.; Manfredi, L.B.; Jacob, P.; and Jones, R.T. Fluoxetine for cocaine dependence in methadone maintenance: Quantitative plasma and urine cocaine/benzoylecgonine concentrations. *J Clin Psychopharmacol* 13:243-250, 1993.

Cohen, J. A coefficient of agreement for nominal scales. *Educ Psychol Measure* 20:37-46, 1960.

Cone, E.J., and Dickerson, S.L. Efficacy of urinalysis in monitoring heroin and cocaine abuse patterns: Implications in clinical trials for treatment of drug dependence. In: Jain, R.B., ed. *Statistical Issues in Clinical Trials for Treatment of Opiate Dependence*. National Institute on Drug Abuse Research Monograph 128. DHHS Pub. No. (ADM)92-1947. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 46-58.

Cone, E.J.; Menchen, S.L.; and Mitchell, J. Validity testing of the TDx® cocaine metabolite assay with human specimens obtained after intravenous cocaine administration. *Forensic Sci Int* 37:265-275, 1988.

Covi, L.; Hess, J.M.; Kreiter, N.A.; and Haertzen, C.A. Three models for the analysis of a fluoxetine placebo controlled treatment in cocaine abuse. In: Harris, L.S., ed. *Problems of Drug Dependence, 1993.* National Institute on Drug Abuse Research Monograph 141. DHHS Pub. No. (ADM)94-3749. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1994. p. 138.

Department of Health and Human Services. "Mandatory Guidelines for Federal Workplace Drug Testing Programs." *Federal Register*, June 9, 1994.

Gawin, F.H., and Kleber, H.D. Cocaine abuse treatment. Open pilot trial with desipramine and lithium carbonate. *Arch Gen Psychiatry* 41:903-909, 1984.

Jain, R.B. Design of clinical trials for treatment of opiate dependence: What is missing? In: Jain, R.B., ed. *Statistical Issues in Clinical Trials for Treatment of Opiate Dependence*. National Institute on Drug Abuse Research Monograph 128. DHHS Pub. No. (ADM)92-1947. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 46-58.

Kolar, A.F.; Brown, B.S.; Weddington, W.W.; Haertzen, C.C.; Michaelson, B.S.; and Jaffe, J.H. Treatment of cocaine dependence in methadone maintenance clients: A pilot study comparing the efficacy of desipramine and amantadine. *Int J Addict* 27:849-868, 1992.

Levin, F.R., and Lehman, A.F. Meta-analysis of desipramine as an adjunct in the treatment of cocaine addiction. *J Clin Psychopharmacol* 11:374-378, 1991.

Magura, S.; Goldsmith, D.; Casriel, C.; Goldstein, P.J.; and Lipton, D.S. The validity of methadone clients' self-reported drug use. *Int J Addict* 22:727-749, 1987.

McCarthy, J. Quantitative urine drug monitoring in methadone programs: Potential clinical uses. *J Psychoactive Drugs* 26:199-206, 1994.

Saxon, A.J.; Calsyn, D.A.; Haver, V.M.; and Delaney, C.J. Clinical evaluation of urine screening for drug abuse. *West J Med* 149:296-303, 1988.

Sherman, M.F., and Bigelow, G.E. Validity of patients' self-reported drug use as a function of treatment status. *Drug Alcohol Depend* 30:1-11, 1992.

Silverman, K.; Higgins, S.T.; Brooner, R.K.; Montoya, I.D.; Schuster, C.R.; and Preston, K.L. Differential reinforcement of sustained cocaine abstinence in intravenous polydrug abusers. In: Harris, L.S., ed. *Problems of Drug Dependence, 1994*. National Institute on Drug Abuse Research Monograph 153. DHHS Pub. No. (NIH) 95-3883. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1995.

Skog, O.-J. The validity of self-reported drug use. *Br J Addict* 87:539-548, 1992.

Weddington, W.W.; Brown, B.S.; Haertzen, C.A.; Hess, J.M.; Mahaffey, J.R.; Kolar, A.F.; and Jaffe, J.H. Comparison of amantadine and desipramine combined with psychotherapy for treatment of cocaine dependence. *Am J Drug Alcohol Abuse* 17:137-152, 1991.

ACKNOWLEDGMENTS

This research was funded by the National Institute on Drug Abuse Intramural Research Program. Quantitative assays were performed by Christopher Sheppard and Rosalind Jones; data were analyzed by Chris Johnson and Nancy Kreiter.

AUTHORS

Kenzie L. Preston, Ph.D.
Chief, Clinical Trials Section

Edward J. Cone, Ph.D.
Chief, Chemistry and Drug Metabolism Section

Intramural Research Program
National Institute on Drug Abuse
Addiction Research Center
P.O. Box 5180
Baltimore, MD 21224

Kenneth Silverman, Ph.D.
Assistant Professor
Department of Psychiatry and Behavioral Sciences
Johns Hopkins University School of Medicine
5510 Nathan Shock Drive
Baltimore, MD 21224

Charles R. Schuster, Ph.D.
Professor
Department of Psychiatry and Behavioral Neurosciences
Director
Clinical Research Division on Substance Abuse
Wayne State University
2751 East Jefferson
Detroit, MI 48207

The Forensic Application of Testing Hair for Drugs of Abuse

Mark L. Miller, Brian Donnelly, and Roger M. Martz

ABSTRACT

Hair testing is only used by the Federal Bureau of Investigation (FBI) when other information exists that indicates drug use and can remove a person from suspicion or associate them with criminal activity. The detection of cocaine in hair has been the FBI's first priority in hair testing for drugs of abuse because of its prevalence. Several cases when hair testing was used are reported in this chapter. Further, analysis of over 100 samples was performed on hair obtained from a medical examiner's random autopsy collection. Sixty-five percent of the samples tested positive for cocaine or opiates. The results of hair testing for drugs of abuse were found to be consistent with autopsy toxicology reports. The analysis of hair washes and nails from the autopsy samples suggests external contamination of hair with drugs is not widespread.

INTRODUCTION

The forensic testing of hair for drugs of abuse is a recently acquired law enforcement tool that can be used to ascertain the truth about an individual's consumption of drugs. Lying to an FBI special agent about drug use (or any other matter) is illegal. Yet it can be anticipated that truthful information about self-admitted drug use is not frequently encountered by law enforcement. Alternative methods such as hair analysis are therefore needed to measure the past use of drugs.

One of the primary reasons for a person's lack of candor with law enforcement is the fear of criminal prosecution. People involved in criminal activity frequently conceal, distort, or falsify the truth. In fact, upon initial investigation, no suspect has confessed to the abuse of drugs

---

This chapter was previously published as Federal Bureau of Investigation Laboratory Division Publication No. 94-19.

in the cases the FBI Laboratory has dealt with in the testing of hair for drugs.

Reluctance to admit drug use to law enforcement personnel can occur for reasons other than incrimination. For example, even in instances when drug use has been surveyed with promises of anonymity and confidentiality among those arrested on criminal charges, it has been found through biological tests (to ascertain the accuracy of the responses) that there is a tendency to conceal or underreport the short- and long-term use of drugs (Mieczkowski and Newel 1993). One of the primary reasons for under-reporting may be to hide the extent of abuse. Moreover, the ability to accurately recollect and self-report may be impaired when the user has been under the influence of a mind-altering drug. Additionally, purchased street drugs are often of unknown purity and composition, and users may unintentionally give inaccurate reports.

It is difficult for drug abusers to accurately self-report which drugs and how much drug they have used when they are frequently consuming illicit substances that may have been obtained from unreliable sources. For example, a recent Drug Enforcement Administration (DEA) publication (DEA 1994) cited several instances of street drugs having a very different composition than their represented contents. In the first case, a small lump of a waxy black solid sold as tar heroin was found to be part of a black crayon. In another instance, a white powder purported to be cocaine was analyzed and found to be ephedrine and caffeine. A substance sold as crack was identified as a mixture of dextrose and paraffin wax. An alleged fentanyl sample was revealed to contain not only the suspected drug but also heroin and nicotinamide. As can be seen from these examples, drug abusers can be consuming very different drugs than intended, or, in extreme cases, no drug at all.

The development of drug-specific hair tests devised in the FBI Laboratory has been driven by the type of drug analysis requests received, which concurs with criminal justice survey data on the high prevalence drugs (i.e., cocaine). According to the 1992 National Institute of Justice annual report on Drug Use Forecasting (DUF), in 24 major U.S. cities, anywhere from 48 to 85 percent (depending on the location) of male or female booked arrestees tested positive for various drugs by urinalysis (Department of Justice 1993). Cocaine was found to be the most prevalent drug at 22 of the 24 test sites, and accounted for as much as 72 percent of the positive drug results in Manhattan for females. Marijuana was the leading drug at two of the sites and was the second most detected drug overall;

38 percent of male arrestees in Omaha tested positive. The third most frequently detected type of substance revealed by urinalysis results came from the opiate class of drugs. The highest percentage of opiate positives from the 24 locations was in Manhattan, with 24 percent of females testing positive. The arrestees in this study were booked on a variety of charges (mostly felony), not just drug offenses. These results serve to illustrate the link between crime and drug abuse.

Results of the DUF study suggest cocaine is the most commonly abused drug. For this reason it can be understood why the FBI Laboratory has established cocaine testing in hair as its first priority for this type of analysis. The detection of marijuana, the second most prevalent abused drug among arrestees according to the DUF study, has not been pursued in hair by the FBI Laboratory because of its low concentration in this tissue and the persistence of its metabolites in urine. Urinalysis permits detection of marijuana use up to several weeks after its consumption (Liu 1992; Cone, this volume). The FBI Laboratory is developing hair tests for opiates/heroin because of their prevalence and use in society as illustrated in the DUF study.

Hair testing has distinct advantages over other forms of toxicological sampling and analysis. For example, distinguishing heroin use from other opiates via blood or urine samples is more problematic than it is in hair testing because of the short half-life of heroin and its primary metabolite, 6-monoacetylmorphine (6-MAM), in these fluids. Heroin and 6-MAM are detectable in urine for only a few hours. Morphine and codeine are secondary metabolites of heroin and are more persistent in biological fluids than heroin or 6-MAM. In contrast, 6-MAM is the major marker of heroin use in hair. The differentiation of opiate use is important because morphine and codeine can be licitly consumed in foods such as poppy seeds or prescribed in medications such as cough syrups (ElSohly and Jones 1989; Liu 1992). Therefore, one of the largest incentives for the determination of heroin use from hair is the ability to differentiate its use from other opiates via the presence of its unique identifying metabolite.

Due to the rapid metabolism and elimination of most drugs and their metabolites, it is difficult to analyze and quantitate them in body fluids 2 days or more after use. In contrast, cocaine and heroin use can be detected in hair samples collected months after the drugs are consumed. Another advantage of hair testing is the noninvasive nature of sampling compared with the collection of blood or urine.

APPLICATION OF HAIR TESTING

Hair testing for drugs of abuse has enhanced the ability of law enforcement to corroborate the truthfulness of testimony on drug use. The historical information on drug consumption attainable from testing hair gives it a distinct advantage over urine drug testing because of the extended detection window. The data obtained from hair testing have had an impact in investigations on a wide variety of offenses. Hair analysis is only used by the FBI Laboratory when there is evidence that drug abuse has occurred and it has a bearing on a case. The results of hair testing can associate subjects with criminal offenses or remove a person from suspicion. Generally, hair testing for drugs is needed as a confirmation technique when there is a disputed positive urinalysis (for example, claims of sample mislabeling or of a single occurrence of drug use, false positives), allegations of criminal activity, parole violations, or a history of drug abuse.

Some cases that have used hair testing at the FBI Laboratory and involve drug-related offenses include a drug smuggler, military personnel, Government employees, law enforcement personnel, prison inmates, parolees, and public officials. A prominent mayor, an attorney, and a prosecutor are included on the hair analysis list of public officials who were suspected of drug abuse. Hair testing for drugs of abuse also has made a critical difference in the outcome of casework seemingly unrelated to the use of drugs, such as investigations of murder, rape, and product tampering.

The FBI Laboratory has processed approximately 76 requests for hair testing related to casework since the first analysis in 1987 for an investigation involving a cocaine smuggler (records are kept according to how many cases have requested hair testing). The number of case samples steadily rose from 1987 to 1992, when it peaked at 35 investigations involving hair testing (figure 1). The numbers have tapered off recently as some requests have been referred to other laboratories to prevent casework overload.

The court cases that have used FBI results of cocaine hair testing have been successful, beginning with the smuggler's trial in 1987. Nearly half the cases have been military personnel faced with courts martial over drug abuse. Convictions were obtained in all but one case. Most defendants have pleaded guilty when confronted with combined positive urinalysis

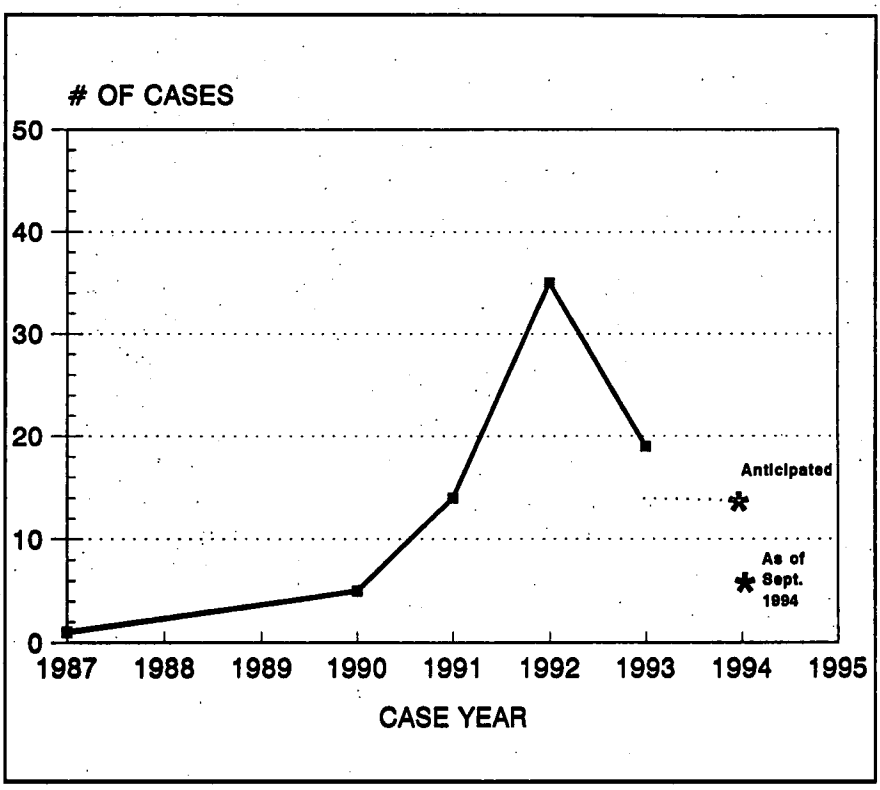


FIGURE 1. *Annual number of hair testing cases during 1987 to 1994 for the FBI Laboratory Chemistry-Toxicology Unit.*

and hair testing results. In cases of nonmilitary Federal employees, they have either been found negative and cleared or resigned their positions.

TESTING HAIR FOR DRUGS

The FBI Laboratory performs hair testing for cocaine. The testing of hair for drugs begins with the voluntary or court-ordered collection of approximately 100 hairs from the vertex of the contributor. To maintain sample integrity, the hair is transferred from the collection official to the laboratory through a documented chain of custody.

Hair is tested for cocaine and its major metabolite benzoylecgonine using mass spectrometry for the determination of cocaine abuse. Both compounds have been detected in the majority of cases. The anticipated hair test for heroin use focuses on the detection of its primary metabolite

6-MAM. The appearance of heroin and/or 6-MAM is a prerequisite for a declaration of heroin use determination via hair analysis. The presence of morphine and codeine are also examined, but a positive finding is not necessarily an indicator of heroin use.

Details of the procedure for analyzing drugs in hair can be found in the previous publications by the authors, but the method is briefly described here (Martz et al. 1991; Miller et al., in press). A 5 milligram (mg) hair sample is cleaned by washing it twice with solvent (methanol) to remove potential contaminating drugs on the hair surface. Baumgartner and colleagues (1993) established that solvent washing readily removes drugs on the surface of hair.[1] The drugs in the hair are extracted with acid (for cocaine analysis only) or solvent at above ambient temperature after internal standards are added to the solution. Deuterated analogs of the target drugs or related compounds are used as the internal standards for the purpose of quantitation. After extraction and sample preparation, the final concentrate is analyzed by tandem mass spectrometry (MS/MS) or electrospray ionization liquid chromatography/mass spectrometry (ESI LC/MS).

HAIR TESTING CASES

Examples of how testing hair for drugs can be used in a forensic environment are given below for illustrative purposes. One of the earlier high publicity cases involved the victim of a product tampering by international smugglers (Martz et al. 1991). This case fell under Federal jurisdiction as a consumer product-tampering offense. In July of 1990, a Miami man became extremely ill after drinking an imported malted beverage from Colombia. After drinking the contents of the bottle, the subject thought he may have been poisoned; he stated the beverage tasted bad, and his mouth and tongue were numb. The man went into a coma immediately after making the statement and was rushed to the hospital. At the hospital he was diagnosed as suffering from acute cocaine intoxication after a urinalysis test.

Cocaine was detected in the residue of the bottle consumed by the victim. The subject was maintained alive for 24 days until his life support system was shut off. A recall of the malt beverage found an average of 30 grams of cocaine per bottle of the tampered product.

After the victim died, hair samples were collected to determine whether he was a regular cocaine user who had overdosed or the victim of a product tampering (during the period after the incident but before his death, the victim's hair grew approximately 1 to 1.5 centimeters (cm) (Chatt and Katz 1988)). Historical information on his drug usage was gathered by conducting segmental analysis on the victim's 2.5 cm length hair. The hair was cut into half-centimeter segments and analyzed (figure 2). The hair segments contained a peak concentration of almost 100 nanograms (ng) per mg at a time period that corresponds to the ingestion of the suspect beverage (segment 1-1.5 cm). The high level of cocaine in the two segments at the tip of the hair (segments 1.5-2.5 cm) indicate the victim was a user of cocaine before the incident.[1] Witness interviews substantiated results of the segmental hair analysis during the investigation, which revealed the victim was a chronic cocaine user.

In the next example, a rape investigation was aided by hair analysis for cocaine. A request was made for hair analysis by a small town's police department to contest the alibi of a suspect after a woman reported an

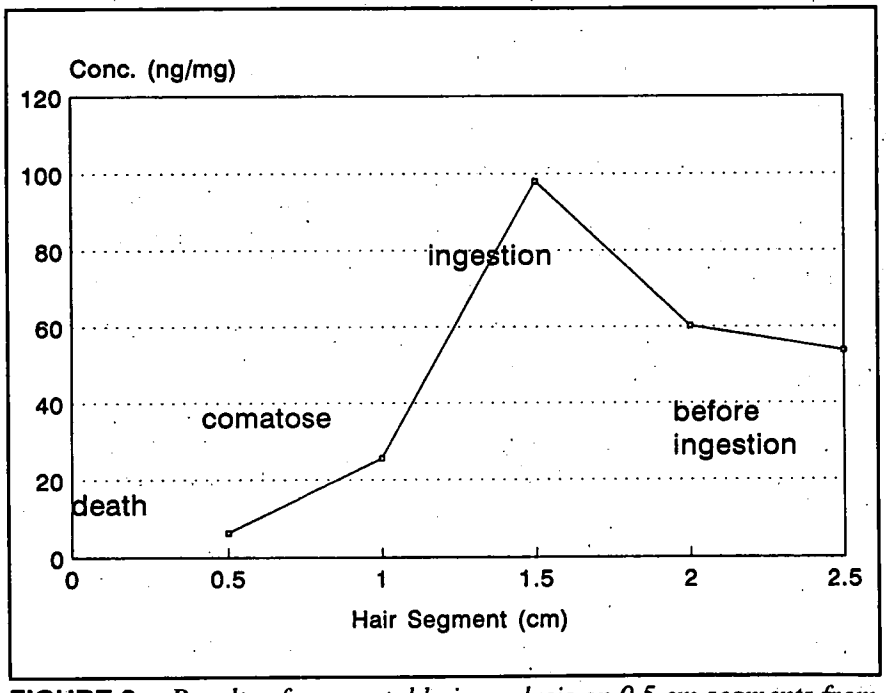


FIGURE 2. *Results of segmental hair analysis on 0.5 cm segments from a lethal intoxication of cocaine in imported beverage. The data points represent hair cocaine concentrations.*

acquaintance had raped her in her own home. The suspect stated he and the victim were dating, engaging in sex, and had used crack cocaine together on numerous occasions. She denied his allegations and proof was needed to refute or confirm his alibi. Since the suspect was positive for cocaine and the victim was negative for use of cocaine over the previous several months, hair testing was effective in contradicting the alibi.

The use of hair testing also has been effective in accidental or man-slaughter death investigations. In one case, a child died as a result of cocaine intoxication while in the care of his mother and her common-law husband. The mother indicated that her husband was a cocaine user and the husband implicated the mother as a drug user. Results of hair testing revealed the father was positive for cocaine while the mother was found to be negative. This implicated the father as a user and possible owner of the cocaine ingested by the child and resulted in his confession as being the possessor of the cocaine.

Members of the military are routinely tested for drug usage via urinalysis. Those found to be using drugs are court martialed and discharged from the service. In several instances, hair testing has been used to corroborate positive urine tests as well as other investigative information such as adulterated urine specimens. In one particular case, a military man near retirement whose urine and hair tested positive for cocaine was exonerated from court martial in spite of this evidence. He claimed his wife had spiked his food with cocaine. His wife, who was divorcing him, initially refused to corroborate his story, but later testified to spiking his food several times. Because he was considered a victim of tampering, the jury found him innocent.

RESEARCH ON DRUGS IN HAIR

A project at the FBI's Forensic Science Research Unit screened random hair samples collected from autopsies conducted by a medical examiner. These samples consisted mainly of homicide, suicide, and accident victims. A small proportion of the people autopsied died of medical illness, drug overdose, or exposure. Thus far, 115 hair samples have been analyzed for cocaine, benzoylecgonine, 6-MAM, morphine, and codeine. Preliminary results for cocaine (58 percent positive) and opiates (29 percent positive) screening suggest abuse of these substances is high in the sampled population. The positives range from 16 to 72 percent for cocaine and 1 to 24 percent for opiates in the 1992 DUF Annual Report.

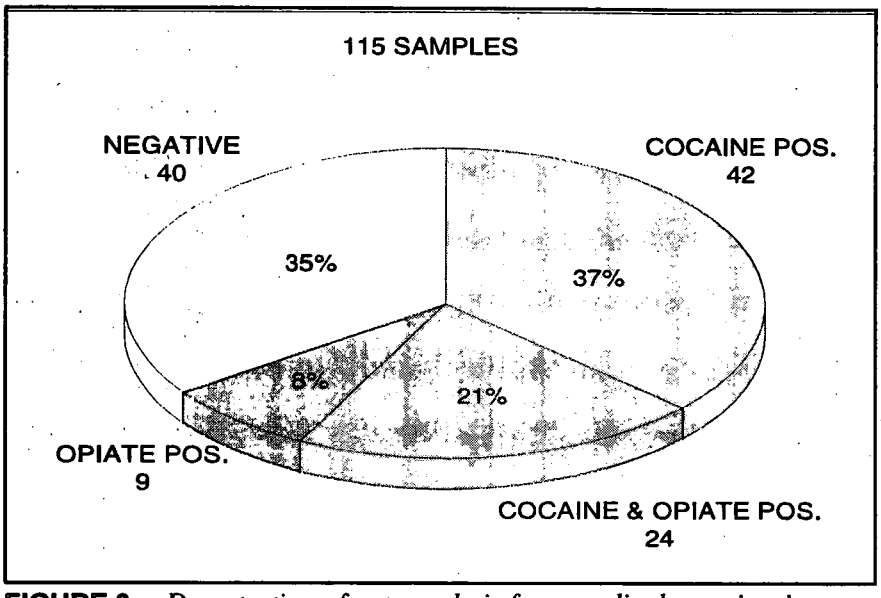


FIGURE 3. *Drug testing of autopsy hair from medical examiner's office. Percentage breakdown of autopsy hair which tested negative or positive for cocaine, benzoylecgonine, 6-MAM, morphine, and codeine.*

Only 35 percent of the autopsy samples tested negative for all 5 drugs (figure 3). This observation is consistent with the autopsy results; all of the subjects in this negative group for whom cause-of-death data were obtained had died of either accidents, illnesses, or gunshot wounds. A larger proportion tested positive for cocaine use only (37 percent) and 8 percent tested positive for opiates only. More than one-fifth (21 percent) of the subjects tested positive for both an opiate and cocaine.

A compilation of the 66 cocaine-positive hair samples netted an average concentration of 30 ng/mg of hair for cocaine and 4.6 ng/mg of hair for its metabolite, benzoylecgnine (figure 4). The median values of both drugs are much smaller, indicating most of the concentrations are at the low end of the range. The large standard deviations reveal a wide distribution in the minimum and maximum values obtained.

Results of the limited number of samples positive for opiates show the average and median values are approximately 1 ng/mg of hair or less (figure 5). The 6-MAM, morphine, and codeine average concentrations are all much smaller than the average levels of cocaine and benzoylecgnine

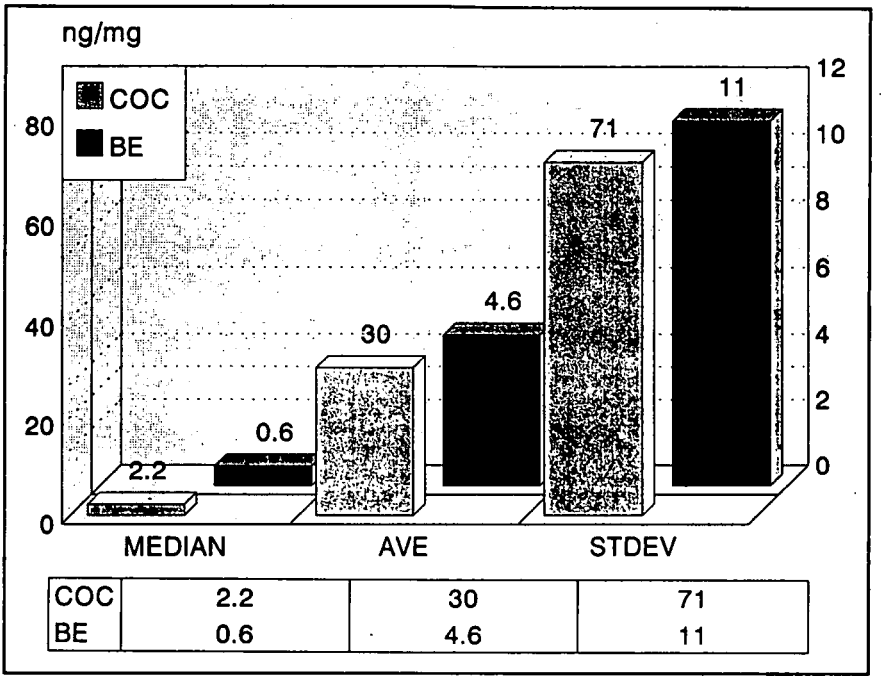
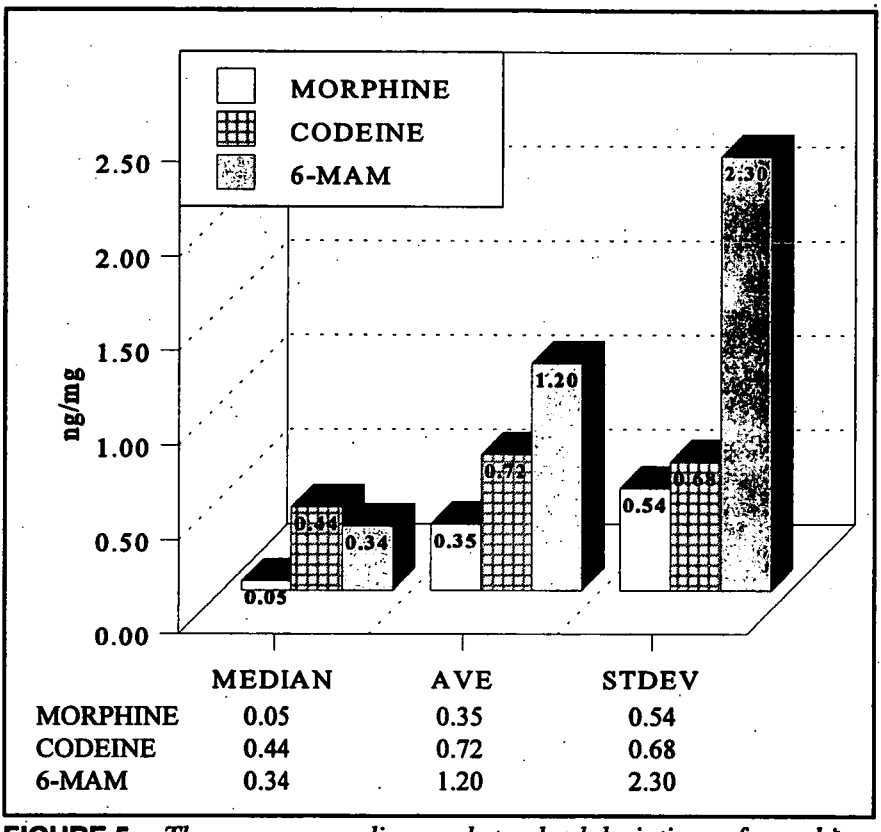


FIGURE 4. *The average, median, and standard deviation of cocaine and benzoylecgnine concentrations in 66 cocaine-positive autopsy hair samples. The left and right axes scales apply to the cocaine and benzoylecgnine concentrations, respectively.*

in the hair samples. Most of the values for the cocaine and 6-MAM-positive samples are single digit or smaller (figure 6). However, there were four samples with cocaine concentrations over 100 ng/mg.

The possibility of surface contamination of hair samples with drugs is one of the more controversial subjects in the field. It has been proposed that contamination and incorporation into the hair can result from environmental exposure to drugs, and thus sampling does not necessarily detect use of drugs. However, Baumgartner and associates (1989) have found that most hair samples do not exhibit any external contamination. They further state that drugs on the surface of hair are removed by washing with shampoo.[1] It has also been suggested by Fritch and colleagues (1992) that not all cocaine found in washes is due to external contamination. At a minimum, hair testing is still useful in forensics even if contamination exists because it is an indicator of exposure to a drug environment.



	MEDIAN	AVE	STDEV
MORPHINE	0.05	0.35	0.54
CODEINE	0.44	0.72	0.68
6-MAM	0.34	1.20	2.30

FIGURE 5. *The average, median, and standard deviations of morphine, codeine, and 6-MAM concentrations in 12, 11, and 22 drug positive autopsy hair samples, respectively.*

Results of the 115 autopsy samples indicate surface contamination of the hair is not a major problem. With an average of 0.16 for all samples that were drug positive, the cocaine wash-to-extract concentration ratio is very low and indicates most of the drug is in the interior of the hair. The median wash-to-extract ratio was 0.01; this reveals that half of the samples had less than 1 percent of their cocaine on the exterior of the hair. A total of 40 percent of the cocaine-positive hairs showed no detectable traces of cocaine in the wash, and 77 percent had a wash-to-extract ratio of no more than 0.1.

Another argument against the contamination issue is the proportionately large presence of metabolites such as benzoylecgnine and 6-MAM in hair. If surface contact were the mechanism for incorporation, unless

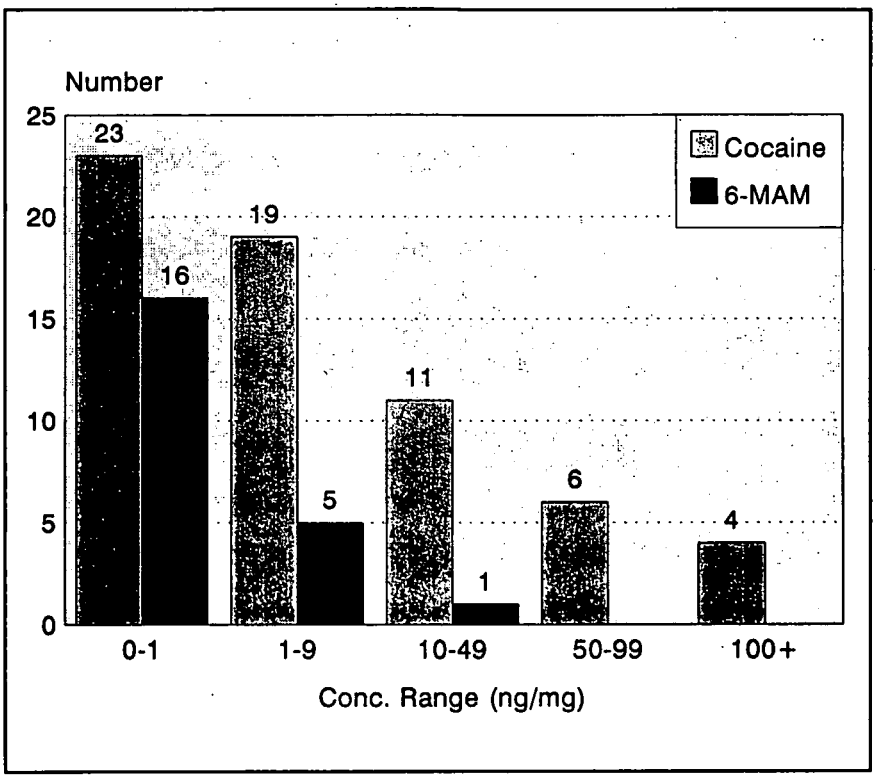


FIGURE 6. *Concentration histogram of cocaine and 6-MAM-positive autopsy hair samples.*

degradation had occurred, only original drugs would be readily detectable in contaminated hair. The contact of parent drugs with hair does not result in the formation of metabolites (Baumgartner and Hill 1993). In the authors' study of autopsy hair, only samples with traces of cocaine (< 0.3 ng/mg of hair) had undetectable levels of benzoylecgnine.

In drug abusers, toenails are less likely than hair to become externally contaminated in the daily handling of illicit drugs. A study of 20 autopsy toenails was conducted; cocaine-positive results were found in 15 of the 16 nail samples that had hair positive for cocaine. The one exception had a cocaine level of only 0.1 ng/mg in the hair. It is not surprising that the nail was negative when it is considered that nails have much lower drug concentrations than hair. In addition, cocaine metabolites benzoylecognine and cocaethylene were found in both hair and the corresponding nails, which suggests that it is unlikely the hairs are routinely contaminated by environmental sources of drugs. Therefore,

positive results in the authors' laboratory for the determination of cocaine in both hair and toenails suggests the controversy over contamination is overstated.[1]

Sample adulteration has been an issue in urine testing for some time, and may also become a concern for the validity of hair-testing results if a method were found to remove drugs from hair in vivo. In 1994, the FBI Laboratory participated in a round-robin test organized by the National Institute of Standards and Technology for the determination of drugs in hair. The blind test samples contained two sets of hair that had been spiked by soaking the specimens in solutions of drugs. Before the results of the round-robin test were known, the test samples were examined microscopically. It was observed that two of the samples had a higher sheen than the others (reddish-brown color). When the results were released, it turned out that these two samples were the adulterated preparations. The higher sheen may be the inadvertent effect of the solvent's cleansing the hairs as they were being soaked in drug solution. This observation could be of use in discovering adulterated hair specimens during testing by looking for this characteristic sheen. The scientific community has yet to agree on how to establish that hair has been adulterated or contaminated.

SUMMARY

The testing of hair for drugs has been an invaluable aid and often a necessary tool for law enforcement. It has given the forensic investigator a glimpse into the past. In conjunction with the use of urinalysis, hair testing can give a more detailed drug history on a test subject. The two tests should be considered complementary. Hair testing results have helped to incriminate those with hair positive for drugs as well as lessen suspicion for subjects with drug negative hair. Findings from a project on autopsy hair samples are internally consistent and show a positive rate for cocaine within the same range found in other survey data from booked arrestees on the prevalence of drug abuse.

ENDNOTE

1. Refer to the Technical Note at the end of the Introduction (p. 13).

REFERENCES

Baumgartner, W.A.; Black, C.T.; Jones, P.F.; and Blahd, W.H. Radioimmunoassay of cocaine in hair. *J Nucl Med* 23(9):790-792, 1982.

Baumgartner, W.A.; Hill, B.S.; and Blahd, W.H. Hair analysis for drugs of abuse. *J Forensic Sci* 34(6):1433-1453, 1989.

Baumgartner, W.A., and Hill, V.A. Sample preparation techniques. *Forensic Sci Int* 63(1-3):121-135, 1993.

Chatt, A., and Katz, S.A. *Hair Analysis: Applications in the Biomedical and Environmental Sciences.* New York: VCH, 1988. pp. 11-12.

Department of Justice. *Drug Use Forecasting 1992 Annual Report: Drugs and Crime in America's Cities.* National Institute of Justice Research in Brief, Department of Justice (DOJ) Pub. No. (NCJ)142973. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1993.

Drug Enforcement Administration. *The Microgram.* (Washington, DC) Vol. XXVII, no. 7, pp. 207-208. July 1994.

ElSohly, M.A., and Jones, A.B. Morphine and codeine in biological fluids: Approaches to source differentiation. *Forensic Sci Rev* 1(1):13-22, 1989.

Fritch, D.; Groce, Y.; and Rieders, F. Cocaine and some of its products in hair by RIA and GC/MS. *J Anal Toxicol* 16:112-114, 1992.

Liu, R.H. Important considerations in the interpretation of forensic urine drug test results. *Forensic Sci Rev* 4(1):51-65, 1992.

Martz, R.; Donnelly, B.; Fetterolf, D.; Lasswell, L.; Hime, G.W.; and Hearn, W.L. The use of hair analysis to document a cocaine overdose following a sustained survival period before death. *J Anal Toxicol* 15:279-281, 1991.

Mieczkowski, T., and Newel, R. An evaluation of racial bias in hair assays for cocaine: Black and white arrestees compared. *Forensic Sci Int* 63(1-3):85-98, 1993.

Miller, M.L.; Donnelly, B.; and Martz, R.M. Research at the Federal Bureau of Investigation in hair analysis for drugs of abuse. *Hair Testing for Drugs of Abuse: International Research on Standards and Technology.* Rockville, MD: National Institute on Drug Abuse, in press.

ACKNOWLEDGMENT

Microscopic examination of the round-robin hair specimens was conducted by Douglas W. Deedrick of the FBI Laboratory. Autopsy

samples were contributed by Edward McDonough, M.D., Office of the Chief Medical Examiner, State of Connecticut, Farmington, CT.

AUTHORS

Mark L. Miller, Ph.D.
Research Chemist
FBI Laboratory Division
Forensic Science Research and Training Center
FBI Academy
Quantico, VA 22135

Brian Donnelly, Ph.D.
Supervisory Special Agent

Roger M. Martz, B.S.
Unit Chief

FBI Laboratory Division
Chemistry-Toxicology Unit
Washington, DC 20535

167347

Patterns of Concordance Between Hair Assays and Urinalysis for Cocaine: Longitudinal Analysis of Probationers in Pinellas County, Florida

Tom Mieczkowski and Richard Newel

ABSTRACT

This chapter reports on a field trial involving the application of hair assays to a probation population. The objectives were to evaluate the general reactions of probation officers and probationers to the collection of hair samples, to compare the outcomes of the hair samples with the outcomes of urinalyses (which the probationers undergo routinely), to note and react to differences in the prevalence as indicated by the two assay types, and to assess the general monitoring potential for hair assays in a correctional setting. In general, hair assays showed an increased capability of detecting cocaine exposure when compared to urinalysis. The detection of cannabis was, however, problematic for hair. The hair assays, using urine as a comparator, appeared to result in several apparent false negatives for cannabinoids. There were no false negatives for cocaine, and an approximately fourfold increase in the detection rate when compared to urine. The collection of hair samples was not difficult and the cooperation of the probationers was quite good. Probation officers appear to prefer the use of hair specimens to urine specimen collection, and appeared enthusiastic about the potential use of hair analysis in their routine monitoring of clients.

INTRODUCTION

This chapter reports on a pilot study evaluating probationers' use of illicit drugs. Normally, probationers would undergo drug testing by urinalysis alone, but they were also monitored by hair assays. Among the major objectives of the project were to evaluate the differences, if

any, in drug prevalence rates as measured using both hair and urine specimens and assess the clinical utility of using hair assays as a supplement to urine testing in evaluating the likelihood of drug use or exposure in this sample group.

This chapter focuses primarily on the concordance outcomes of cocaine assays for hair and urine specimens. Data are presented on the overall concordance of hair and urine assays and the configurations of individual case assay results. The authors discuss the possible interpretation of these outcomes as they bear on the potential utility of hair analysis in various field settings.

BACKGROUND

Criminal justice and correctional agencies are often required by law or executive mandate to do drug testing on persons under their control. Consequently, persons who are convicted of a crime and sentenced to probation frequently are required to submit to on-demand random drug testing. Urinalysis testing, based on low-cost, rapidly readable, immunoassay technology, often is done with small portable kits read directly by the case officer; it has become universally used by correctional agencies.

Refraining from use of illegal drugs is a typical condition imposed on probationers. In the attempt to monitor convicted persons and their potential use of drugs, correctional agencies are often the most active users of drug-testing services within State criminal justice agencies. The Bureau of Justice Statistics (1992), for example, recently estimated that approximately 500,000 urinalysis tests for illicit drug use are conducted annually by correctional agencies.

Drug testing also has been shown to be effective in reducing drug consumption when implemented in probationary settings and a useful, even critical, component of treatment (Speckart et al. 1989). It also helps classify incoming offenders into particular programs (Deschenes and Anglin 1992), and can be used to verify claims of drug addiction or to monitor for exposure to methadone (Brewer 1993). Having accurate data on prevalence of drug use by type may also aid officials seeking a more effective use of system resources. It must be remembered that probation officers do not automatically issue a violation to probationers who test drug positive (+) by a bioassay. They will view the occurrence of a (+) test in a larger context and may choose to ignore it, offer some

warning or minor operational penalty, or even write an official violation. The utility of a drug assay for a probationary setting is directly tied to the extent to which it reveals accurate and refined information about a probationer's drug activities. Officers use assays in an investigatory manner in making judgments about probationers and their involvement with drugs. But it is important to stress that the assay outcomes are not judgments in and of themselves from which punitive consequences inevitably flow.

Cocaine, by a very large order of magnitude, continues to be one of the most prevalent abused, illicit drugs within the criminal justice system. Cocaine arrests nationwide, for example, occur at rates 2 to 3 times those of other popular drugs such as marijuana, heroin, or lysergic acid diethylamide (LSD) (National Institute of Justice 1993). If one examines self-reported prevalence in Pinellas County, Florida, cocaine ranks second only to marijuana among criminal offenders as the most prevalent drug of choice (Mieczkowski and Newel 1993). Its use is twice (or more) that of other hard drugs in national prevalence in criminal justice populations at all levels of processing (Bureau of Justice Statistics 1992).

Because cocaine and its metabolites are rapidly excreted from the body via urine, evasion of detection by urinalysis is a widespread problem for agencies concerned with drug monitoring. Being drug positive can result in punitive action, so probationers generally do not want to reveal drug use to their probation officer, and will normally attempt to evade detection. Random testing, which can make evasion difficult, is often problematic and expensive to effectively implement on a wide scale. This is to a large degree a result of the typically large caseloads of probation officers (see, for example, Mieczkowski et al. 1994). Users can often enhance their chances of defeating the testing with a variety of simple tactics. For example, skipping an appointment and receiving even a 24-hour delay in providing a urine specimen dramatically increases the probability of falling below the cutoff value for cocaine. Another frequently used tactic, often combined with the first, is to consume large quantities of fluids during the delay period. There is also a thriving retail trade that sells a number of organic and natural urinalysis-defeating compounds.

Evasion Tactics for Urinalysis

As a consequence, hair analysis has been suggested as a supplement to urine testing because it offers a long retrospective window of detection

and is more difficult to evade. Hair can reveal cocaine exposure from approximately 1 week to several months after it has occurred, provided the person has hair of sufficient length. It has also been suggested, although there is controversy about this, that hair assay values may correlate with the amount of cocaine ingested, and thus might be used to evaluate both qualitative and quantitative exposure to the drug (Mieczkowski et al. 1991). Several studies have established correlations between self-reported cocaine use and aggregate hair assay values (Hoffman et al. 1993; Magura and Kang, in press; Mieczkowski and Newel 1993) and between mother and neonatal hair levels (Callahan et al. 1992). Others, however, have reported inconsistent correlation outcomes with controlled-dose cocaine-administration trials with human volunteers (Henderson et al. 1993).

Preliminary research shows that it is difficult to remove sequestered drugs from hair in sufficient amounts to defeat a sensitive assay entirely (Allgood et al. 1991). Hair also has other advantageous properties: It is relatively inert, low in septic potential, easy to transport, and easy to store. Hair assay is thus an appealing technology in correctional settings.

Hair analysis has other potential uses in settings beyond drug monitoring in correctional settings. It has forensic utility, for example, in evaluation of suspicious deaths (Staub 1993). It has potential utility in drug epidemiology, especially for validation of data based on drug use self-reports. The Committee on Government Operations of the House of Representatives (1993) has recently recommended that in major drug use surveys conducted by the Federal Government, researchers investigate ways to evaluate the study's validity by using hair assays. Hair analysis has proven useful in medical contexts, both as a diagnostic tool for determining exposure to cocaine (Marques et al. 1993; Welch et al. 1990), and a therapeutic tool in drug treatment settings (Brewer 1993; Mieczkowski et al. 1994).

A review of the basic literature on hair assay technology is beyond the scope of this chapter. It has been done at length previously, and the size of this literature has now grown so large that such a discussion would fill scores of pages. Several excellent articles comprehensively review the technology of hair assays (Chatt and Katz 1988; Harkey and Henderson 1989; Mieczkowski 1992; NIDA 1995).

ASSUMPTIONS ABOUT HAIR ASSAY TECHNOLOGY

The following assumptions regarding hair assay technology have underpinned the preparation of this chapter.

1. Hair assays are able to detect cocaine and its principal metabolites benzoylecognine, ecognine methyl ester (EME), norcocaine, and several other metabolic cocaine byproducts. Detection of cocaine is possible by several different analytic techniques and can be done at high levels of sensitivity and specificity (Harkey et al. 1991). Hair assay technology for cocaine is effective whether or not the underlying technique is an immunoassay-based procedure or a chromatographic and spectrometric procedure. In effect, there is no major scientific disagreement about whether cocaine can be detected in hair. However, the appropriate interpretation of the assays has engendered controversy, a few examples of which follow. Can sufficient cocaine be acquired through casual environmental contact to confound the interpretation of the test? Does externally applied cocaine bond strongly enough to hair to defeat washing or wash-to-analyte ratios as criteria for passive versus active exposure?

2. Although individual variation of dose-assay values has not been widely studied in controlled environments, existing epidemiological data support the observation that with aggregated data sets, groups of persons who on average are more intensely using cocaine (large amounts, frequently consumed) will have higher average hair assay values than groups of persons using cocaine in smaller amounts less frequently (Graham et al. 1989; Hoffman et al. 1993; Mieczkowski and Newel 1993). However, no average dosage consumption can be quantitatively determined by reference to the quantitative value of a hair assay outcome. The authors have elsewhere recommended that quantitative hair assay data be treated as rank-order data and comparisons of repeat assays be used only intrasubjectively in clinical applications (Mieczkowski and Newel 1993).

3. Because the range of individual biovariability for cocaine assays of hair is not known, the comparison of assay values across subjects is done with substantial risk of accurate interpretation. But the comparisons of assay values taken over time for a specific individual appear to be a useful method in many circumstances for determining relative intensity of exposure over time (Brewer 1993; Martz et al. 1991).

4. Hair assays, like all other assays of tissues and fluids, measure only exposure to a substance. Generally, assays cannot themselves determine the actual method or conditions under which the exposure took place. They can only provide limited information. Decisions regarding the volitional ingestion of illicit drugs will always require human judgment. Biological assays can help support or refute particular judgments but cannot make them.

5. Passive contamination is an important consideration in making decisions about the nature of drug exposure in any assay procedure, including urine, blood, or other tissues. Researchers, for example, have reported that they cannot completely remove passively applied cocaine from the hair surface after in vitro vapor contamination (Cone et al. 1991). However, the distinction between external contamination and ingestion is sometimes clinically irrelevant. It has been proposed that passively exposed hair and hair from cocaine users can be distinguished on the basis of the ratios of wash assay values to analyte values, and, when possible, of endogenous metabolites (Baumgartner and Hill 1990, 1992). If this is correct, then a complete removal of external contaminants may not be required in many clinical circumstances. Koren and colleagues (1992) have reported on an application of this procedure that allowed them to readily distinguish passive from active contamination. Cone (1994) has recently suggested that cocaine-to-benzoylecognine ratios greater than 0.05 nanograms/milligram (ng/mg) may distinguish use from contamination, and that norcocaine and cocaethylene may, in some circumstances, act as definitive markers of cocaine ingestion as opposed to environmental exposure and surface contamination.[1]

In this chapter, the authors hold the view that the distinction between inadvertent casual exposure and meaningful, frequent contact via consumption can be made with a relatively small chance of error in most clinically relevant circumstances. External contamination versus internal (inadvertent or unknown) contamination can be evaluated by using both wash kinetic procedures and relying, when possible, on detecting endogenous metabolites (Koren et al. 1992).[1] Furthermore, the use of conservative cutoff values for evidentiary applications can help further reduce the likelihood of a false determination. Walsh (unpublished data), for example, has done a long-term quality assurance study of Baumgartner's assay technique. Walsh found that during the submission of more than 900 blind samples using both positive and negative standards, no false positive assays (i.e., reporting the presence of a drug

in a negative control) were reported; there were only five false negatives (i.e., failure to detect a drug in a known positive standard) and these were all in samples categorized as low-concentration standards.

METHODOLOGY

The present study relied upon volunteer participation by both probation officers and probationers. It is thus a convenience sample, and there are no statistically meaningful ways that these data can be generalized. The sample was created with the permission and cooperation of the Florida Department of Corrections. A detailed description of the study methods can be found in Mieczkowski and colleagues (1994).

A solicitation to all active probation officers in the Pinellas/Pasco County region was issued by the research team to recruit 20 volunteer officers as participants. As an incentive, the volunteer officers were given a training stipend of $200, a commendation and recognition plaque, and a letter of recognition for their files on completion of their participation. Each volunteer officer was asked to identify and recruit 8 to 10 probationers in his or her caseload who were currently undergoing regular monthly urinalysis. Their task was to enlist the cooperation of these persons during a 6-month project in which the officer would collect a monthly urine and hair specimen from each probationer. Probationers who volunteered received the incentive of having the project pay for the routine urinalysis (which they would normally have to pay for themselves), which represented a cost savings to them of approximately $36 in laboratory fees (note that these probationers had to undergo monthly urinalysis as a normal condition of probation, regardless of their participation in the project).

Probationers who volunteered also underwent a special, one-time interview at the project startup, administered by their case officer. This interview queried them about, among other things, their drug use history, their hair hygiene habits, and several aspects of their activities, such as recreation and water sports, that have been suggested as having possible impact on the outcome of the hair assays. Outside of this intake interview, all other interactions between probation officer and probationer were designed to be as they would occur routinely. The objective was to make the hair assay protocol as unobtrusive and natural to the normal operational context as possible. While urinalysis outcomes were reported to probation officers (as would be true of the normal routine), hair assay

values were not. No decisions of any sort were made on the basis of using hair assays to establish abstinence or exposure to illicit drugs.

The urinalyses were done by Operation PAR's certified laboratory using enzyme-multiplied immunoassay technology (EMIT) and employing current National Institute on Drug Abuse (NIDA)-endorsed cutoffs for urinalysis. Hair segments were collected; the first 2.6 centimeters (cm) were used in the assay. (Hair samples roughly correspond to behavior over the past 60 days.) Only about 1 percent of hair specimens were of shorter length, and that length ranged from 1.4 to 2.0 cm. The hair assays were analyzed using thresholds recommended by the testing laboratory for epidemiological research work. For cocaine, this threshold is 2 ng/10 mg of hair specimen. In field applications, a higher cutoff value of 5 ng/10 mg is generally recommended. Tandem mass spectrometry confirmations were done on a number of cannabinoid cases (approximately 75). Data on the outcomes of these confirmations have been reported elsewhere (Mieczkowski 1995).

DATA

The volunteer officers were able to recruit 152 probationers, and over the course of the 6-month project 62 were lost for a variety of reasons. By the end of the project, there were 89 probationers who had been enrolled since the first month. Recruitment and retention of probationers and the number of hair and urine samples retrieved each month are reported in table 1.

Of the 89 cases with 100 percent participation, 36 were negative on all assays (both hair and urine) for all drugs; 53 had at least one drug (+) assay on at least one sample. Thirty-six completed cases were drug (-) for all assays and all specimens, as were 26 incomplete cases. Thus, "double-drug negatives" was the most common outcome. The second most frequent outcome was "double positive," that is, if one specimen were positive, it was highly likely that the other specimen would be positive as well. There were 33 such complete cases with at least one (+) assay for each specimen. The complete series of these 33 outcomes for cocaine, cannabinoids, and opiates is listed in appendix 1.

TABLE 1. *Summary count of samples.*

Number of hair and urine samples	Number of cases	Percent (rounded)
1	152	21.7
2	135	19.3
3	117	16.7
4	104	14.9
5	101	14.3
6	89	13.0
Total	698	100.0

The least likely outcome was the occurrence of a positive urine assay with a negative hair assay, and that was equally true for complete and incomplete cases. The final alternative, a (-) urine assay but a (+) hair assay, was also less likely than the double (+) or double (-) outcomes, but more frequently occurring than hair (-)/urine (+) outcomes.

Table 2 is a cross-tabulation that compares dichotomous hair and urine outcomes for cases with six pairs of specimens. There was a loss of some samples due to insufficient mass, leaving the number of assayable urine and hair sample pairs at 503. Table 2 presents the joint outcome distribution of the hair assays for any drug in the hair panel, and any drug in the urine panel for which both hair and urine specimens were tested (n.b., this excludes benzodiazepines and amphetamines, which were a part of the urinalysis panel but not included in the hair assay). The single most frequent outcome is the concordance between double negative cases (N = 260), while the least frequent outcome is a urine (+)/hair (-) (N = 12). Hair (+)/urine (-) cases constitute the second most frequent combination (N = 145), and double (+) cases the third most prevalent (N = 86).

The basis of the analytic approach here is to assume that different outcome probabilities are associated with different cells. These differential outcome likelihoods are based on what the concordant and nonconcordant cells are likely to represent in clinical reality. In addition,

TABLE 2. *Contrasting hair and urine samples for any assayed drug.*

Hair assay for any drug	Urinalysis for any drug (-)	(+)	Row total
(-)	260	12	272 (54.1%)
(+)	145	86	231 (45.1%)
Column total	405 (80.5%)	98 (19.5%)	503 (100.0%)

these assumed probabilities reflect the experiences of the authors' earlier work and the outcome patterns found in approximately 2,000 cases they have previously examined.

CONCORDANT CASES

In any given criminal justice population, some number of persons will test negative by both assays, cases the authors characterize as "double (-)'s." In Pinellas County, this "double (-)" pattern has consistently been the most prevalent of all possible cell outcomes. Generally, the authors believe that the most plausible clinical interpretation of this outcome is that it indicates a person who is not exposed, or is exposed below the measurable limit of detection or cutoff value for the assayed drug for both chronic and acute time frames.

The authors have usually found subjects who are (+) on both hair and urine assays ("double (+)'s") to be third in ranking the prevalence of cell frequencies for the 2-by-2 tables. The most plausible interpretation of this finding seems to be that it indicates chronic exposure to the assayed substance. The authors have also found in earlier work with cocaine users that persons in this category who show a high concentration of cocaine in their hair assay are very likely to test urine (+) for cocaine (Mieczkowski and Newel 1993). Research on arrestee populations in Pinellas County showed that when the concentration of cocaine exceeds 10 ng/mg of hair, the likelihood of being simultaneously urine positive for cocaine approaches 90 percent (Mieczkowski and Newel 1994).

NONCONCORDANT CASES

There are two possible nonconcordant outcomes: hair (+) and urine (-) or hair (-) and urine (+). While each of these outcomes is nonconcordant, each implies quite different interpretative possibilities. The authors have found in previous work that, with cocaine, there are substantial numbers of cases that are hair (+)/urine (-) and few urine (+)/hair (-) cases.

The authors have interpreted this general pattern—that over many cocaine assays one should find substantially more hair (+)/urine (-) outcomes—as an indicator of the ability of the hair assay to accurately detect cocaine for a longer retrospective time than urinalysis. However, this capability and its exact relationship can obviously be influenced by many factors, including the amount of drug consumed, the potential to become heavily environmentally contaminated, the purity of drug consumed, and the use of particular cutoff values for the assay procedures.

Considering cocaine in particular, one category of nonconcordant cases, hair (-)/urine (+), is of special significance. These cases are of particular interest because one expects to find very few, if any, such cases in these sample populations. Because cocaine is rapidly excreted from the urine, the plausible ways by which a person can become hair negative and urine positive are limited. Previous work has supported this conjecture. The authors believe that for a drug rapidly excreted via urine (e.g., cocaine) it would be difficult to explain a high rate of frequency for these cases, especially in a criminal justice-based population with a substantial history of drug involvement. While one would expect a few persons to be assayed as urine (+)/hair (-) for cocaine, large numbers would be an indication of the failure of the hair assay. The authors have previously published hair and urine data on cocaine prevalence rates within criminal justice populations that have corroborated the expectations of few hair (-)/urine (+) cases. In the authors' previous work, these cases appear at rates of less than 1 out of every 100 persons tested.

As table 2 has shown, considering any drug for which both the urine and hair were assayed, 12 samples derived from 9 cases fall into the "least plausible" category (cell II) of being hair (-) but urine (+). These cases are termed "paradoxical" given the reasons outlined above. Table 3 is a listing of the 9 cases from which these 12 paradoxical samples were derived and the substance detected.

As table 3 shows, of the 12 samples, 10 are (+) for cannabinoids and 2 are positive for opiates. It is important to note that none of these cases involves cocaine. If the analysis is expanded to include probationers who

TABLE 3. *Urine (+)/hair (-) cases.*

Case #	Sample #	Substance detected
2-8	2	Cannabinoids
	3	Cannabinoids
	6	Cannabinoids
4-4	1	Cannabinoids
5-4	1	Cannabinoids
11-9	1	Cannabinoids
12-11	1	Cannabinoids
	2	Cannabinoids
13-5	6	Opiates
13-8	1	Cannabinoids
15-6	1	Opiates
17-3	1	Cannabinoids

did not complete the study, one finds two cases that have single cocaine (+) urine and no cocaine (+) hair. Table 4 displays the concordance of hair and urine cocaine assays for all cases, both completed and non-completed.

Table 4 includes all 698 hair and urine specimens from the 152 original probationers, including specimens from cases that did not complete the project. In cell II, one finds two cocaine (+) urines that have corresponding hair (-) assays for cocaine, both coming from incomplete cases.

In both situations, the cocaine (+) urine was obtained on the last probationer visitation, so no subsequent hair samples were gathered to evaluate whether the hair in later assays would test cocaine (+). Remember that cocaine detected in the urine at time t_1 would not be detected in the hair for at least 7 days. Because of this time differential for the two specimens, one would not expect the hair assay to detect very recent cocaine use. Had additional hair specimens been taken from these persons, the assay

TABLE 4. *Contrasting hair and urinalysis outcomes for cocaine.*

Hair assay cocaine	Urinalysis for cocaine (-)	(+)	Row total
(-)	592	2	594 (85.2%)
(+)	80	24	104 (14.8%)
Column total	672 (96.3%)	26 (3.7%)	698 (100.0%)

might have detected the cocaine indicated by the urine. Pertinent information on these two paradoxical cases (#18-8 and #4-5) is listed below. Table 5 shows the concentration of drug in hair (in ng/10 mg hair analyte) except for marijuana, which is dichotomized. The case tables also show whether the urinalysis was positive or negative, and, if positive, for what drug or drugs. The last column shows the time interval in weeks between each specimen collection. Self-reported drug use is not shown in the tables.

TABLE 5. *Findings for two paradoxical cases.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
Case #18-8							
1	0	0	QNS	(+)	Cannabis	0	
2	0	0	QNS	(+)	Cannabis	0	2
3	0	0	QNS	(+)	Cocaine	0	4
Case #4-5							
1	0	0	(+)	(-)	0	0	
2	5	0	(+)	(-)	0	0	6
3	19	0	QNS	(-)	0	0	7
4	0	0	QNS	(-)	0	0	6
5	0	0	(+)	(+)	Cocaine	0	2

KEY: Coc = cocaine, Ops=opiates, Mj=cannabinoids, hr = hair, QNS = hair sample quantity insufficient for analysis. These abbreviations also apply to subsequent case tables.

Case #18-8 was a noncompleted case with a cocaine (+) urinalysis on the final urine specimen. Case #4-5 has a similar configuration to case #18-8. Both had a single cocaine positive urine on their last collected sample. But notice that for case #4-5, cocaine had been detected in the hair in earlier samples (2 and 3).

If one considers the cocaine outcomes using only cases that completed the entire 6 months of the study, there are no paradoxical outcomes in the data set; that is, no cases that had a urine (+) for cocaine, but a (-) hair assay for cocaine. In short, when cocaine was found in the urine, it was always found in the hair.

LOOKING AT INDIVIDUAL CASES: ALL COMPLETED PERSONS TESTING (+) FOR A DRUG

Considered next are the 33 cases that have the common characteristic that they tested (+) for a drug in one or more specimens, either hair, urine, or both. These cases and their respective assay outcomes are listed in appendix 1. Recall that of the 89 completed cases, 56 were (-) for any drug, and 33 were (+) for at least one drug in at least one specimen.

Of these 33 cases, 17 were (+) for a drug other than cocaine. That is, although these cases were (+) for one of the screened drugs, they were (-) for cocaine in all urine and hair specimens. Sixteen cases were cocaine (+) in one or both specimens. The complete breakdown of these 33 cases is illustrated in figure 1.

Figure 1 shows that 6 cases had one or more cocaine (+) hair assays, but had no cocaine (+) urine outcomes and 10 cases tested cocaine (+) in both the hair and urine specimens. In no cases were there more (+) urine outcomes than (+) hair outcomes. In every case but one, the hair assay detected cocaine more frequently than did urine. In a single case cocaine was detected once by each specimen.

THE SIX COCAINE (+) CASES IDENTIFIED BY HAIR ASSAYS ONLY

The following series of tables summarizes and describes the six cases that had no cocaine (+) urine outcomes but had one or more cocaine (+)

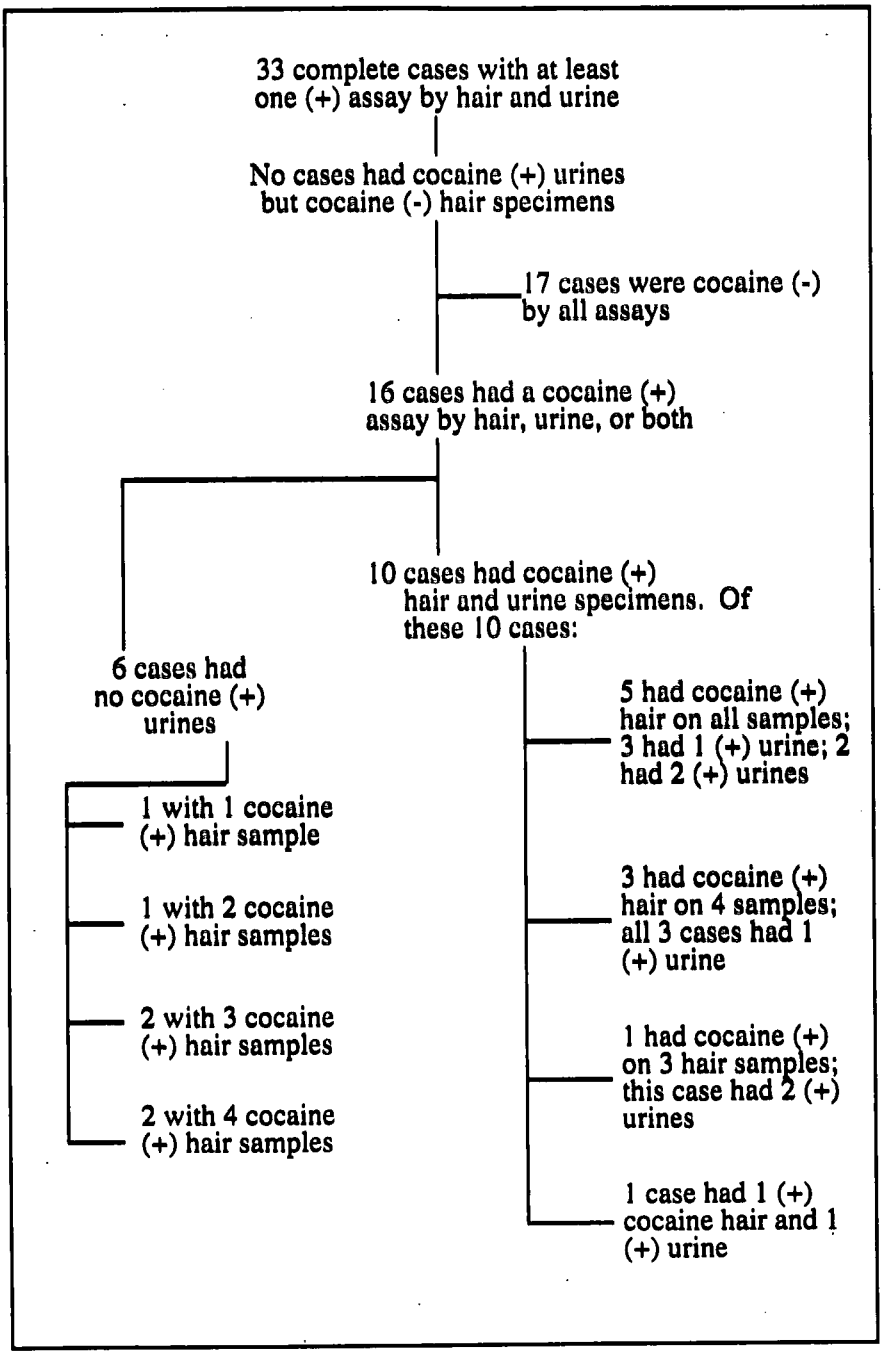


FIGURE 1. *Outcomes for cases with one or more (+) assays.*

hair assays. These probationers would have been identified as cocaine (+) if hair assays were part of the monitoring program.

Case #3-5 (table 6) had a self-reported history of cocaine and marijuana use, but assay outcomes seem to indicate abstinence during the study. Cocaine appears in the first two hair samples, but never appears in the urine. Diazepines, however, were detected in the final urinalysis.

Case #11-9 (table 7) presents an interesting pattern. This person self-reported a history of alcohol and cocaine use. Initial urine assay showed that the person tested (+) for cannabinoids at intake, but tested (-) in all five subsequent urinalyses. For samples 4, 5, and 6, the person tested hair (+) for cocaine at very high levels, but did not test urine (+) for cocaine. Also notice that the level of cocaine in the hair specimen dropped in each

TABLE 6. *Case #3-5. Urine (-), hair (+) cocaine assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	40	0	(-)	(-)	0	0	
2	9	0	(-)	(-)	0	0	4
3	0	0	QNS	(-)	0	0	2
4	0	0	QNS	(-)	0	0	3
5	0	0	(-)	(-)	0	0	4
6	0	0	(-)	(+)	Diazpn.	0	4

hair assay by roughly half over each test period, even though the testing time interval was shortened for samples 5 and 6. These reductions may indicate abstinence or markedly reduced cocaine use after the time of harvesting the fourth sample.

Case #12-1 (table 8) refused to provide any self-report information on drug use. The outcome pattern is somewhat like case #11-9 (table 7). This person tested (+) for diazepines on five out of six urinalyses, but did not test (+) for any other drug in the urine. However, every hair assay

TABLE 7. *Case #11-9. Urine (-), hair (+) cocaine assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	0	0	(-)	(+)	Cannabis	0	
2	0	0	(-)	(-)	0	0	4
3	0	0	(-)	(-)	0	0	4
4	561	0	(-)	(-)	0	0	4
5	361	0	(-)	(-)	0	0	2
6	153	0	(-)	(-)	0	0	3

TABLE 8. *Case #12-1. Urine (-), hair (+) cocaine assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	0	142	(-)	(+)	Diazpn.	0	
2	0	96	(+)	(+)	Diazpn.	0	4
3	53	850	(+)	(-)	0	0	4
4	113	75	(+)	(+)	Diazpn.	0	4
5	64	79	(+)	(+)	Diazpn.	0	4
6	26	145	(+)	(+)	Diazpn.	0	4

was opiate (+), and the quantitative values for the test were very elevated. As well, results were cocaine (+) for hair on four consecutive samples (3, 4, 5, and 6).

Case #12-8 (table 9) also refused to provide any self-report information on illicit drug use. Hair and urine samples 1 and 2 were (+) for cannabinoids, and hair samples 4 and 5 were cannabinoid (+) as well. Hair samples 1 and 2 were confirmed for cannabinoids by gas chromatography/mass

TABLE 9. *Case #12-8. Urine (-), hair (+) cocaine assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	0	0	(+)	(+)	Cannabis	0	
2	0	0	(+)	(+)	Cannabis	0	4
3	0	0	QNS	(-)	0	0	4
4	30	0	(+)	(-)	0	0	5
5	25	0	(+)	(-)	0	0	4
6	5	0	QNS	(-)	0	0	4

spectrometry/mass spectrometry (GC/MS/MS). Notice that hair samples 4, 5, and 6 were cocaine (+), but no cocaine was ever detected in the urine.

Case #13-5 (table 10) refused to provide any information on drug use and was negative for all assays on intake. However, there was a very large time gap (14 weeks) between the first and second sample collection. The second hair sample tested (+) for cocaine, but at a low level. All subsequent hair assays were (-), and only the final urine specimen has a (+) outcome

TABLE 10. *Case #13-5. Urine (-), hair (+) cocaine assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	0	0	(-)	(-)	0	0	
2	9	0	(-)	(-)	0	0	14
3	0	0	(-)	(-)	0	0	3
4	0	0	(-)	(-)	0	0	4
5	0	0	(-)	(-)	0	0	4
6	0	0	(-)	(+)	ops.	0	2

for opiates. Since there were no subsequent hair samples, the appearance of opiates in the hair following this (+) urine cannot be evaluated.

Case #14-5 (table 11) represents the last case of those persons who had at least one cocaine (+) hair sample but no cocaine detected in the urine. This person self-reported use of marijuana, but did not report any use of cocaine or opiates. As the table indicates, the person had three urine (+) outcomes for cannabinoids and two for diazepines. This person tested cannabinoid (Mj) (+) by hair assay for every hair specimen collected during the study. Additionally, the person had two low-level opiate (+) hair samples (2 and 4) and four consecutive cocaine (+) hair specimens. Neither of these substances was ever detected in the urine.

TABLE 11. *Case #14-5. Urine (-), hair (+) cocaine assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	30	0	(+)	(+)	Cannabis	0	
2	26	5	(+)	(-)	0	0	4
3	7	0	(+)	(+)	Cannabis	0	4
4	26	3	(+)	(+)	Cannabis	0	5
5	0	0	(+)	(+)	Diazpn.	0	6
6	0	0	(+)	(+)	Diazpn.	0	4

CASES HAVING BOTH HAIR AND URINE COCAINE (+) SPECIMENS

As noted in figure 1, 10 cases were cocaine (+) in both their hair and urine specimens. In the following section the authors examine these 10 cases and their outcome configurations.

Cases With All Hair Assays Cocaine (+)

Five cases had all six hair specimens as cocaine (+) and either one, two, or three urine specimens as cocaine (+). The following set of tables

presents the outcomes of these five cases. The consistently (+) cocaine hair assays support an interpretation of cocaine use, or very substantive and consistent exposure to cocaine. If a person with this pattern of assays denies using cocaine, one would certainly want to explore how these exposure levels could be attained, especially for those who have (+) urinalyses as well as consistently (+) hair outcomes.

Although case #2-4 (table 12) self-reported use of cocaine and marijuana, it was not detected in any hair or urine specimens provided by the subject. However, 4 of the 6 samples were QNS for cannabinoid assays. Cocaine was consistently detected in every hair sample at moderate levels, and was also detected in urine sample 6.

TABLE 12. *Case #2-4. All hair assays cocaine (+).*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	35	0	(-)	(-)	0	0	
2	14	0	QNS	(-)	0	0	5
3	21	0	(-)	(-)	0	0	4
4	32	0	QNS	(-)	0	0	4
5	16	0	QNS	(-)	0	0	4
6	38	0	QNS	(+)	Cocaine	0	5

Case #3-2 (table 13) self-reported use of marijuana and heroin, but did not report use of cocaine. Neither opiates nor cannabinoids were ever detected in any samples during the course of the study. Cocaine was detected in every hair specimen at low to moderate levels, and was detected twice in the urine (samples 2 and 4). Again, the cocaine (+) urinalyses linked with the consistent testing of the hair as cocaine (+) are indicative of cocaine use or exposure.

In case #3-11 (table 14), the person refused to provide any information on illicit drug use. This person tested cocaine (+) in hair on every sample at elevated values, and also tested cannabinoid (Mj) (+) on every hair

TABLE 13. *Case #3-2. All hair assays cocaine (+).*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	11	0	(-)	(-)	0	0	
2	34	0	QNS	(+)	Cocaine	0	6
3	33	0	QNS	(-)	0	0	4
4	29	0	QNS	(+)	Cocaine	0	6
5	53	0	(-)	(-)	0	0	4
6	28	0	(-)	(-)	0	0	4

TABLE 14. *Case #3-11. All hair assays cocaine (+).*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	527	0	(+)	(+)	Diazpn.	0	
2	901	0	(+)	(+)	Cocaine	0	4
3	550	0	(+)	(+)	0	0	4
4	330	0	(+)	(-)	0	0	8
5	399	0	(+)	(+)	Cocaine	0	4
6	265	0	(+)	(-)	0	0	4

sample. Only 2 cocaine urinalyses were positive (2 and 5), and there were no cannabinoid (+) urinalyses. In this situation, one sees an outcome very similar to case #3-2 (table 13), only here the cocaine hair assay values are much higher.

In case #5-2 (table 15), the person refused to provide any information on illicit drug use. This person tested cocaine (+) on every hair assay at moderate levels, and also tested cocaine (+) on a single urinalysis

TABLE 16. *Case #5-2. All hair assays cocaine (+).*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	37	0	(-)	(-)	0	0	
2	25	0	(-)	(-)	0	0	6
3	66	0	QNS	(+)	Cocaine	0	8
4	52	0	QNS	(-)	0	0	3
5	21	0	QNS	(-)	0	0	6
6	17	0	(+)	(-)	0	0	4

(sample 3). A single cannabinoid hair sample was positive (sample 6), and half the hair samples were too small to permit a cannabinoid assay.

In case #8-3 (table 16), the person refused to provide any information on illicit drug use. The person tested cocaine (+) at moderate to high levels for every hair sample, and tested cocaine (+) for a single urinalysis (sample 2). The quantitative values are consistent in samples 1 through 4, then increased by almost twofold in samples 5 and 6. This individual tested (-) for all other drugs.

In the authors' view, these cases demonstrate either failure to detect, or sporadic detection of, cocaine by urinalysis with unreliable self-reports to the probation officer. This stands in contrast with the consistent detection of cocaine by hair assay. This analytic result suggests that hair analysis can be a useful comparison for urine outcomes.

Cases With Four or Fewer Cocaine (+) Hair Assays

There are three cases where four of the six hair samples tested cocaine (+). In all three of these cases, there was only one cocaine (+) urine specimen. The following tables present the outcomes for these three cases.

TABLE 16. *Case #8-3. All hair assays cocaine (+).*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	112	0	(-)	(-)	0	0	
2	117	0	(-)	(+)	Cocaine	0	4
3	103	0	(-)	(-)	0	0	4
4	134	0	(-)	(-)	0	0	3
5	222	0	(-)	(-)	0	0	3
6	207	0	(-)	(-)	0	0	5

Case #6-12 (table 17) self-reported use of cocaine. The pattern demonstrated is interesting in that it is compatible with desistence from use at the outset of the study and a binge episode detected by the fourth

TABLE 17. *Case #6-12. Four cocaine (+) hair assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	43	0	QNS	(-)	0	0	
2	0	0	(-)	(-)	0	0	4
3	0	0	QNS	(-)	0	0	2
4	230	0	QNS	(+)	Cocaine	0	8
5	120	0	QNS	(-)	0	0	6
6	11	4	QNS	(-)	0	0	3

hair and urine samples. Notice the 8-week gap between samples 3 and 4, and the high corresponding cocaine value for sample 4. The drop in hair assay values over the following two samples is interesting and consistent

with the possibility that abstinence or marked reduction of cocaine use occurred after the fourth sample was collected.

Case #13-2 (table 18) self-reported use of cocaine and exhibited fairly consistent cocaine (+) values in hair. Notice the 7-week gap between samples 2 and 3, and then the consequent detection of cocaine in both hair and urine specimens. Note as well that while cocaine continued to be detected in hair samples 4, 5, and 6, all subsequent urinalyses were negative. Case #17-1 (table 19) self-reported use of opiates and cocaine. Cocaine was detected in the second hair and urine samples, with the hair assays showing several sequential (+) outcomes, although the quantitative measure of the subsequent samples diminishes to very low levels by the fourth sample. Also note that although opiates appear in three urine specimens, they are never detected in the hair specimens at the same time.

TABLE 18. *Case #13-2. Four cocaine (+) hair assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	0	0	(-)	(-)	0	0	
2	0	0	(-)	(-)	0	0	3
3	33	0	(-)	(+)	Cocaine	0	7
4	84	3	(-)	(-)	0	0	3
5	104	0	(-)	(-)	0	0	5
6	50	0	(-)	(-)	0	0	3

One case (table 20) had three cocaine (+) hair samples and two cocaine (+) urinalysis. One case (table 21) had a single cocaine (+) hair assay and a single cocaine (+) urinalysis. The outcomes of these two cases are presented in the following tables.

Case #20-5 (table 20) self-reported use of marijuana only and was (+) for cannabinoids in hair for every sample taken. Although the initial cocaine hair sample was positive, the simultaneously taken urine sample tested (+) for opiates but negative for cocaine and cannabinoids.

However, note that for samples 5 and 6 the person tested cocaine (+) by both hair and urine; the timespan between these two samples was relatively short.

TABLE 19. *Case #17-1. Four cocaine (+) hair assays.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	32	0	(-)	(+)	Opiates	0	
2	34	0	(-)	(+)	Cocaine	0	4
3	11	0	QNS	(-)	0	0	4
4	5	QNS	QNS	(-)	0	0	4
5	QNS	QNS	QNS	(+)	Opiates	0	5
6	0	0	QNS	(+)	Opiates	0	4

TABLE 20. *Case #20-5. Cocaine (+) hair samples, 2 cocaine (+) urinalyses.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	43	0	(+)	(+)	Opiates	0	
2	0	0	(+)	(-)	0	0	6
3	0	0	(+)	(-)	0	0	8
4	0	0	(+)	(-)	0	0	5
5	53	0	(+)	(+)	Cocaine	0	4
6	5	0	(+)	(+)	Cocaine	0	2

This case also shows that the hair assays failed to detect the opiate (+), which should have appeared in a later hair sample. The ability to evaluate opiates, and specifically heroin, in hair and urine has been

constrained by a number of factors, the most important of which is the very low numbers of opiates in the samples, less than 0.5 percent cumulatively for all the authors' sampling over the past 5 years. Furthermore, opiate detection by immunoassay is problematic because so many codeine-based opiates and opiate analogs are used in legitimate medications. The hair assay reagent used in this study is insensitive to codeine-based opiates, and optimized for morphine sensitivity in order to recognize heroin exposure. Of course, identification of a specific opiate compound requires the use of a nonimmunoassay-based GC/MS analytic procedure.

Case #6-6 (table 21) self-reported use of marijuana and cocaine. A single (+) initial urinalysis indicated the presence of both cocaine and cannabinoids, but the first three hair specimens were of insufficient mass to be tested for cannabinoids. The third hair sample tested as a low (+) for cocaine, which was the only substance detected by the hair assays, approximately 8 weeks after the initial cocaine (+) urine result.

TABLE 21. *Case #6-6. A single cocaine (+) hair and urine assay.*

Sample #	Coc hr	Ops hr	Mj hr	Urine assay	Urine1 result	Urine2 result	Interval (weeks)
1	0	0	QNS	(+)	Cocaine	Cannabis	
2	0	0	QNS	(-)	0	0	4
3	10	0	QNS	(-)	0	0	4
4	0	0	(-)	(-)	0	0	10
5	0	0	(-)	(-)	0	0	3
6	0	0	(-)	(-)	0	0	4

DISCUSSION

The use of hair assays as a drug-monitoring technique, as noted in the introduction, offers several potential advantages not available with urinalysis. Experiences during this pilot project indicate that hair assays can be used in probationary drug monitoring without major impediments to their introduction. Based on survey and interview data with the field

officers, the authors believe the assays would be well received by both correctional officers and probationers themselves.

Basic Detection Capabilities

The data collected in this project, in the authors' view, demonstrate a consistent and recognizable outcome pattern for cocaine. As elaborated in the body of the chapter, the authors believe these configurations support an interpretation of the efficacy of the hair assay for cocaine analysis. Occurrences of hair (-)/urine (+) outcomes (which the authors have termed the "paradoxical" type) continue to be rare. This is true not only for the data presented here for probationers; over the past 5 years in analyses of slightly more than 2,000 cases, only a dozen or so cases of this type have been identified. Furthermore, this pattern has been reported by others, including Wish (1994), Feucht and colleagues (1994), Magura and Kang (in press), Mieczkowski and coworkers (in press), and Baer and colleagues (1991). Because cocaine is rapidly excreted in the urine, and if the hair assay reliably detects exposure to cocaine, then the patterns of outcomes must generally conform to the type delineated here.

Findings related to marijuana are not presented in this chapter, but it is mentioned here in passing that the marijuana assay patterns also support the authors' interpretation of the critical role excretion rate plays. When one looks at marijuana, which has a much longer half-life in the urine than does cocaine (i.e., it is excreted much more slowly), one can see a marked lessening of the effect consistently seen with cocaine. That is, a considerable number of cases are cannabinoid (+) in urine but (-) in hair. The authors believe that this is due to the compound effect of urine being a particularly good medium for cannabinoids and hair being a weak one. For several reasons, and ones that are not well researched, cannabinoids concentrate relatively poorly in the hair. For example, while nanograms are the typical unit of measure for cocaine, picograms and femtograms (one quadrillionth of a gram) are the ranges in which marijuana is typically assayed.

It is commonly recognized that an indirect approach such as used here is not the ideal or optimal method to evaluate hair assay technology. However, it is a useful and pragmatic approach if one considers the constraints upon any researcher seeking to use a controlled-dose administration method. In fact, such an approach has been done (Henderson et al. 1993) and, as noted earlier, it produced ambiguous results. However, the

researchers were compelled to use low doses of cocaine relative to typical consumption levels because of limitations imposed by the use of human subjects. Doses in Henderson and colleagues' study were many times lower than what would be considered normal for heavy and chronic users of cocaine in criminal justice populations. It is important to bear in mind that at the lowest recommended clinical cutoff value of 5 ng/10 mg of hair, not a single hair segment in the Henderson and colleagues' study would meet the standard required by the present research to be called a clinical positive.

An epidemiological and clinical approach represents the only realistic way to determine the outcomes of hair assays in consistent, chronic, and high-dose users of cocaine and crack cocaine. It is unlikely (and rightly so) that the sorts of conditions that prevail in the cocaine and crack subculture regarding quantities and modes of drug administration will ever be duplicated under laboratory conditions, or would ever be permitted to be done in a laboratory setting. Cocaine users on the street have relatively open access to cocaine, constrained only by their financial resources. In the authors' experience with binge users of cocaine, it is common that they may consume several grams a day.

The general experiences of this project also lead to the conclusion that the hair assays in probationary field settings could be both feasible and useful in communities with high cocaine prevalence rates. Indeed, it has already been done and continues to be done in a variety of settings. The data show that it would be welcomed in some circumstances if it would reduce the demand on correctional officers for obtaining urine specimens from their cases. (In Florida, at any rate, officers in this study expressed extreme distaste for observing urination and would much prefer to take hair specimens.) Furthermore, the authors believe many probationers would prefer giving hair specimens to observed urinations. Hair assays, for example, could be used as an initial screening device to assign probationers to risk pools with different rates of urinalysis testing. It is likely that this would be well received in the field.

Difficulties in Implementation

The most significant problem facing implementation is the lack of widely recognized threshold or cutoff values for the hair assay for cocaine. Currently, individuals and institutions that use hair assays arrive at their own standards, typically in consultation with the analyzing laboratory. Cutoffs for any assay procedure using any sort of specimen

are ultimately fixed at the technique's limit of detection (LOD). However, since cutoffs as they are used with urinalysis, for example, reflect a concern with passive environmental exposure and inadvertent microingestion, they are typically set higher—and sometimes much higher—than the LOD in order to accommodate some quantity of detectable drug that may be present due to inadvertent exposure.

Certainly this can also be done for hair. The authors have used several cutoff points to rank order cases along a continuum of exposure. While recognizing that a person can be passively exposed and may attain detectable quantities of cocaine in the hair, the authors believe a conservative threshold, perhaps something in the range of 5 to 10 ng of cocaine/mg of hair, is an acceptable value. While there has been much speculation about passive contamination as a meaningful clinical problem, there have not been substantial published findings suggesting this would prove to be an insuperable obstacle for hair analysis. Even in the work of those most sensitive to passive contamination as a problem in the utilization of hair assays (Cone et al. 1991; Goldberger et al. 1991), experimental findings have never failed to distinguish negative controls from positive users. Furthermore, recent work by Maloney and colleagues (1994) has demonstrated that casual physical contact of cocaine-contaminated objects by drug-abstinent persons does not result in the transfer of cocaine to their person in quantities detectable even at the lowest limit of detection by GC/MS. The authors' view is that there is no compelling evidence that environmental contamination is an unresolvable clinical problem for hair analysis of cocaine, provided one is willing to accept that marginal cocaine use, because of high cutoff values, may be classified as passive contamination. In effect, by adopting very high cut-offs one accepts some false negative assays as inevitable. This is precisely the approach currently used for interpreting cocaine detection by urinalysis. Cocaine can readily be detected at levels more than 10 times lower than the current Federal guidelines of 300 ng/mL of urine. Persons who fall below this value are considered drug negative, even though they may have readily detectable amounts of cocaine metabolite in the urine.

Finally, the authors caution that the facile use of bioassays can also create a false sense of certainty about the meaning and utility of biological testing of any kind. All bioassays require prudent use and careful interpretation. When they are used solely for epidemiological estimations or other work that does not have potentially negative individual consequences, the level of error tolerance is greater than if one had to make punitive decisions

based on an assay result. In fact, it seems apparent that using both hair and urine assays in combination would be an inherently safer approach in these contexts. It is noteworthy that much of the criticism directed at hair assays is not unique to hair as a testing matrix; it is equally applicable to urinalysis, yet the sensitivity to the potential misinterpretation of urinalysis seems relatively muted in comparison.

Urinary excretion curves, for example, change as people age, yet the same concentration criteria are applied to human subjects of urinalysis for cocaine without using age-graded cutoffs. Research also shows that the excretion rate of cannabinoids is quite variable, and can result in dramatic fluctuations in the presence of cannabinoids in the urine. In some cases cannabinoids may appear months after the cessation of active use (Dackis et al. 1982; Ellis et al. 1985). Even in regard to cocaine excretion via urine, it has been reliably 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 concentration levels may move back and forth across the 300 ng/mL cutoff threshold (Burke et al. 1990; Weiss and Gawin 1988). Thus an abstinent person subject to a urinalysis could be defined as a recent user of the substance when, in fact, use may have ceased well before the conventionally accepted 72-hour window.

These concerns with the clinical use of bioassays are well founded, in the authors' view, because one is apt to treat the bioassay as the behavior that is presumed to underlie the bioassay result. Hair assays, especially for cocaine with its potentially long retrospective period, make persons more vulnerable to detection than urinalysis. But relying in any clinical situation on any single assay is, the authors believe, unwise. Bioassays should always be viewed as pieces of information that can help a person make a clinical inference, but not as a substitute for an inference. The toleration for error in the assay procedure is tied to the consequences of the inference. When high degrees of certainty are required, repetition of tests, testing by multiple technologies, use of multiple specimens from the same subject, and other such steps should be employed. Drug assay outcomes are only pieces of information. They must be integrated into a meaningful whole and interpreted in the exercise of human judgment. Clinical applications of bioassays, including hair assays, have as their objective the provision of information. Ideally, this information will 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.

Hair assays should certainly be used when the outcome cannot put the person undergoing the testing in jeopardy. It is hard to imagine why this should be objectionable. For example, hair assays could be readily used in epidemiological work where personal identification is not obtained. Furthermore, it seems that hair assays could be used in clinical settings to determine the absence of exposure to cocaine, since a false negative represents no legal encumbrance to the person being tested. Clearly hair assays can be used when those tested have given their permission to use them as a component, for example, to voluntary admission to a treatment program.

The additional benefits to this approach are that as they are so used, knowledge regarding their interpretation will broaden. As these first uses unfold, they will provide a larger database from which further refinements and more profound understanding will emerge about this new technology.

ENDNOTE

1. Refer to the Technical Note at the end of the Introduction (p. 13).

REFERENCES

Allgood, C.C.; Sniegoski, L.; and Welch, M. The analysis of human hair for drugs of abuse. *Proceedings of the 39th American Society of Mass Spectometry Conference on Mass Spectrometry and Allied Topics.* Nashville, TN, May 19-24, 1991.

Baer, J.; Baumgartner, W.; Hill, V.; and Blahd, W. Hair analysis for the detection of drug use in pretrial, probation, and parole populations. *Federal Probat* 55(1):3-10, 1991.

Baumgartner, W., and Hill, V. Hair analysis for drugs of abuse: Decontamination issues. *Proceedings of the 2nd International Congress of Therapeutic Drug Monitoring and Toxicology.* Barcelona, October 9-12, 1990.

Baumgartner, W., and Hill, V. "Hair Analysis for Drugs of Abuse: Forensic Issues." Paper presented at the International Symposium on Forensic Toxicology, FBI Academy, Quantico, VA, June 15-19, 1992.

Brewer, C. "Hair Analysis as a Tool for Monitoring and Managing Patients on Methadone Maintenance." Paper presented at the First International Meeting on Forensic Hair Analysis, Genoa, Italy, December 6-8, 1993.

Bureau of Justice Statistics, Office of Justice Programs. *Drugs, Crime and the Justice System.* Pub. No. NCJ#-133652. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992.

Burke, W.; Ravi, N.; Dhopesh, V.; Vendegrift, B.; and Maany, I. Prolonged presence of metabolite in urine after compulsive cocaine use. *J Clin Psychiatry* 51(4):145-148, 1990.

Callahan, C.; Grant, T.; Phipps, G.; Clark, G.; Novack, A.; Streissguth, A.; and Raisys, V. Measurement of gestational cocaine exposure: Sensitivity of infant's hair, meconium, and urine. *J Pediatr* 120(5):763-768, 1992.

Chatt, A., and Katz, S. *Hair Analysis: Applications in the Biomedical and Environmental Sciences*. New York: VCH Publishers, 1988.

Committee on Government Operations, House of Representatives. *Report to the Chairman: Drug Use Measurement: Strengths, Limitations, and Recommendations for Improvement.* U.S. General Accounting Office Pub. No. GAO/PMD-93-18, Gaithersburg, MD, 1993.

Cone, E. "The Pharmacology and Pharmacokinetics of Cocaine in Hair." Paper presented at the Society of Forensic Toxicology Special Conference on Drug Testing in Hair, Tampa, FL, October 29-30, 1994.

Cone, E.; Yousenejad, D.; Darwin, W.; and Maguire, T. Testing human hair for drugs of abuse. II. Identification of unique cocaine metabolites in hair of drug abusers and evaluation of decontamination procedures. *J Anal Toxicol* 15(5):250-255, 1991.

Dackis, C.; Pottash, A.; Annitto, W.; and Gold, M. Persistence of urinary marijuana levels after supervised abstinence. *Am J Psychiatry* 139(9):1196-1198, 1982.

Deschenes, E., and Anglin, M.D. Effects of legal supervision on narcotic addict behavior: Ethnic and gender influences. In: Mieczkowski, T., ed. *Drugs, Crime, and Social Policy*. Boston: Allyn & Bacon, 1992. pp. 167-196.

Ellis, G.; Mann, M.; Judson, B.; Schramm, T.; and Tashchian, A. Excretion patterns of cannabinoid metabolites after last use in a group of chronic users. *Clin Pharmacol Ther* 38(5):572-578, 1985.

Feucht, T.; Stephens, R.; and Walker, M. Drug use among juvenile arrestees: A comparison of self-report, urinalysis, and hair assay. *J Drug Issues* 24(1&2):99-116, 1994.

Goldberger, B.; Caplan, Y.; Maguire, T.; and Cone, E. Testing human hair for drugs of abuse. III. Identification of heroin and 6-acetylmorphine as indicators of heroin use. *J Anal Toxicol* 15(5):226-232, 1991.

Graham, K.; Koren, G.; Klein, J.; Schneiderman, J.; and Greenwald, M. Determination of gestational cocaine exposure by hair analysis. *JAMA* 262(23):3328-3330, 1989.

Harkey, M., and Henderson, G. Hair analysis for drugs of abuse. In: Baselt, R., ed. *Advances in Analytical Toxicology*. Chicago: Year Book Medical Publishers, 1989. pp. 298-330.

Harkey, M.; Henderson, G.; and Zhou, C. Simultaneous quantitation of cocaine and its major metabolites in human hair by gas chromatography/chemical ionization mass spectrometry. *J Anal Toxicol* 15(2):260-266, 1991.

Henderson, G.; Harkey, M.; and Jones, R. "Final Report: Hair Analysis for Drugs of Abuse." Final report of project activities under National Institute of Justice grant no. 90-NIJ-CX-0012, 1993.

Hoffman, J.; Wish, E.; Koman, J.; Schneider, S.; Flynn, P.; and Luckey, J. "Hair, Urine, and Self-reported Drug Use Concordance at Treatment Admission." Paper presented at the American Society of Criminology Annual Meetings, Phoenix, AZ, October 27-30, 1993.

Koren, G.; Klein, J.; Forman, R.; and Graham, K. Hair analysis of differentiation between systemic exposure and external contamination. *J Clin Pharmacol* 32:671-675, 1992.

Magura, S., and Kang, S. Measuring cocaine use by hair analysis among criminally-involved youth. *J Drug Issues*, in press.

Maloney, B.; Barbato, L.; Ihm, B.; Nipper, H.; and Cox, R. The qualitative determination of trace amounts of cocaine obtained through casual contact. *Microgram* 27(6):185-187, 1994.

Marques, P.; Tippetts, A.; and Branch, D. Cocaine in the hair of mother-infant pairs: Quantitative analysis and correlations with urine measures and self-reports. *Am J Drug Alcohol Abuse* 19(2):159-175, 1993.

Martz, R.; Donnelly, B.; Fetteroff, D.; Lasswell, L.; Hime, G.; and Hearn, W. The use of hair analysis to document a cocaine overdose following a sustained survival period before death. *J Anal Toxicol* 15:279-281, 1991.

Mieczkowski, T. New approaches in drug testing: A review of hair analysis. *Ann Am Acad Polit Soc Sci* 521(May):132-150, 1992.

Mieczkowski, T. A research note: The outcome of GC/MS/MS confirmation of hair assays on 93 cannabinoid (+) cases. *J Forensic Sci Int* 70(1-3):83-91, 1995.

Mieczkowski, T., and Newel, R. Comparing hair and urine assays for cocaine and marijuana. *Federal Probat* 57(2):59-67, 1993.

Mieczkowski, T.; Barzelay, D.; Gropper, B.; and Wish, E. Concordance of 3 measures of cocaine use in an arrestee population: Hair, urine and self-report. *J Psychoactive Drugs* 23(3):241-249, 1991.

Mieczkowski, T.; Mumm, R.; and Connick, H. The use of hair analysis in a pre-trial diversion program in New Orleans. *Int J Offender Ther Comp Criminol*, in press.

Mieczkowski, T., and Newel, R. "Epidemiological Data on Hair Analysis: Current Findings on Self-reported Drug Use, Urinalysis, and Hair Assays." Paper presented at the Society of Forensic Toxicology Special Conference on Drug Testing in Hair, Tampa, FL, October 29-30, 1994.

Mieczkowski, T.; Newel, R.; Allison, G.; and Coletti, S. "Hair Assays for Drugs of Abuse in a Probation Population: Implementation of a Pilot Study in a Field Setting." Final report of activities under National Institute of Justice grant no. 92-IJ-CX-K010, 1994.

National Institute of Justice. *Drug Use Forecast System 1992 Annual Report*. Washington, DC: National Institute of Justice, 1993.

National Institute on Drug Abuse. Cone, E.; Welch, M.; and Babecki, M., eds. *Hair Testing for Drugs of Abuse: International Research on Standards and Technology*. USDHHS, NIDA, NIH Pub. No. 95-3727, 1995.

Staub, C. "Is Hair Analysis a Useful Tool in Forensic Toxicology? The Situation in Switzerland." Paper presented at the First International Meeting on Hair Analysis as a Diagnostic Tool for Drugs of Abuse, Genoa, Italy, December 6-8, 1993.

Speckart, G.; Anglin, M.D.; and Deschenes, E. Modeling the longitudinal impact of legal sanctions on narcotics use and property crime. *J Quant Criminology* 5(1):33-56, 1989.

Weiss, R., and Gawin, F. Protracted elimination of cocaine metabolites in long-term, high-dose cocaine abusers. *Am J Med* 85:879-880, 1988.

Welch, R.; Martier, S.; Ager, J.; Ostrea, E.; and Sokol, R. Radio-immunoassay of hair: A valid technique for determining maternal cocaine abuse. *Subst Abuse* 11(4):214-217, 1990.

Wish, E. "What We Thought We Knew and What We Know: What Drug Testing Taught Us about Criminal Justice Populations." Paper presented at Hair Testing for Illicit Drugs: Application for Criminal Justice, New Orleans, December 2, 1994.

ACKNOWLEDGMENTS

This paper was prepared for the National Institute on Drug Abuse Technical Review on Improving the Accuracy of Self-Reported Surveys, September 8 and 9, 1994. This research was funded by the National Institute of Justice (#92-IJ-CX-K010), whose support is gratefully acknowledged. The authors would also like to acknowledge Operation PAR, Inc. for their support and assistance.

AUTHORS

Tom Mieczkowski, Ph.D.
Associate Professor
Department of Criminology
University of South Florida
140 7th Avenue South
St. Petersburg, FL 33701

Richard Newel, B.S.
Director
Research Division
Operation PAR, Inc.
10901-C Roosevelt Boulevard
St. Petersburg, FL 33716

APPENDIX I. *The 33 cases with at least one (+) assay are displayed below. A (+) sign indicates the assay was positive for that drug, a (-) indicates the opposite. An asterisk (*) means that the specimen could not be analyzed due to insufficient quantity of hair. Abbreviations are for cocaine, marijuana, and opiates.*

Case #1-3			Case #3-4		
hair	coc	------	hair	coc	------
	mrj	+++++		mrj	--++++
	ops	+		ops	------
		--+---			
urine	coc	------	urine	coc	------
	mrj	+-----		mrj	---+-+
	ops	------		ops	------
Case #2-4			Case #3-5		
hair	coc	+++++	hair	coc	++----
	mrj	+		mrj	------
	ops	-*--**		ops	------
		------			
urine	coc	-----+	urine	coc	------
	mrj	------		mrj	------
	ops	------		ops	------
Case #2-8			Case #5-2		
hair	coc	------	hair	coc	+++++
	mrj	+--++-		mrj	+
	ops	------		ops	--*-*+
					------
urine	coc	------	urine	coc	--+---
	mrj	+++++		mrj	------
	ops	+		ops	------

Case #3-2		
hair	coc	+ + + + +
	mrj	+
	ops	− * * *− −
		− − − − − −
urine	coc	− + − + − −
	mrj	− − − − − −
	ops	− − − − − −
Case #6-6		
hair	coc	− − + − − −
	mrj	* * * − − −
	ops	− − − − − −
urine	coc	+ − − − − −
	mrj	+ − − − − −
	ops	− − − − − −
Case #6-12		
hair	coc	+ − − + + +
	mrj	* − * * * *
	ops	− − − − − −
urine	coc	− − − + − −
	mrj	− − − − − −
	ops	− − − − − −
Case #12-8		
hair	coc	− − − + + +
	mrj	+ + * + + *
	ops	− − − − − −
urine	coc	− − − − − −
	mrj	+ + − − − −
	ops	− − − − − −
		− − − − − −
Case #8-3		
hair	coc	+ + + + +
	mrj	+
	ops	− − − − − −
		− − − − − −
urine	coc	− + − − − −
	mrj	− − − − − −
	ops	− − − − − −
Case #13-2		
hair	coc	− − + + + +
	mrj	− − − − − −
	ops	− − − + − −
urine	coc	− − + − − −
	mrj	− − − − − −
	ops	− − − − − −
Case #11-7		
hair	coc	− − − − − −
	mrj	+ + + * * +
	ops	− − − − − −
urine	coc	− − − − − −
	mrj	− − + + + −
	ops	− − − − − −
Case #13-3		
hair	coc	− − − − − −
	mrj	+ + * + + *
	ops	− − − − − −
urine	coc	− − − − − −
	mrj	+ + + + − −
	ops	− − − − − −
Case #11-9		
hair	coc	− − − + + +
	mrj	− − − − − −
	ops	− − − − − −
urine	coc	− − − − − −
	mrj	+ − − − − −
	ops	− − − − − −
Case #13-4		
hair	coc	− − − − − −
	mrj	+ + + − + +
	ops	− − − − − −
urine	coc	− − − − − −
	mrj	+ + + − − +
	ops	− − − − − −

Case #12-1			Case #13-5		
hair	coc	−−+++	hair	coc	−+−−−
	mrj	−+++++		mrj	−−−−−
	ops	+++++		ops	−−−−−
		+			
urine	coc	−−−−−	urine	coc	−−−−−
	mrj	—−−−−		mrj	−−−−−
	ops	−−−−−		ops	−−−−+
Case #13-6			Case #15-6		
hair	coc	−−−−−	hair	coc	−−−−−
	mrj	++−−−−		mrj	−−−−*
	ops	−−−−−		ops	−−++++
urine	coc	−−−−−	urine	coc	−−−−−
	mrj	+−−−−		mrj	−−−−−
	ops	−−−−−		ops	+−−−−
Case #14-1			Case #16-1		
hair	coc	−−−−−	hair	coc	−−−−**
	mrj	+++++		mrj	+*****
	ops	+		ops	−−−−**
		−−−−−			
urine	coc	−−−−−	urine	coc	−−−−−
	mrj	+++++		mrj	+−−−−
	ops	+		ops	−−−−−
		−−−−−			
Case #14-3			Case #16-2		
hair	coc	−−−−−	hair	coc	−*−−*−
	mrj	+++++		mrj	******
	ops	+		ops	−*++*+
		−−−−−			
urine	coc	−−−−−	urine	coc	−−−−−
	mrj	+++++		mrj	+−−−−
	ops	+		ops	−+−+−+
		−−−−−			

Case #14-4			Case #16-3		
	coc	------		coc	---** -
hair	mrj	+++++-	hair	mrj	*+****
	ops	------		ops	---** -
	coc	------		coc	------
urine	mrj	+-----	urine	mrj	++----
	ops	------		ops	------
Case #14-5			Case #17-1		
	coc	++++--		coc	++++*-
hair	mrj	+++++	hair	mrj	--****
	ops	+		ops	----*-
		-+-+--			
	coc	------		coc	-+----
urine	mrj	+-++--	urine	mrj	------
	ops	------		ops	+---++
Case #17-2			Case #20-2		
	coc	----*-		coc	------
hair	mrj	+-****	hair	mrj	++-+-*
	ops	------		ops	--+---
	coc	------		coc	------
urine	mrj	--+---	urine	mrj	-++--+
	ops	------		ops	------
Case #18-10			Case #20-5		
	coc	----*-		coc	+---++
hair	mrj	*****+	hair	mrj	+++++
	ops	----*-		ops	+
					------
	coc	------		coc	----++
urine	mrj	------	urine	mrj	------
	ops	----+-		ops	+-----

167348

The Validity of Self-Reports of Drug Use at Treatment Admission and at Followup: Comparisons With Urinalysis and Hair Assays

Eric D. Wish, Jeffrey A. Hoffman, and Susanna Nemes

ABSTRACT

Studies conducted in the 1970s and early 1980s concluded that people will provide valid information about their illicit drug use when research interviews are conducted under appropriate conditions. Recent studies of treated and untreated populations using improved urinalysis techniques as well as hair analysis techniques indicate that the validity of respondents' self-reports of recent drug use may be considerably less than previously reported and may differ according to a number of factors. Results are presented from a study of clients participating in the Washington, DC, Treatment Initiative study who were assessed for drug use by interview, urinalysis, and hair analysis. At intake, almost all clients who tested positive had reported their use of heroin but fewer clients had reported their cocaine use. At posttreatment followup, clients underreported both heroin and cocaine use. Findings from treatment outcome studies that fail to validate and adjust their estimates of self-reported recent drug use should be interpreted with considerable caution.

INTRODUCTION

The measurement of drug use by structured research interviews is an established technique in the social sciences. Numerous studies conducted in the 1970s and early 1980s concluded that respondents will provide valid information about their illicit drug use when the interviews are conducted by trained interviewers in a nonthreatening setting and when the respondents feel reasonably secure that their disclosures will not result in adverse consequences (Harrell 1985; Hubbard et al. 1989). Indeed, the Federal Government spends millions of dollars on surveys of household members and student populations that rely on respondents' willingness to report their illicit drug use accurately (General Accounting Office 1993).

There are three important reasons why conclusions from the early literature supporting the validity of self-reports must be reevaluated. First, most of the validity studies were based primarily on indirect measures of validity, usually assessments of internal consistency or the construct validity of responses. If a respondent's reports of drug use were internally consistent or correlated with other variables in theoretically expected ways (construct validity), the findings were interpreted as supporting the validity of the drug use self-reports. However, an important limitation of such indirect estimates of validity is that a respondent who lied consistently during the interview would have been judged to be providing valid responses. Thus, a person who underreported both drug use and other deviant behaviors would have exhibited the expected correlation between low drug use and low deviance. (See Magura et al. 1987 for an example of such a spurious relationship.) The same spurious association would be found if respondents were prone to overreporting deviance and drug use.

Even attempts to validate self-reported drug use by comparisons with official record information may lead to what at first glance appears to be evidence of the validity of self-reported drug use information. For example, Wish (1988) found the expected relationship between self-reported drug dependence and the number of previous drug arrests in respondents' criminal justice records; this was in an arrestee cohort in which there was considerable underreporting of recent drug use in comparison with the urine test results.

The second reason that conclusions of earlier validity studies should be reassessed involves the substantial improvements that have been made in the sensitivity of biological measures of recent drug use. The development of objective measures of recent drug use based on biological assays has provided researchers with tools to measure recent drug use directly and to avoid many of the problems described above. However, while urine test results have been used by researchers for almost 25 years to validate self-reports of drug use, the technology has improved so much that it casts doubt on the usefulness of early validity studies (Mieczkowski 1990).

The early urine tests used a process called thin layer chromatography (TLC), a very time-consuming and subjective laboratory test. As tests were perfected and became more sensitive and easier to interpret, it became clear that TLC had greatly underdetected the recent use of drugs, especially cocaine and opiates (Wish et al. 1983). Because TLC

underdetected the use of these drugs, the concordance between self-reported use and the urine tests was inflated in a group of people who were concealing their drug use. Drug users who reported that they had not used a drug appeared to be telling the truth because the TLC failed to detect the drug. The early urinalysis-based validity studies conducted before the advent of the more sensitive immunoassay screening tests were therefore likely to have overestimated the validity of the self-reports of drug use. Moreover, if hair analyses prove to be a more sensitive measure of drug use than current-day urine tests, the validity research using even today's sensitive urinalyses also may prove to have overestimated the validity of self-reported drug use.

A third reason for questioning the conclusions of earlier validity studies is the secular changes that have occurred with regard to attitudes toward illicit drug use. Since the beginning of the cocaine and crack epidemic and related street violence in the early 1980s and the emerging acquired immunodeficiency syndrome (AIDS) epidemic among injecting drug users (IDUs), the public has become more intolerant of drug use (Musto 1991). Earlier studies of the validity of self-reports of drug use were conducted at times when individuals may have been more likely to reveal their drug use in a research interview, which could have led to greater accuracy in self-report measures than is achieved today.

Researchers have begun to reassess the limitations and determinants of self-report measures of drug use with the more sensitive urinalysis and hair analysis (Magura and Kang 1995). The weight of the evidence suggests that the relationship between a respondent's self-reports of drug use and actual drug use behavior is more complex and variable than had been understood. For example, the evidence is overwhelming that people under the supervision of the criminal justice system greatly underreport their recent use of illicit drugs even when they are interviewed by researchers under conditions of anonymity and confidentiality (Dembo et al. 1990; Mieczkowski et al. 1991; Wish and Gropper 1990). Even arrested youth interviewed 6 months after their release in the community by experienced research interviewers, under conditions of confidentiality, have been found to conceal their recent drug use (Magura et al. 1995).

It may be expected that individuals who are interviewed while they are under the supervision of the criminal justice system or after release may never feel secure enough to disclose their illicit drug use in research interviews. However, studies of noncriminal populations have also found underreporting of recent drug use. Of the patients seeking

treatment in a medical clinic who tested positive for cocaine by urinalysis, only 28 percent reported recent use of the drug in the nurse-administered medical intake interview (McNagy and Parker 1992). Marques and colleagues (1993) studied a sample of infants and their postpartum mothers using interviews and urine and hair analyses. They found that while the cocaine levels in infant hair were correlated with analyses of maternal urine ($r = 0.28$) and hair ($r = 0.43$), the maternal self-reports of cocaine use did not correlate ($r = 0.06$) with the infant hair results. The authors concluded that self-reported drug use information routinely collected by interviewers should be interpreted cautiously.

Cook and associates (1995, this volume) found that less than one-half of the employees of a steel manufacturing plant who tested positive by urine or hair analysis reported their drug use in anonymous research interviews or group-administered questionnaires. The largest amount of under-reporting was found for cocaine/crack use. A study of occupants of shelters and residents of single-occupancy hotels in New York City and State found that only one-third of those who tested positive for cocaine by hair analysis reported ever using the drug in the telephone research interview, even though all had been informed that they would be tested (Appel 1995). Underreporting of recent drug use in comparison with urinalysis results was also reported by another study of the homeless in New York City (New York City Commission on the Homeless 1992).

While the evidence suggests that traditional interview studies in which a researcher conducts a one-time interview or periodic interviews with a research subject may be open to underreporting, it has been suggested that more sustained, ethnographic, community-based interview procedures may obtain more valid self-reports of drug use. Weatherby and associates (1994) found that when community outreach workers recruited admitted drug injectors to participate in an AIDS risk-assessment study, the urine test results confirmed their self-reported drug use. However, Wish and Mieczkowski (1994) pointed out that because the study's findings came from people recruited and interviewed because they had previously reported their drug use to the recruiter and had been informed of the impending urine test, the likelihood that the urine tests would detect underreporting in the research interview was diminished. Moreover, Falck and colleagues (1992) found considerable underreporting of cocaine and opiate use in their study of a similar sample of not-in-treatment, nonincarcerated IDUs who were not given advance notice of the urine test.

It could be argued that people in contact with the criminal justice system, the homeless, and employees may have significant reasons for under-reporting their drug use, even in confidential research interviews. One might expect, however, that drug abuse treatment clients would find little reason to conceal their drug use, especially at admission to treatment. Assessment and diagnostic tools generally rely upon the person's accurate reporting of recent drug use and associated problems. Moreover, treatment evaluation studies often depend on self-report measures of drug use at intake and at followup to assess treatment outcomes. Systematic under-reporting of drug use would greatly bias the results of such studies.

The evidence suggests that even drug abuse treatment clients may systematically underreport their drug use. Magura and associates (1987) found that only 35 percent of those receiving treatment at methadone programs who tested positive for opiates by enzyme-multiplied immuno-assay technique (EMIT) reported using the drug in the previous 30 days. Reporting was higher for cocaine (85 percent) and benzodiazepines (61 percent). These results underestimated the level of potential under-reporting, however, because clients were classified as having used a drug if they reported current use or use in the past 30 days, rather than use in the past 2 or 3 days, the period to which the urine tests were sensitive.

A comparison of the urinalysis results and self-reported drug use for clients in the Treatment Outcome Prospective Study (TOPS) 24 months after treatment found that only 33 percent of those positive for opiates reported using heroin in the previous 3 days (Research Triangle Institute (RTI) 1994). That study also found that only 40 percent of the cocaine-positive clients reported using the drug in the previous 3 days.

More recently, the Early Retrospective Study of Treatment Outcomes (RTI 1994), a study of clients receiving treatment for cocaine as a subset of Drug Abuse Treatment Outcome Study (DATOS) programs, found that only 26 percent of the 109 clients who tested positive for cocaine by urinalysis at followup 12 months after treatment reported using the drug in the previous 72 hours. Less than one-half (43 percent) of the cocaine-positive clients admitted using the drug in the past 2 weeks. Even when the researchers expanded their measure to compare the concordance between any drug-positive urine test and a self-report of the use of any drug in the past 72 hours, they reported that "... still two-thirds of those who tested positive for any drug did not report use of any drug in the past 72 hours" (RTI 1994, p. 4).

Magura and associates (1992) obtained interview, urine, and hair test information to investigate the validity of hair analysis among clients receiving methadone treatment. They found that 81 percent of clients positive for cocaine by urinalysis and 73 percent positive by hair analysis reported using the drug in the confidential research interview. The numbers were smaller for heroin, however—57 percent and 64 percent, respectively.

Hinden and colleagues (1994) found that most of those who tested positive by hair analysis for heroin (96 percent) or cocaine (89 percent) at the inception of residential treatment had reported their use of these drugs during the admission interview. However, at the posttreatment interview, only 67 percent of those positive for heroin and 51 percent of those positive for cocaine reported using the drugs. The authors speculated that people may be less likely to report drug use after treatment or when not in the protected treatment environment.

An experiment to assess the benefit of giving interim methadone maintenance to individuals on a waiting list at three methadone treatment programs provided additional information about client underreporting of recent drug use (Sowder et al. 1993). Each of these clients had been randomly assigned to an experimental or control condition. Experimental subjects were provided low doses of methadone and some support services while waiting for admission to the full program; control subjects remained on the waiting list without receiving methadone. A baseline interview was conducted with each subject at entry to the research, and a followup interview was conducted about 4 months later, but before entry to formal treatment. Urine specimens were obtained at the baseline and followup interviews.

The study found that at baseline virtually all of the experimental (97 percent) and control subjects (99 percent) who tested positive for opiates reported using an opiate during the previous 48 hours. However, slightly more than half of those testing positive for cocaine (53 percent and 62 percent, respectively) reported use of the drug in the past 48 hours. Most of the cocaine positives (over 80 percent) did report using cocaine in the past 30 days. The authors speculated that at baseline those who wanted to obtain methadone had an incentive for reporting their recent heroin use. No such incentive was present for reporting cocaine, and to some persons there may have been a disincentive to report use of drugs other than heroin.

While the experimental and control group subjects had similar rates of underreporting at baseline, marked differences were found at followup. Eighty percent of the control clients who tested positive for opiates at followup reported using the drug in the past 48 hours, but only 56 percent of the opiate-positive experimental clients reported such use ($p < 0.05$). The results for cocaine were even more disparate: 63 percent versus 33 percent ($p < 0.05$). Thus, while all subjects tended to underreport use of each drug at followup, the experimental subjects were more likely to conceal their drug use. The researchers suggested that experimental subjects may have had an incentive (e.g., social desirability) to show that the treatment they had participated in had some benefit. Although these findings need to be replicated, they suggest that treatment followup studies that rely solely on self-reported drug use to assess outcome run the risk of reporting reductions in drug use among treated versus untreated clients that may largely reflect systematic differences in underreporting. Similar concerns have been raised by Magura and Kang (1995) in their review of studies of the validity of respondent self-reports in drug treatment research studies.

In sum, the recent research literature raises important questions regarding the validity of self-report measures of drug use in studies of drug abuse treatment. At treatment admission, the validity of self-reports of drug use may depend upon the type of drug and the treatment modality. Cocaine use frequently goes unreported; people seeking methadone treatment may report the recent use of heroin even as they underreport cocaine use. Moreover, those who have completed some treatment may have special motivation to underreport all recent drug use in the posttreatment period. The remainder of this chapter presents findings relevant to some of these issues using information from research interviews, urinalyses, and hair analyses for a subsample of people participating in the Washington, DC, Treatment Initiative (DCI) study. The next section provides an overview of the DCI study and the validity substudy. The third section presents the results of the validity substudy, following which the implications of the findings and the literature for studies of treatment outcome are discussed.

THE DCI STUDY AND VALIDITY SUBSTUDY

The DCI is an experiment designed to test the efficacy of providing enhanced inpatient or outpatient treatment to clients seeking treatment in the District of Columbia. People who sought treatment at the Central

Intake Division (CID) run by the DC Alcohol and Drug Abuse Administration (ADASA) or who were ordered by the court to obtain treatment were eligible to volunteer to participate in the DCI. Volunteers were sent to the DCI Diagnostic Unit, where research staff administered a battery of interviews and psychological measures. The Individual Assessment Profile (IAP), developed for the DCI by researchers at RTI (Flynn et al. 1992), was administered to all participants before they were assigned to treatment. The IAP is a structured interview based on the longer DATOS protocol; it asks about many aspects of the client's life, including demographic information, drug use, treatment history, and criminal history. Based on the results of a clinical assessment, clients were assigned to the appropriate residential therapeutic community or outpatient treatment modality. The research staff then randomly assigned clients to either the enhanced or standard treatment program for their modality. Clients were interviewed periodically after admission and a small subsample was interviewed over the telephone or in person as part of a 3-month postdischarge followup study. More extensive followup interviews are currently being conducted with all persons assigned to one of the two residential therapeutic community programs. (A more complete description of the DCI appears in Hoffman et al. 1995.)

Intake Data Collection

To assess the validity of self-reports of drug use obtained in the IAP interview, a validity study was undertaken with all clients appearing at the diagnostic unit between September 29, 1991, and February 18, 1993. The intent was to compare the self-reports of opiate and cocaine use with the analysis results of a urine specimen and a hair sample collected by staff. Each measure is described below.

Self-Reports of Drug Use. This information was obtained from the IAP questions regarding lifetime use, frequency of use, and past month use of heroin, opiates, and cocaine. The IAP was administered by trained research interviewers at the initial in-person interview. All research participants were asked for their informed consent and told that all study data were protected by a Federal Certificate of Confidentiality.

Urine Tests. Specimens were obtained by CID staff as part of the routine medical screening at intake and analyzed by the ADASA laboratory for the presence of opiates and cocaine using standard immunoassay screening tests (e.g., EMIT). Standard National Institute

on Drug Abuse (NIDA) laboratory cutoff levels were used. Confirmation of positive results was not attempted. Both the urine and hair tests are sensitive to the class of opiate drugs or to a metabolite of cocaine, rather than cocaine itself, but for simplicity, throughout this chapter reference is made to cocaine or opiate test results. The minority of persons who self-reported use of opiates also reported using heroin. Opiate test results are therefore compared with self-reports of heroin in the remainder of this chapter.

Hair Tests. At the initial assessment, each client was asked to provide a hair sample for analysis after completion of the IAP. Clients who provided the hair specimen were given a food voucher for $10. Research staff cut a sample of hairs as close to the scalp as possible near the crown of the head, using the standard procedures established by the Psychemedics Corporation. The hair samples were sent to Psychemedics for testing for cocaine and opiates using their standard radioimmunoassay of hair (RIAH) test procedures (Psychemedics Corporation 1991). The length of the hair was cut to a maximum of 3.9 centimeters (cm), representing about 3 months of growth (Saitoh et al. 1967). Confirmation of positive RIAH results was not conducted.

Postdischarge Followup

Toward the end of the project, an attempt was made to reinterview clients who had been discharged from treatment for at least 3 months; a comprehensive followup study was not possible at the time. Clients were interviewed over the telephone or in person using a modified followup version of the IAP. All respondents were asked to provide a hair specimen for analysis, for which they were paid $10. All those who had been interviewed over the telephone were asked to go to the research office to provide the hair specimen. No urine specimens were collected. While a larger number completed the posttreatment interview, this chapter focuses on the 39 clients who also went to the research office to provide a hair specimen. Questions about drug use in the past 90 days were added to the IAP followup interview so the self-report period would correspond to the period to which the hair analysis results were sensitive.

Limitations

A number of limitations should be noted in reviewing the results. First, none of the positive urine or hair test results was confirmed. Research has found that the greatest threat to the validity of these tests is the

presence of false negatives. That is, the tests are more likely to fail to detect recent drug use than to erroneously detect drug use in a nonuser (Visher and McFadden 1991). Once the drug is extracted from the hair, the RIAH test used with the resulting solution is equivalent to that used in urinalysis. Thus, the limitations to the validity of urinalysis apply to RIAH. In other research using hair analysis (with confirmation) for high-risk populations, the current authors have found that in virtually every instance an initial positive result for cocaine or opiates by hair analysis was confirmed by gas chromatography/mass spectrometry (GC/MS), the ultimate standard for identifying drugs.

There is some controversy with regard to the possibility that clients who are exposed to external drug contamination (e.g., drugs smoked by others) may test positive by hair analysis (Mieczkowski 1992). There has also been some controversy about the impact of melanin concentrations in the hair on drug absorption and the possibility that drug metabolites in sweat may be deposited along the hair and thus complicate estimates of time of use (Harkey and Henderson 1988; Mieczkowski 1993). The laboratory used for the RIAH test analyzes wash kinetics to ensure that external drugs are removed from the hair before drugs are extracted from inside the hair. While some disagree about whether these laboratory techniques completely eliminate external contamination, the concentrations of drugs detected in the hair specimens of the research subjects in this study tend to be much higher than those detected from external contamination.[1] Further, the overwhelming majority of clients in this study who tested positive for cocaine also tested positive for opiates, which increases the likelihood that they had actually used the drugs.

Given the acknowledged high rates of false negative urine (and hair) test results, these types of toxicologic tests tend to underestimate recent drug use. This does not represent a large limitation, however, because the analyses are principally concerned with whether persons who did test positive also reported using the drugs detected.

A second limitation stems from the availability of hair and urine specimens for only 22 percent of the clients assessed during the time of the validity study. Analyses presented later in this chapter show that those who provided both specimens were likely to be older heroin users, while the remaining respondents tended to be young crack users. Had specimens been obtained from these crack-using youth, the level of underreporting might have exceeded that found among the older heroin users. Thus, the levels of underreporting of drug use presented here

could be considerably below what would be expected in a more representative sample of all persons seeking treatment.

A third limitation involves the comparability of the postdischarge followup results and those from the intake validity sample. Some of the 39 clients in the followup sample were interviewed postdischarge by telephone and some in person. Given the finding that household surveys conducted by telephone produce somewhat lower estimates of recent drug use than in-person interviews (Gfroerer and Hughes 1992), one might expect more underreporting in the followup sample than in the intake sample. However, clients interviewed on the telephone had to make a special trip to the research office to provide a hair specimen. Such compliance with the research procedures may have been related to more accurate disclosure of drug use. Another limitation of the followup component is that only eight clients in the discharge followup sample were included in the intake validity sample. Analyses presented below show that the 39 clients interviewed posttreatment differed from those interviewed at intake primarily with regard to age and heroin use. Clients in the followup sample were less likely to report daily heroin use at intake and were younger. Both factors could have been associated with greater underreporting of drug use in the followup sample. For these reasons, differences in the level of reporting of drug use between the intake validity sample and the discharge sample can only be considered as suggestive pending further replication. The ongoing, larger followup study of all inpatient DCI clients will permit a more systematic comparison of the validity of self-reports of drug use at intake and postdischarge.

RESULTS

Intake Validity Sample

During the period of the validity study, 487 people were processed by the diagnostic unit. Table 1 shows that 56 percent provided a urine specimen and 33 percent provided a hair specimen. A hair or urine specimen was obtained from 67 percent of the sample, and both specimens were obtained from 106 persons, or 22 percent of the sample. It was not clear why urine and hair specimens were not obtained for more sample members. However, if a person came to the diagnostic unit without going to the CID, a urine specimen would not have been collected. Also, hair specimens could not be obtained from the many persons who had

TABLE 1. *Percentage of interviewed clients who provided urine or hair specimen.*

Provided	N		%	
No hair or urine specimen	161		33	
Urine specimen only	165		34	
Hair specimen only	55	33%	11	56%
Urine and hair specimen	106		22	
	487		100%	

hair styles so short that a sufficient specimen could not be obtained with scissors. An estimate of the percentage of respondents from whom a hair specimen could not be obtained is not available. However, other research indicating that people are more likely to provide hair than urine samples leads the authors to believe that many of the missing hair specimens were due to short hair rather than refusal to provide a sample.

Because the analyses of the intake validity sample focus exclusively on the minority of individuals who provided both urine and hair specimens, potential differences between these individuals and the rest of the target sample were examined. Table 2 presents comparisons of the four groups formed according to whether they provided urine or hair specimens (provided neither specimen, hair only, urine only, or both). Three characteristics differentiated the groups. Clients who provided urine only or hair and urine specimens were 4 to 5 years older (mean age 38.1 to 39.2 years) and most likely to have reported heroin use in the past year (75 to 79 percent). Clients who provided both specimens were least likely to have reported daily use of crack cocaine. Ethnicity, gender, education, previous arrest, previous alcohol or drug treatment, and use of powder cocaine did not differ in the four groups. These findings suggest that the clients who provided both hair and urine specimens were older heroin users, perhaps those seeking methadone treatment. This conclusion is consistent with the fact that the CID is much more likely to obtain urine specimens to verify heroin use from individuals seeking methadone treatment.

TABLE 2. *Client characteristics by specimens provided (N = 487 clients).*

		Subjects who provided			
	(N)	No hair/ urine (161)	Hair only (55)	Urine only (165)	Hair and urine (106)
Male		73%	43%	75%	59%
Mean age		34.3*	33.5*	38.1*	39.2*
African American		95%	95%	93%	94%
Less than 12 years education		64%	57%	58%	59%
High school diploma/GED		60%	57%	52%	53%
Ever arrested		75%	71%	84%	85%
Used daily in past year					
Cocaine		28%	27%	26%	33%
Crack		32%**	36%**	27%**	15%**
Heroin		61%*	23%*	75%*	79%*
Previous alcohol/drug treatment		73%	64%	75%	84%

NOTE: Numbers (Ns) vary slightly because of missing information.
KEY: * = $p < 0.05$; ** = $p < 0.01$.

Hair Versus Urine Tests Results at Intake

The length of the hair specimens varied from 0.5 cm to 3.9 cm. This means that the window of detection for drug use by RIAH extended from 1 to 3 weeks before the interview to as long as 3 months before the interview. (Hair takes about 7 days to grow out to the scalp level (Harkey and Henderson 1988)). Thus, a cutting at the scalp represents drug use that occurred about 1 week earlier. For most drugs, therefore, the sensitivity period of hair analysis does not overlap with that of urinalysis.) Given that the urine specimens detect use of opiates and

cocaine in the 24 to 72 hours before the specimen is provided, one would expect that even in a group of chronic users, the hair would detect more users. Ninety-one percent of the clients in the intake validity sample tested positive for opiates by hair and 83 percent by urinalysis, a nonsignificant difference (table 3). The hair tests did detect much more cocaine use, however—93 percent versus 69 percent ($p < 0.01$).[1]

Both the urine and hair test results indicated considerable multiple drug use by the sample clients. Seventy percent of clients whose urine tested positive for opiates also had a positive urine test for cocaine. Eighty-five percent of those with a positive urine test for cocaine had a urine test positive for opiates. The numbers were even higher for the hair tests. Almost all clients (97 percent) who tested positive for opiates by RIAH had a cocaine-positive test and 94 percent with a hair test positive for cocaine tested positive for opiates.

TABLE 3. *Estimates of drug use by self-report, urinalysis, and hair analysis at intake interview (N = 106 clients who provided urine and hair specimens).*

	Self-report			
	Ever used ≥ 5 times	Used past 30 days	Urinalysis	Hair
Opiates/heroin	93%	91%	83%	91%
Cocaine	90%	71%	69%*	93%*

KEY: * = $p < 0.01$.

Estimates of Cocaine and Heroin Use at Intake

Because the IAP did not include questions regarding drug use in the past 24 to 72 hours or past 90 days, direct comparisons of self-reported use and urinalysis and RIAH results during their exact detection periods were not possible. Comparisons were therefore made with respect to self-reported use in the past 30 days or lifetime use of the drug on five or more occasions. The results in table 3 show fairly similar estimates for heroin/opiate use based on all four measures.

The greater reporting of opiate use is clearly shown in table 4. Between 96 percent and 100 percent of the clients who tested positive for opiates by hair or urinalysis reported use of the drug on at least one of the three self-report measures at intake. The numbers were lower for cocaine.

TABLE 4. *Percentage of clients positive for opiates or cocaine by urine or hair at intake who reported using the detected drug.*

	(N) positive for opiates by		(N) positive for cocaine by	
	Hair (97)	Urine (88)	Hair (99)	Urine (73)
Reported using detected drug five or more times in life	97%	100%	91%	95%
In past year	97%	100%	79%	87%
In past 30 days	96%	99%	75%	82%

NOTE: Ns may vary slightly because of missing information.

While there was some underreporting of cocaine use, there was an association between the self-reported frequency of cocaine or heroin use in the previous month and the likelihood that the person tested positive (table 5). Most clients (87 to 100 percent) who reported using opiates or cocaine on 26 to 30 days in the past month tested positive for the reported drug by hair analysis or urinalysis. The hair tests were much more likely than the urine tests to detect drug use in clients who reported using the drugs less frequently. For example, three times as many people who reported no opiate use in the past 30 days tested positive by RIAH as by urinalysis (30 percent versus 10 percent). The differences were smaller (83 percent versus 49 percent) but in the same direction for those who reported no cocaine use in the past month.

A strong association was also found between the self-reported frequency of drug use and the concentration of drugs found in the hair.[1] The median concentration of opiates in the positive hair specimens was 45 nanograms (ng) per 10 milligrams (mg). However, the average concentration detected varied from 4.4 ng/10 mg for people who reported no use of heroin in the past 30 days to 59.8 ng/10 mg for people

who reported daily use (figure 1). The standard deviations were quite large relative to the means, indicating considerable variation within each group. However, these computations include people who had concentration levels in their hair that were below the detection thresholds routinely used by the laboratory to designate the presence of drugs. If these negative results had been removed, the standard deviations would have been smaller. The median concentration of cocaine metabolite in positive hair specimens was 115 ng/10 mg. Again, the concentration was greatest (404.4 ng/ 10 mg) among self-reported daily users (figure 2).

TABLE 5. *Percentage of clients who tested positive by hair analysis or urinalysis, by self-reported number of days used during the past month.*

	Self-reported number of days used in past month		
	0	1-25	26-30
Tested positive for opiates (N)	(10)	(11)	(85)
Urine	10%	73%	93%
Hair	30%	82%	99%
Tested positive for cocaine (N)	(35)	(48)	(23)
Urine	49%	75%	87%
Hair	83%	98%	100%

Self-Reports and Hair Tests at the Postdischarge Interview

The postdischarge followup study yielded 39 clients who completed the telephone or in-person interview and provided a hair specimen. Eight of these clients had also been included in the intake validity sample. (The remaining 31 clients had been interviewed at intake but not in the same time period as the validity sample.) Table 6 presents characteristics at intake for the intake validity sample and the postdischarge followup sample. The followup sample differed from the intake validity sample with regard to age and past-year use of heroin. They tended to be

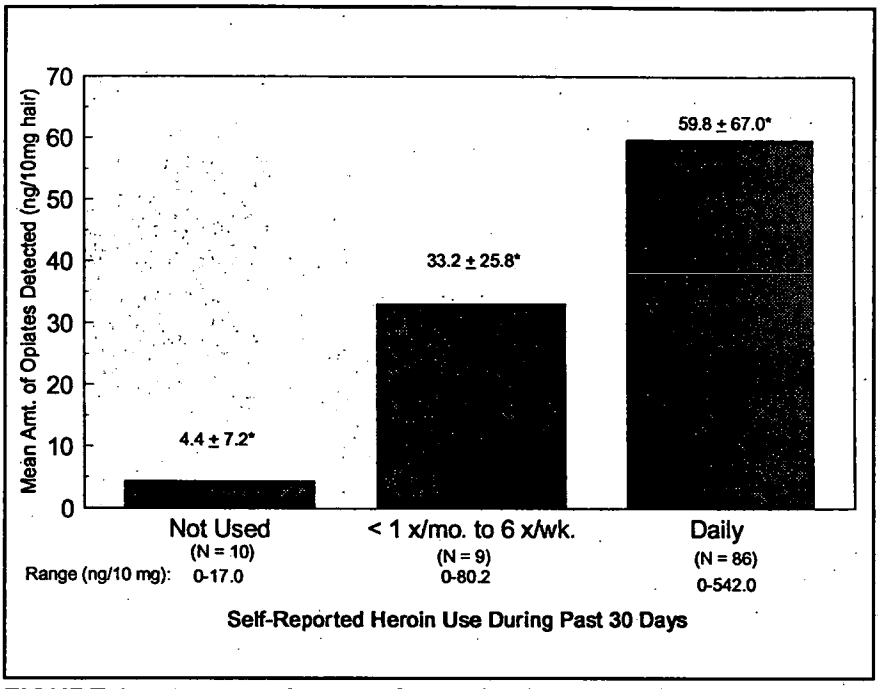


FIGURE 1. *Amount of opiates detected in hair at intake, by self-reported use in the past 30 days (N = 105 clients).*

NOTE: Information was missing for 1 of the 106 clients.
KEY: * = standard deviation.

younger and less likely to have used heroin daily in the year before intake. However, the two samples were similar in terms of education, ethnicity, previous arrest, use of crack/cocaine, and previous drug or alcohol treatment. In view of the similarity of the two groups, differences between them in self-reports and hair tests may reflect differences in how people self-report at intake compared with followup, rather than differences in the composition of the samples.

The followup interview included questions about drug use in the past 90 days that would permit direct comparison with the window of sensitivity of the hair analyses. (Hair specimens had again been cut to a maximum of 3.9 cm, and 72 percent of the sample had hair specimens of this length, representing drug use in the previous 7 to 90 days. The findings indicated considerable differences in estimates of drug use from self-reports and the hair tests. While 62 percent of the followup sample tested positive for opiates by RIAH, only 36 percent reported using

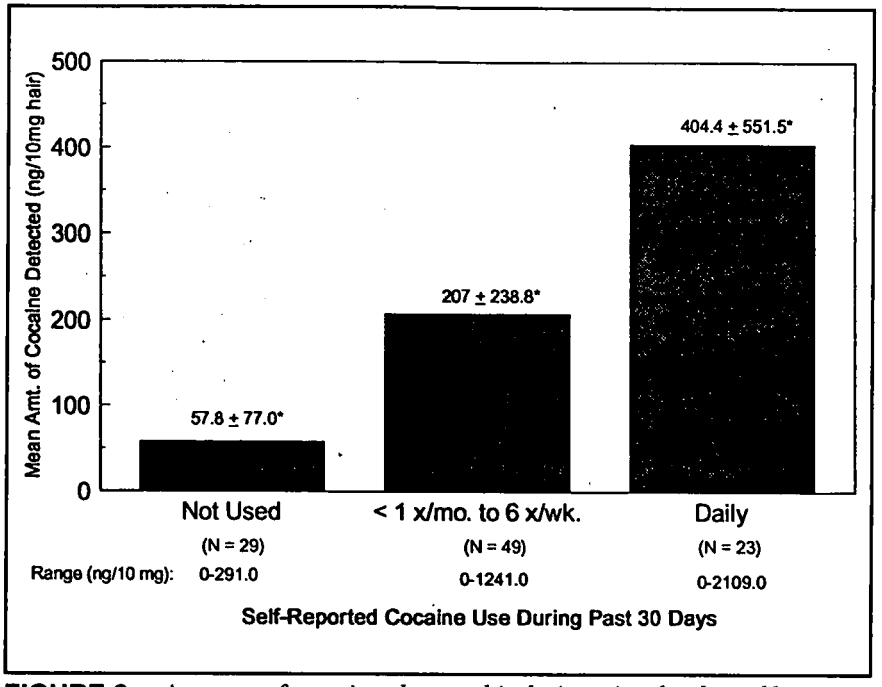


FIGURE 2. *Amount of cocaine detected in hair at intake, by self-reported use in the past 30 days (N = 101 clients).*

NOTE: Information was missing for 5 of the 106 clients.
KEY: * = standard deviation.

opiates in the past 90 days. Similar differences were found with respect to cocaine—80 percent positive by RIAH, 41 percent by self-report. Only about half of the clients who tested positive by hair analysis for opiates (46 percent) or cocaine (52 percent) reported using the drug in the past 90 days. While not exactly comparable, these numbers are considerably below similar analyses of self-reports and hair tests at intake, reported in table 4. At intake, 96 percent of those with a hair test positive for opiates and 75 percent of those positive for cocaine reported using the drug in the past month.

To determine whether the degree of self-reporting at followup was related to the level of use, the followup sample was divided into high or low levels of drug detected in the hair. Clients above the median concentration (31.2 ng/10 mg for opiates and 105.0 ng/10 mg for cocaine) were classified as heavier users of that drug.[1]

TABLE 6. *Characteristics of intake validity sample and followup sample at intake.*

Characteristic (from intake interview)	Intake validity sample[a] (N = 106)	Followup sample[b] (N = 39)
Male	59%	51%
Mean age	39.2*	36.4*
African American	94%	90%
Less than 12-year education	59%	58%
High school diploma/GED	53%	66%
Ever arrested	85%	90%
Used daily in past year		
Cocaine	33%	26%
Crack	15%	21%
Heroin	79%**	57%**
Prior alcohol/drug treatment	84%	71%

KEY: a = Clients who provided hair and urine specimens; b = eight followup clients were also among the 106 clients in the intake validity study sample; * = $p < 0.05$; ** = $p < 0.01$.

While the sample sizes were small, the results show that clients who had the larger concentrations of a drug in their hair were significantly more likely to have reported use of the drug in the past 90 days (figure 3). Approximately three-quarters of clients with the higher concentration of drugs in their hair reported using the detected drug in the past 90 days, compared with one-third or fewer of the persons with less of the drug in their hair.

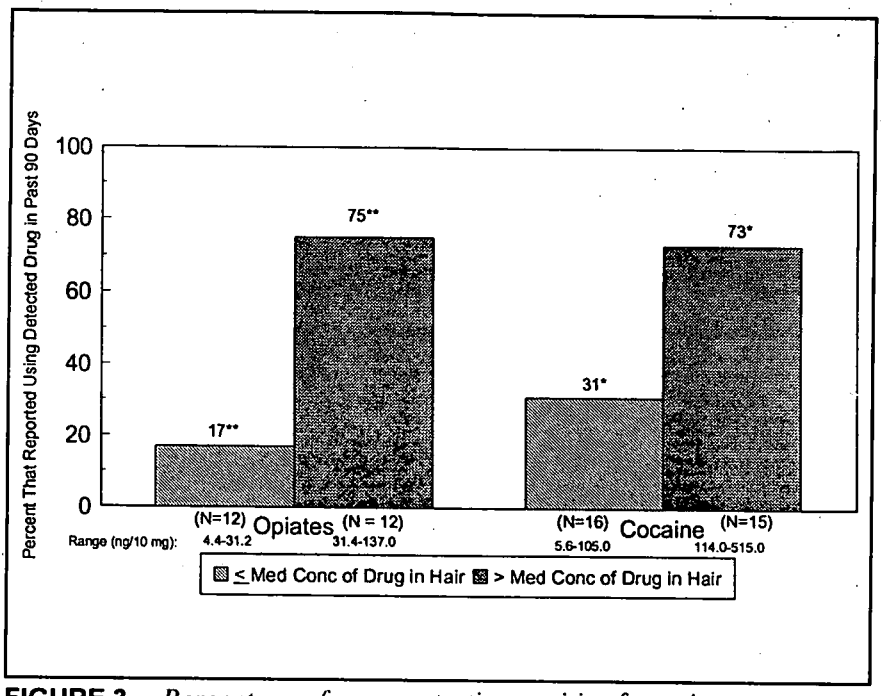


FIGURE 3. *Percentage of persons testing positive for opiates or cocaine at followup who reported using the detected drug in the past 90 days, by amount of drug detected in hair.*

NOTE: Clients were grouped according to whether they fell above or below the median concentration of drug detected among all tested followup clients. Median concentrations were 31.2 ng/10 mg for opiates and 105.0 ng/10 mg for cocaine.

KEY: * = $p < 0.05$; ** $p < 0.01$.

DISCUSSION AND IMPLICATIONS

The results of this study have considerable implications for drug abuse treatment research and for clinical practice. Each of the main themes is discussed below.

- *The validity of client self-reports of drug use may differ by drug.* The overwhelming majority of clients tested positive for opiates by urinalysis or hair analysis at intake, and virtually all reported use of heroin in the previous 30 days. Clients' readiness to report recent heroin use is perhaps not surprising in view of the analyses suggesting that the sample who provided urine and hair specimens

included many who were seeking methadone treatment. The finding that clients who tested positive for cocaine at intake were less likely to report recent use of cocaine is consistent with the findings of Sowder and associates (1993) and with the possibility that heroin users seeking methadone treatment may perceive a disincentive for reporting cocaine use. These findings suggest that discussions of the validity of drug use among drug treatment clients must be framed in the context of the specific drug used.

- *Multiple drug use may go undiagnosed by self-report measures.* The fact that 97 percent of clients who tested positive for opiates by RIAH also tested positive for cocaine has important implications for research as well as clinical management. If clients are relatively less likely to report their recent cocaine use at treatment intake, clinical or research interviews that rely solely on self-reports might underdiagnose multiple drug use. In this study, more than 90 percent of the cocaine-positive clients (by either hair analysis or urinalysis) did report use of cocaine five or more times in their lifetime, even though they denied use in the past month or year. By asking less threatening questions about lifetime drug use, it might be possible to identify clients at risk for current multiple drug use who should receive additional testing or study.

- *Hair analysis detected more cocaine use than did urinalysis.*[1] This finding is consistent with extensive research showing that RIAH's greater window of sensitivity (up to 90 days in this study) leads to the identification of more cocaine users than does urinalysis. Hair analysis did not detect more heroin users in this sample, which contained 79 percent self-reported daily heroin users. With such frequent heroin use, most users can be identified by the 24 to 72 hour sensitivity period of urinalysis.

- *While some clients underreport drug use, their disclosures of extensive drug use may still have substantial validity.* Clients who reported daily use of heroin or cocaine were more likely to test positive for these drugs by urinalysis or RIAH. Self-reported daily users also had the highest concentrations of the reported drug in their hair. These findings suggest that when clients do report extensive drug use, the information is likely to be valid. These findings are consistent with those of Wish (1988), who found that in a sample of people underreporting their recent drug use, those who did report drug dependence had higher rates of

drug-related arrests and expected associations with other correlates of serious drug use.

- *Hair analysis may offer some diagnostic utility.* The finding that daily users of heroin and cocaine had the highest concentrations of drug detected in their hair raises the possibility that hair analysis may be useful in identifying people with the most serious drug abuse problems.[1] As hair analysis techniques are improved, research should be conducted to determine the relationship between quantitative hair test results and clinical and research diagnoses.

- *The validity of self-reports of recent drug use may be less at followup than at intake.* Clients who tested positive for cocaine or heroin were much less likely to self-report use of these drugs at postdischarge followup than at intake. These findings are consistent with those reported for treated (experimental) clients in a program designed to provide methadone to clients while they were waiting to enter the full treatment program (Sowder et al. 1993). The findings are also consistent with the underreporting at treatment followup reported by Hinden and associates (1994) and the RTI (1994). While it is possible that the underreporting found in this study at followup occurred because some followup interviews were conducted over the telephone and only a small number of clients were followed up, much of the underreporting may be the result of the respondents' intention to conceal their current drug use from the researchers. If this is the case, treatment evaluations that compare self-reports of drug use at intake and followup may show reductions in drug use largely as an artifact of the greater underreporting at followup. Until this issue is settled, treatment outcome evaluations that measure drug use solely by self-reports should be interpreted with caution.

- *Underreporting may be less of a problem among the most serious substance abusers.* The fact that about 70 percent of clients with higher concentrations of cocaine or opiates in their hair reported their recent drug use suggests that underreporting may present less of a problem when the goal is to identify the most severe users. Individuals with the greatest drug abuse problems may be most likely to admit their problem in a research or clinical interview. This finding warrants further study and replication by others.

CONCLUSION

The findings reported here contribute to those of other studies that have questioned the validity of self-reports of recent drug use among drug abuse treatment clients. For years researchers have discussed the more obvious determinants of a respondent's willingness to report drug use, including response style, interviewer characteristics, social desirability, and the nature of the interview setting. Researchers must now become sensitive to a host of other factors that may influence a respondent's willingness to report recent illicit drug use, such as: type of drug; whether the person is assigned to a treatment or comparison group; whether the interview occurred at intake, in treatment, or postdischarge; and the severity of the respondent's drug use. Researchers should consider these factors in designing and interpreting treatment outcome studies. Most important is to include toxicologic measures of drug use in all treatment outcome research to validate respondents' self-reports of recent drug use and adjust for underreporting. In the absence of such adjustments, estimates of treatment outcome based on self-reports should be interpreted with caution.

ENDNOTE

1. Refer to the Technical Note at the end of the Introduction (p. 13).

ACKNOWLEDGMENT

This chapter was supported in part by grant U18 DA07082 from the National Institute on Drug Abuse and the Center for Substance Abuse Treatment.

REFERENCES

Appel, P.W. "Substance Abuse Among Adults in Transient Housing in New York State: Validation of Self-Report by Use of Hair Analysis." Albany: New York State Office of Alcoholism and Substance Abuse Services and Research Institute on Addictions, 1995.

Cook, R.F.; Bernstein, A.D.; Arrington, T.L.; Andrews, C.M.; and Marshall, G.A. Methods for assessing drug use prevalence in the workplace: A comparison of self-report, urinalysis and hair analysis. *Int J Addict* 30(4):403-426, 1995.

Dembo, R.; Williams, L.; Wish, E.D.; and Schmeidler, J. *Urine Testing of Detained Juveniles to Identify High-Risk Youth.* National Institute of Justice Research in Brief. Washington, DC: National Institute of Justice, 1990.

Falck, R.; Siegel, H.A.; Forney, M.A.; Wang, J.; and Carlson, R.G. The validity of injection drug users' self-reported use of opiates and cocaine. *J Drug Issues* 22(4):823-832, 1992.

Flynn, P.M.; Hubbard, R.L.; Forsyth, B.H.; Fountain, D.L.; Smith, T.K.; and Hoffman, J.A. "The Individual Assessment Profile (IAP): Standardizing the Assessment of Substance Abusers." Paper presented at the 100th Annual Meeting of the American Psychological Association, Washington, DC, August 15, 1992.

General Accounting Office. *Drug Use Measurement.* Washington, DC: General Accounting Office, 1993.

Gfroerer, J.C., and Hughes, A. Collecting data on illicit drug use by phone. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 277-295.

Harkey, M.R., and Henderson, G.L. "Hair Analysis for Drugs of Abuse: A Critical Review of the Technology." Report submitted to California Department of Alcohol and Drug Programs. Winters, CA: Henderson-Harkey, Inc., 1988.

Harrell, A.V. Validation of self-report: The research record. In: Rouse, B.A.; Kozel, N.J.; and Richards, L.G., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity.* National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985. pp. 12-21.

Hinden, R.; McCusker, J.; Vickers-Lahti, M.; Bigelow, C.; Garfield, F.; and Lewis, B. Radioimmunoassay of hair for determination of cocaine, heroin and marijuana exposure: Comparison with self-report. *Int J Addict* 29(6):771-789, 1994.

Hoffman, J.A.; Schneider, S.J.; Koman J.J., III; Flynn, P.M.; Luckey, J.W.; Cooley, P.C.; Wish, E.D.; and Diesenhaus, H.I. The centralized intake model for drug abuse treatment: The role of computerized data management. *Computers Hum Behav* II(2):215-222, 1995.

Hubbard, R.; Marsden, M.E.; Rachal, J.V.; Harwood, H.J.; Cavanaugh, E.R.; and Ginzburg, H.M. *Drug Abuse Treatment: A National Study of Effectiveness.* Chapel Hill, NC: University of North Carolina Press, 1989.

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

Magura, S.; Goldsmith, D.; Casriel, C.; Goldstein, P.J.; and Lipton, D.S. The validity of methadone clients' self-reported drug use. *Int J Addict* 22(8):727-749, 1987.

Magura, S.; Freeman, R.; Siddiqi, Q.; and Lipton, D.S. The validity of hair analysis for detecting cocaine and heroin use among addicts. *Int J Addict* 27(1):54-69, 1992.

Magura, S.; Kang S.Y.; and Shapiro, J.L. Measuring cocaine use by hair analysis among criminally-involved youth. *J Drug Issues* 25(4):683-701, 1995.

Marques, P.R.; Tippetts, A.S.; and Branch, D.G. Cocaine in the hair of mother-infant pairs: Quantitative analysis and correlations with urine measures and self-report. *Am J Drug Alcohol Abuse* 19(2):159-175, 1993.

McNagy, S.E., and Parker, R.M. High prevalence of recent cocaine use and the unreliability of patient self-report in an inner-city walk-in clinic. *JAMA* 267(8):1106-1108, 1992.

Mieczkowski, T. The accuracy of self-reported drug use: An evaluation and analysis of new data. In: Weisheit, R., ed. *Drugs, Crime and the Criminal Justice System.* Cincinnati, OH: Anderson Publishing, 1990. pp. 275-302.

Mieczkowski, T. New approaches in drug testing: A review of hair analysis. *Ann Am Acad Polit Soc Sci* 521:132-150, 1992.

Mieczkowski, T. An evaluation of patterns of racial bias in hair assays for cocaine: Black and white arrestees compared. *Forensic Sci Int* 63:85-98, 1993.

Mieczkowski, T.; Barzelay, D.; Gropper, B.; and Wish, E.D. Concordance of three measures of cocaine use in an arrestee population: Hair, urine and self-report. *J Psychoactive Drugs* 23(3):241-246, 1991.

Musto, D. Opium, cocaine and marijuana in American history. *Scientific American,* July 1991. pp. 40-47.

New York City Commission on the Homeless. *The Way Home, A New Direction in Social Policy.* New York: New York City Commission on the Homeless, 1992.

Psychemedics Corporation. *Policies and Procedures Manual.* Santa Monica, CA: Psychemedics Corporation, 1991.

Research Triangle Institute. "Early Retrospective Study of Cocaine Treatment Outcomes, Overview and Findings." Draft technical report no. 1, vol. I, submitted to the National Institute on Drug Abuse. Research Triangle Park, NC: Research Triangle Institute, 1994.

Saitoh, M.; Uzuka, M.; and Sakamoto, M. Rate of hair growth. *Adv Biol Skin Hair Growth* 9:183-194, 1967.

Sowder, B.J.; Beschner, G.M.; Ergh, T.; Wish, A.; Wish, E.D.; Zhang, Y.; and Sowder, J. "Waiting List Community Intervention Project." Summary report submitted to the National Institute on Drug Abuse. Bethesda, MD: NOVA Research Corporation, 1993.

Visher, C., and McFadden, K. *A Comparison of Urinalysis Technologies for Drug Testing in Criminal Justice*. National Institute of Justice Research in Action. Washington, DC: National Institute of Justice, 1991.

Weatherby, N.L.; Needle, R.; Cesari, H.; Booth, R.; McCoy, C.B.; Watters, J.K.; Williams, M.; and Chitwood, D.D. Validity of self-reported drug use among injection drug users and crack cocaine users recruited through street outreach. *Eval Program Plan* 17(4):347-355, 1994.

Wish, E.D. Identifying drug-abusing criminals. In: Leukefeld, C.G., and Tims, F.M., eds. *Compulsory Treatment of Drug Abuse: Research and Clinical Practice*. National Institute on Drug Abuse Research Monograph 86. DHHS Pub. No. (ADM)88-1578. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1988. pp. 139-159.

Wish, E.D., and Mieczkowski, T. Comment on Weatherby et al. *Eval Program Plan* 17(4):356-357, 1994.

Wish, E.D., and Gropper, B.A. Drug testing by the criminal justice system: Method, research and applications. In: Wilson, J.Q., and Tonry, M., eds. *Drugs and Crime*. Chicago: University of Chicago Press, 1990. pp. 321-339.

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

AUTHORS

Eric D. Wish, Ph.D.
Director

Susanna Nemes, M.A.
Research Associate

Center for Substance Abuse Research
University of Maryland at College Park
4321 Hartwick Road, Suite 501
College Park, MD 20740

and

Jeffrey A. Hoffman, Ph.D.
Vice President for Operations
Koba Associates, Inc.
1156 15th Street, NW, Suite 200
Washington, DC 20005

The Validity of Self-Reported Cocaine Use in Two High-Risk Populations

Stephen Magura and Sung-Yeon Kang

ABSTRACT

Self-reports of drug use are extensively employed in research on drug use and in evaluations of drug abuse treatment and human immunodeficiency virus (HIV) prevention interventions. The chapter first summarizes recent research addressing the validity of drug use self-reports in high-risk populations. The results of two self-report validity studies are then compared, one for a sample of patients in methadone maintenance and the other for a sample of criminally involved young adults. Cocaine use was more accurately reported by the methadone patients; the possible reasons for this are explored.

INTRODUCTION

There is a continuing need to obtain more valid estimates of illicit drug use, both for the general population and for specific population groups believed to be at high risk for use. All broad-based surveys and the great majority of individual research studies have relied on self-reporting of drug use. Previous research with populations at risk for drug use indicates that the validity of self-reporting varies widely among studies (Magura et al. 1987). Although biological specimens (such as urine, hair, saliva, breath, and blood) can be very useful as objective indicators of drug and alcohol use in epidemiological and other research studies, sole reliance on them is often undesirable. Such specimens may be difficult or impossible to obtain in many studies, and all have inherent (although different) limita-tions in measuring the timing, duration, frequency, and intensity of drug use, as well as the routes of administration and social context of use. For example, the most widely used biological test, urinalysis, provides only reliable indications of heroin or cocaine use within the past 48 hours, but no information on route of administration, although the latter may be essential for assessing degree of dependence or HIV risks (e.g., intranasal, injecting, or smoking). Consequently, it is

important to develop a better understanding of the conditions under which valid self-reports of drug use may be obtained, or their degree of validity under various conditions. Conditions might be identified where valid self-reports are unattainable, which would argue for the need to obtain biological specimens if such research is to be done.

Much of the research on drug use self-reporting has focused on either of two high-risk populations: individuals involved with the drug abuse treatment system or with the criminal justice system. Magura and colleagues (1987) reviewed 13 studies published up to 1985 that examined the validity of drug use self-reports among samples of arrestees and past and present drug abuse treatment clients. Only studies that included comparisons of confidential self-reports with a criterion, usually urinalysis, were included. The mean conditional kappa (K_c) among studies for opiates and cocaine was about 0.5. (K_c measures the degree beyond chance to which self-reports agree with the criterion (Bishop et al. 1975); also see note to table 5.)

The authors concluded that:

> "It is difficult to compare the results of the studies because of differences on such variables [as]...the type of population studied, the type and pattern of drug use, and the measurement procedures and conditions. Even when sample sizes were large enough to permit it, many studies failed to break down their data by treatment modality, present treatment status, or legal status" (Magura et al. 1987, p. 734).

Few of the studies examined possible correlates of drug use underreporting. However, there was some suggestion that higher criminality was associated with such underreporting. Inaccurate reports of drug use were found by Eckermann and associates (1971) to be correlated with severity of arrest charge and by Page and colleagues (1977) with number of prior arrests, although McGlothlin and associates (1977) found no correlation with legal status. In their empirical study accompanying the literature review, Magura and fellow researchers (1987) found a bivariate correlation between self-reported criminality and underreporting of drug use.

Studies published since 1985 have tended to support the tentative hypothesis that addicts not in treatment and having more criminal involvement are less accurate reporters of their illicit drug use than

addicts involved with drug abuse treatment. A review of these recent studies will be presented.

The chapter then compares the results of two studies conducted by the first author on the validity of self-reported cocaine use, one for a sample of patients in methadone treatment and the other for a sample of recently arrested young men. The comparison is of interest because the study methodologies, including the biological criterion of drug use (hair analysis), are identical or similar in all respects; thus the remaining sources of variability are the characteristics of the subject populations. This comparison is intended to provide some insights into the characteristics of subject populations associated with inaccurate reporting of illicit drug use even in a confidential research setting.

Review of Studies in High-Risk Populations Since 1985

The review considers, first, studies of persons involved with drug abuse treatment, and then studies of persons involved with the criminal justice system. Magura and colleagues (1987) compared self-reports of drug use with urinalysis for patients currently in methadone treatment in four clinics in New York City (N = 248); both the self-reports and urine tests were confidential. Among subjects who tested positive for each given drug, 65 percent (24/37) did not report opiate (e.g., heroin) use, 39 percent (36/93) did not report benzodiazepine (e.g., diazepam) use, and 15 percent (10/66) did not report cocaine use. Although opiate use was especially underreported, the current use rate as measured by urinalysis was relatively low (15 percent). There were also subjects who reported drug use not detected by urinalysis (e.g., 27 percent of subjects tested had positive cocaine urines versus 42 percent admitting to its use). This might be attributable to the fact that the time period for self-reporting was 1 month, whereas the urines could detect cocaine or opiate use only during the previous 2 or 3 days.

Twenty-five percent of the sample underreported one or more of the three drugs examined. However, of those who tested positive for multiple drugs, only 11 percent failed to report any of them. In a stepwise log-linear analysis, underreporting was associated with interviewer type (professional) and subject's age (30 years and over). The following variables were not independently associated with underreporting in the log-linear analysis: clinic site, number of medication pickup days, and self-reported current criminality.

Wasserman and coworkers (1993) obtained confidential urines and self-reports of drug use twice weekly for about 10 weeks for patients in four methadone treatment clinics in the San Francisco area. Subjects (N = 81) were told results would not be shared with clinic staff. Overall chance-corrected agreement between self-reports and urinalysis was fair to good, with median kappas of 0.61 for opiates and 0.50 for cocaine. However, the disagreements were almost always positive urinalyses that contradicted negative self-reports. Rates of positive urines were twice as high as rates of self-reported use for some of the time periods; cocaine underreporting was higher than for opiates. No demographic or treatment-related predictors of valid self-reporting were found. A current diagnosis of antisocial personality (based on "Diagnostic and Statistical Manual of Mental Disorders," 3d. ed., rev. (DSM-III-R) interview) predicted increased reporting of cocaine use; the authors suggest that this may indicate a general tendency in some patients to admit to socially disapproved behaviors. In support of this, Magura and associates (1987) found that self-reports of recent crime were associated with self-reports of drug use, but not with positive urinalysis.

Zanis and coworkers (1994) compared confidential self-reports of drug use with weekly clinic urinalysis for a sample of patients (N = 66) in methadone treatment for at least 6 months. Only 13 percent and 17 percent of subjects who tested positive for opiates and cocaine, respectively, failed to report using those drugs in the preceding month. In addition, 58 percent and 28 percent of those with negative urines for opiates and cocaine, respectively, did report using those drugs in the last month. As a result, opiate and cocaine use were self-reported more frequently in confidential interviews than were detected by weekly clinic urinalysis.

Zanis and associates (1994) suggest that reporting may have been very accurate in their study because subjects knew that their self-reports would be compared with urinalysis results. However, this knowledge also existed in the studies by Magura and colleagues (1987) and Wasserman and coworkers (1993), where there was substantial underreporting of drug use in some time periods. In addition, clients in treatment often do not admit drug use to their counselors until confronted with a positive urinalysis, and sometimes not even then. Further, since the confidential self-reports were retrospective for the last month, subjects already knew what their clinic urinalysis results had been, and many nonetheless reported drug use even when their urinalyses had been negative.

Hoffman and fellow researchers (1994) compared self-reports, urinalysis, and radioimmunoassay hair analysis for drug treatment clients. At intake to treatment, similar rates of opiate use were identified by self-report for the past 30 days (87 percent), urine (83 percent), and hair (93 percent). Cocaine use appeared to be slightly underreported at intake: self-reporting (67 percent), urine (67 percent), and hair (90 percent). Because most of the applications were apparently for methadone treatment, it was necessary to report heroin use, but not cocaine use. Nevertheless, three-quarters of those who subsequently tested positive for cocaine by urine or hair voluntarily reported using cocaine in the previous month. At followup after leaving treatment, 42 percent reported opiate use during the previous 90 days, but 70 percent were hair positive for opiates; in addition, 45 percent reported cocaine use, but 88 percent were hair positive for cocaine. Those with above-median hair concentrations of cocaine and opiates were more likely to report use than those with below-median hair concentrations. (The median cocaine concentration was 102 nanograms (ng) per 10 milligrams (mg) in hair and the median opiate concentration was 31 ng/10 mg in hair.)

Falck and coworkers (1992) compared self-reports of drug use with urinalysis for out-of-treatment injecting drug users who had participated in an acquired immunodeficiency syndrome (AIDS) prevention outreach project. Subjects were those reporting either daily use of heroin and/or cocaine, or abstinence from both drugs, at the time of the interview. Of those urine positive for opiates or cocaine, 45 percent denied current use of both drugs. Subjects whose primary drugs of choice were injected cocaine and crack (used with equal frequency) and those who were African American were more likely to underreport heroin/cocaine use. The following variables were not associated with underreporting: age, gender, educational level, treatment history, and project intervention assignment. Baumgartner and associates (1989) compared self-reports of drug use with results of urinalysis and hair analysis for a community supervision sample (probationers and parolees). Among those hair positive for cocaine at intake to community supervision, 29 percent failed to report recent use (past 90 days) in confidential interviews. Among those hair positive for morphine, 19 percent failed to report recent heroin or other opiate use.

Mieczkowski and colleagues (1991) compared self-reports of cocaine use with urinalysis and hair analysis for arrestees booked into the Pinellas County (Florida) Jail. The interviews were conducted by a specially trained social worker; all interview and test data were

confidential. Of those with positive urinalysis, 76 percent denied cocaine use in the preceding 48 hours. Of those with positive radioimmunoassay hair analysis for 1 month of hair growth, 72 percent denied cocaine use in the preceding month. Results for heroin (opiate) use indicated even greater underreporting; there were nearly nine times more positive hair test results for opiates than self-reported use within the previous month (8.9 percent versus 1.0 percent) (Mieczkowski et al. 1993).[1] This study did not investigate possible factors associated with underreporting of drug use.

Confidential self-reports of drug use were compared with urinalysis for samples of felony arrestees in 14 U.S. cities participating in the Drug Use Forecasting (DUF) system (National Institute of Justice 1990). Estimates of cocaine use based on urine tests were about twice as high as those based on self-reports. Estimates of heroin use based on urine tests were about 1.5 times as high as those based on self-reports. Feucht and colleagues (1994) found that only 12 percent (6/50) of juvenile arrestees in Cleveland who were found cocaine positive by hair analysis reported recent cocaine use on confidential interviews.[1]

Underreporting of illicit drug use, even under conditions of research confidentiality, continues to be a problem for research studies. The greatest underreporting appears to be within criminal justice populations (as compared with noncriminal justice populations) (Mieczkowski et al. 1991, 1993; National Institute of Justice 1990; Feucht et al. 1994) and for addicts out of treatment (as compared with those in treatment) (Falck et al. 1992; followup sample in Hoffman et al. 1994), although one study of community corrections clients found only minimal underreporting (Baumgartner et al. 1989). Studies of addicts involved with drug abuse treatment have found small to moderate amounts of underreporting, although the results often vary by type of drug. In general, it remains very difficult to compare studies because of the myriad differences in study populations, methodologies, and time periods. A new issue in this regard is the recent development of hair analysis, which has a wider window of detection than urinalysis and thus increases the likelihood of detecting drug use underreporting. Finally, despite continuing research, there has been little explicit attention to factors that may be associated with underreporting in specific high-risk populations.

METHODS

Two studies were conducted recently to examine the validity of cocaine use self-reports in two different populations at high risk for use: methadone maintenance patients and criminally involved, young adult males.

Study 1 Sample

The subjects were 134 patients in two methadone maintenance treatment programs in New York City during 1988-89; they were part of a larger study of the cocaine problem in methadone treatment. Patients were randomly selected from each clinic's census for the study; the sampling was stratified to overrepresent patients with recent cocaine-positive clinic urinalysis. The cocaine-using subjects may not be representative of all cocaine-using methadone patients in New York City. The subjects were interviewed about drug use and other topics and provided hair and urine specimens to the researchers (Magura et al. 1992).

Study 2 Sample

The subjects were 121 male young adults who were originally recruited while in jail in New York City and were followed up in the community during 1992-93, at a median of 5 months after their release from jail. The subjects volunteered for the study while in jail and may not be representative of all young adult jail inmates in New York City. The subjects were interviewed both in jail and later in the community about drug use and other topics, and provided a hair specimen at the time of the community interview (Magura et al. 1995).

Study Procedures

In both studies, subjects were informed that participation was voluntary and all interview and testing information was confidential. Subjects gave written informed consent consistent with Federal human subjects regulations for drug treatment clients, prisoners, and minors in custody. The study was approved by the investigators' institutional review board and the data were protected by a Federal Certificate of Confidentiality. In the context of a 90-minute interview, subjects were asked to report whether they ever tried or used cocaine in any form, and the time of their most recent use. Subjects received a $30 incentive. The interviews were

conducted by trained paraprofessionals who were themselves in recovery.

Hair Sampling and Analysis Methods

A specimen of hair of up to 50 strands was cut at scalp level from each subject. Radioimmunoassay tests for cocaine were conducted on either 1.3 centimeter (cm) or 3.6-cm hair segments as measured from the scalp, each 1.3 cm corresponding to about 1 month of hair growth. Hair analysis for the presence of cocaine was performed by a toxicology laboratory specializing in such assays. The tests were blinded (i.e., the laboratory did not know whether drug use was reported by the subjects). One milligram of hair was washed for 15 minutes in ethanol and then three times for 30 minutes each in water with phosphate buffer. The specimen was then enzymatically digested, the melanin fraction removed, and tested for cocaine and its metabolites by radioimmunoassay. Removing the melanin fraction is intended to prevent differences in hair color from affecting the results. Separate results for cocaine and its metabolites are not yielded by the analytical method used.

Aliquots of the wash solution were also tested and wash kinetic curves from such data were compared to the cocaine levels in the hair digest. Three wash kinetic criteria based on hair from known cocaine users have been developed to distinguish between drug use and external contamination of the hair (Baumgartner and Hill 1993). For example, the first criterion requires that the ratio of the amount of cocaine in the digest to the amount of cocaine in the last wash exceed a value of 10. Specimens are regarded as indicating ingestion of cocaine rather than simply the residue of possible external contamination only when all three wash criteria cutoffs are passed. (Note, however, that there are possible, if unlikely, contamination scenarios that may elude detection by these criteria.)

Hair assay results are given in standardized units (i.e., ng of cocaine (including metabolites) per 10 mg of hair). The cutoff for defining cocaine positives was 2 ng/10 mg.

RESULTS

Sociodemographic characteristics of the two study samples are presented in tables 1 and 2. For the methadone sample, 60 percent self-reported

recent cocaine use and 80 percent were hair positive for cocaine. For the young adult sample, 23 percent self-reported recent cocaine use and 67 percent were hair positive for cocaine. Quantitative levels of cocaine in hair are given in table 3. Considering only the positives, it is clear that the methadone patient sample (median = 125 ng) has a higher hair concentration of cocaine than the young adult sample (median = 17 ng). The difference is particularly large at the greatest concentrations (above 300 ng). Although one must still be cautious about quantitative interpretations of hair analysis, this may indicate more use of cocaine by the methadone patients.[1]

Hair analysis for cocaine was compared with self-reports of cocaine use within each sample (table 4). If a 1.3-cm or shorter hair segment was analyzed, self-reports of use in the last 4 to 6 weeks were compared. (Although 4 weeks should be sufficient to parallel the window of detection of a 1.3-cm hair segment, up to 6 weeks was used because subjects have difficulty reporting within exact time intervals; see also Ehrman and Robbins 1994.) If a 3.6-cm hair segment was analyzed, self-reports of use in the last 3 to 4 months were compared. Forty-six percent of the young adults' hair specimens were less than 1.3 cm long. Because drug use reports were not obtained for a time period of less than the previous month (4 weeks), the reports of subjects with hair less than 1.3 cm cover more time than the hair analysis. It is possible, though unlikely, that this could lead to more positive self-reports than positive hair analyses (i.e., if subjects used cocaine just before, but not within, the window of detection of their hair specimen).

There is moderate agreement beyond chance between self-reports and hair analysis for the methadone patients (kappa = 0.45), and no agreement beyond chance (kappa = 0.00) for the criminally involved young adults.

The most striking finding is that 73 percent of the methadone patients whose hair was positive admitted recent cocaine use, whereas only 23 percent of criminally involved young adults whose hair was positive did so. This is despite the fact that the young adults' self-reports cover more time than their hair analyses in about half the cases.

Since the median levels of cocaine in the hair were very different between the two samples, could this help account for the differences in underreporting? There was a statistically significant and strong association between hair level of cocaine and self-report of cocaine use among the methadone patients, but no significant association for the

young adults (table 5). Even among young adults with the highest concentrations of cocaine (>100 ng), only 29 percent reported any recent cocaine use.[1]

TABLE 1. *Characteristics (percent) of methadone patients (N = 134).*

Gender	
Male	74%
Female	26
Age	
Under 30 years	16%
30 - 34 years	26
35 - 39 years	32
40 and older	26
Ethnicity	
African American	24%
Hispanic	59
White	16
Other	1
Marital status	
Single, never married	48%
Married	21
Other	31
Education	
Less than high school	52%
HS graduate or GED	25
Some college	23
Employed (full- or part-time)	31%
Time in current program	
Under 12 months	37%
12 - 23 months	15
24 - 35 months	11
36 months and over	37

Some further investigation is possible and would lead to a better understanding of the apparent underreporting of cocaine use by the young adults. For example, how reliable are their reports of lifetime cocaine use? They were asked on both the in-jail and subsequent followup interviews whether they had ever tried or used cocaine or crack in their lives. As shown in table 6, the reliability of lifetime reports of cocaine/crack use was high. Self-reports of cocaine use in the 90 days before arrest and in the 90 days before the followup interview were also compared. Although cocaine use potentially could vary between time periods, there was a strong association between reports of use in the two periods.

TABLE 2. *Characteristics (percent) of criminally involved young adult males (N = 121).*

Age	
17	5%
18	13
19	41
20	31
21 and over	10
Ethnicity	
African American	63%
Hispanic	35
Other	2
Arrest charge(s)	
Violence[a]	46%
Property	12
Drugs[b]	32
Weapons	12
Prior arrests[c]	
None	40%
One to four	35
Five or more	25
Drug dealing - ever[d]	75%
Drug dealing - last month[d]	41%
Arrested since release - yes[e]	23%
Education since release - yes[f]	29%
Employed - last month[g]	38%

NOTE: N = 121 except where noted below.

KEY: a = Includes homicide, attempted homicide, robbery, assault, rape, arson; b = Includes dealing and possession; c = Prior to the instant arrest leading to incarceration; d = Over 90% of drug dealing involved cocaine or crack. "Last month" is month before the followup interview; e = Based on 84 subjects with responses; f = Includes attending high school or GED classes at followup; graduated high school or received GED since release. Based on 119 subjects with responses; g = Full- or part-time legitimate employment at followup. Based on 110 subjects with responses.

Cocaine in hair was associated in expected directions with certain other variables, as suggested by past research on drug use and crime (e.g., Jessor and Jessor 1977; Elliot et al. 1985; Fagan et al. 1990; Dembo et al. 1993).

TABLE 3. *Levels of cocaine in hair for two samples (in ng/10 mg).*

	Percent methadone patients	Percent criminally involved young adults
Negative (0.0 to 1.9)	20	33
2.0 to 5.0	12	16
5.1 to 30.0	10	29
30.1 to 100.0	16	11
100.1 to 300.0	13	7
300.1 and over	30	5
	101%	101%
	(N = 134)	(N = 121)

NOTE: Totals do not sum to 100 percent due to rounding.

TABLE 4. *Cocaine self-reports by cocaine in hair for two samples.*

		Percent methadone patients		Percent criminally involved young adults	
Cocaine in hair		No	Yes	No	Yes
Cocaine self-report	No	89	27	77	77
	Yes	11	73	23	23
		100%	100%	100%	100%
		(N=27)	(N=107)	(N=39)	(N=77)
		K = 0.45		K = 0.00	

NOTE: Cohen's Kappa (K) measures the degree of agreement between two classificatory variables (Bishop et al. 1975). Perfect agreement (i.e., all cases classified identically on both variables) is indicated by K = 1 and chance agreement only by K = 0.

TABLE 5. *Cocaine self-reports by level of cocaine in hair for two samples.*

	Methadone patients			Criminally involved young adults		
Cocaine in hair (ng/10 mg)	0.0	2.0 to 9.9	≥10.0	0.0	2.0 to 9.9	≥10.0
Cocaine self-reports						
No	89%	74%	17%	77%	87%	70%
Yes	11%	26%	83%	23%	13%	30%
	100%	100%	100%	100%	100%	100%
	(N = 27)	(N = 19)	(N = 88)	(N = 39)	(N = 31)	(N = 46)

TABLE 6. *Self-reported cocaine use in jail (T_1) versus followup in community (T_2).*

Ever tried cocaine (T_2)		No	Yes
Ever tried cocaine (T_1)	No	77	4
	Yes	5	35
		K = 0.83	

Cocaine use—Last 90 days (T_2)		No	Yes
Cocaine use—Last 90 days (T_1)	No	82	7
	Yes	7	20
		K = 0.66	

Cocaine was more likely to be present, and present at higher levels, in the hair of young adults who had higher numbers of previous arrests (a trend at $p = 0.08$), who were rearrested after release from jail ($p < 0.05$), who failed to continue their education after release from jail ($p < 0.01$),

and who were not engaged in legitimate employment at followup ($p < 0.001$). Young adults who had a pending court case or were on probation or parole were less likely to have cocaine in their hair (a trend at $p = 0.10$).

There were no associations between self-reported cocaine use and any of the above variables (i.e., previous arrests, arrest since release, education after release, employment at followup, and legal status at followup); this is inconsistent with earlier research cited above.

Self-reported cocaine use was associated with respondent ethnicity; 39 percent of Hispanics/others reported cocaine use versus 15 percent of African Americans ($p < 0.01$). Cocaine in hair was not associated with ethnicity.

Self-reported cocaine use was associated with self-reported drug dealing at followup; 36 percent of those denying cocaine use report dealing versus 59 percent of those admitting cocaine use also report dealing ($p < 0.05$). Cocaine in hair was not associated with self-reported drug dealing.

DISCUSSION

In personal interviews in the community, 23 percent of the young adults self-reported some form of cocaine use within the preceding 90 days. Analysis of hair specimens indicated that some amount of cocaine was probably ingested by 67 percent of the young adults within that time period. This latter rate, although high, still might underestimate cocaine use because most of the hair specimens were too short to provide for a full 90-day window of detection. In any event, hair analysis yielded a cocaine use prevalence rate 2.9 times as high as that indicated by confidential self-reports for this population (23 percent versus 67 percent, respectively).

Reliability of reporting was high; young adults who admitted or denied lifetime cocaine use during in-jail interviews were very likely to give the same answers on interviews conducted an average of 1 year later. Similarly, young adults interviewed in jail who reported using cocaine in the 90 days before their arrest were very likely to report using cocaine in the 90 days before their community interview. These patterns would not appear if young adults were answering the cocaine questions randomly.

Nevertheless, it appears that many young adults who are using cocaine fail to report it, and fail to report it consistently.

The zero association between cocaine self-reports and cocaine in hair for the young adults is due partly to the nine who self-reported cocaine use, but whose hair was negative. These nine hair specimens averaged 0.8 cm in length, corresponding to a window of detection of about 3 weeks. Five of these young adults reported currently using cocaine from 2 to 5 days per week, so that the drug was potentially detectable even in these short hair lengths. It may be that these young adults used substances they believed or were told to be cocaine or crack, but actually were not. Misidentification and misrepresentation of substances certainly occurs in the drug subculture. It may also be that the hair analysis failed to detect cocaine in the hair of these subjects. The authors cannot explain this finding at this time.

The substantial underreporting of cocaine use is also indicated by the relatively low percentage of subjects (32 percent at followup) who admit to ever trying cocaine or crack in their lifetimes (table 6); this seems unrealistic for inner-city, young-adult males in the context of the current cocaine/crack epidemic.

The study showed that the number of past arrests as well as rearrests were associated with the presence of cocaine in hair. Moreover, young adults who continued their educations or were legitimately employed at followup were less likely to test positive for cocaine in their hair. While causal inferences cannot be made from this study, it appears that these hair assay results are usually consistent with the findings of previous research on the associations among drug use, criminality, and prosocial behavior. In contrast, self-reports of cocaine use were not associated with the above variables. However, self-reported cocaine use was associated with self-reported drug dealing, seemingly a congruent result, while admitted drug dealing was not associated with cocaine in hair. Possibly subjects who are willing to admit to a serious illegal activity are also more likely to admit to cocaine use.

Young adults with an open legal case tended to be less likely to test positive for cocaine in their hair. These young adults may be influenced by potential sanctions if they are identified as drug users in their contacts with criminal justice personnel (i.e., through drug tests or the appearance of use).

What could explain the large observed difference in the validity of cocaine self-reporting between the two samples? The research conditions were very similar: confidentiality was assured, the interview questions were similar, the interviews were conducted by the same type of staff (minority persons in recovery), and the studies were conducted in the same city. However, the subject populations were very different. Some pertinent differences that may explain differential self-reporting are: the methadone patients were much older than the young adults (mean 35 years versus 19 years); the methadone patients were an identified addict population and in treatment, whereas the young adults were not; some methadone patients may have had previous participation in drug research, or heard about such research, whereas the young adults probably had not; and the young adults were all current or recent criminal justice clients, whereas the methadone patients generally were not. The latter, when combined with a lack of previous exposure to social research, could have resulted in greater reluctance to admit drug use to persons who were not well known to the young adults.

The apparent underreporting of recent and lifetime cocaine use in the young adult group might be attributable to several factors. These criminally involved young adults might have been suspicious or fearful of the research interviewers, even though the interviewers were indigenous to the community and previously had interviewed the young adults in jail without untoward consequences. However, the fear explanation does not seem consistent with the young adults' far more frequent willingness to admit use of another illegal substance, marijuana (60 percent in the last month), and an illegal activity, involvement in drug (mainly crack) dealing (75 percent lifetime and 41 percent in the last month).

A second explanation for underreporting is that cocaine use, and especially crack use, is highly stigmatized in these young adults' reference groups (see also Dunlap and Johnson 1992; Hamid 1992), a view that is supported by interviews conducted with the young adults while they were still in jail. This stigmatization of personal use, however, does not extend to making money from selling crack. Thus many of the young adults might have been reluctant to admit using cocaine or crack in order to project a more favorable image of themselves to the interviewers, whereas crack dealing is a lucrative albeit illegal activity that suggests the young adults' enterprise and self-reliance (Inciardi and Pottieger 1991).

A third possibility is that cocaine use is rather infrequent for many of these young adults and, consequently, they do not define themselves as

using the substance at all. Although the researchers emphasized to the subjects that they were interested in recording "one-time use" or trying "just one time," this may not have been sufficient to elicit accurate self-reports. The relatively low levels of cocaine found in their hair as compared with the methadone treatment sample is consistent with the idea that many of these young adults use the drug infrequently or lightly.[1] Infrequent use may also lead to poor recall.

The finding that African Americans are more likely than other groups to underreport drug use has been noted by several previous studies, including surveys of the general population (e.g., Page et al. 1977; Falck et al. 1992; Fendrich and Vaughn 1994). It has been suggested that matching interviewer and respondents by race/ethnicity might avoid this problem (Campbell 1981). However, both interviewers in the young adult study were African American, and apparently elicited more valid reporting from Hispanics than from African Americans. But note that the interviewers were considerably older (in their forties) than the young adults.

The study findings indicate that accurate estimates of cocaine/crack use among criminally involved, inner-city young male adults cannot rely solely on self-reports, even when obtained under conditions of confidentiality by street-wise indigenous interviewers for research purposes only. The apparent differential accuracy of self-reporting by ethnicity is an added complication in interpreting such self-reports.

The study found that drug use is reported with moderate accuracy to researchers by clients in treatment. Thus self-reports in this population could be useful in providing basic information about patterns of drug use (frequency, intensity, routes of administration) to supplement prevalence data, assuming of course that those who do and do not report their use exhibit similar patterns.

In conclusion, one must be cautious about offering general statements about the validity of drug use self-reports. The degree of accuracy obtained clearly depends upon the specific research conditions and, as this chapter suggests, the characteristics of the populations studied. More methodological research must be conducted on ways to improve the validity of drug use self-reporting in certain populations, particularly high-risk young adults and criminal justice populations. Also, research is needed to better understand the factors that lead to relatively more accurate reporting, such as that shown by methadone patients.

ENDNOTE

1. Refer to the Technical Note at the end of the Introduction (p. 13).

REFERENCES

Baumgartner, W.A., and Hill, V.A. Sample preparation techniques. *Forensic Sci Int* 63 (7-8):121-135, 1993.

Baumgartner, W.A.; Hill, V.A.; and Blahd, W.H. Hair analysis for drugs of abuse. *J Forensic Sci* 34(6):1433-1453, 1989.

Bishop, Y.M.M.; Fienberg, S.E.; and Holland, P.W. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press, 1975.

Campbell, B.A. Race-of-interviewer effects among southern adolescents. *Public Opin Q* 45:231-44, 1981.

Dembo, R.; Williams, L.; Schmeidler, J.; and Christensen, C. Recidivism in a cohort of juvenile detainees: A 3 ½- year followup. *Int J Addict* 28(7):631-658, 1993.

Dunlap, E., and Johnson, B.D. The setting for the crack era: Macro forces, micro consequences 1960-92. *J Psychoactive Drugs* 24:307-322, 1992.

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

Ehrman, R.N., and Robbins, S.J. Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. *J Consult Clin Psychol* 62(4):843-850, 1994.

Elliott, D.S.; Huizinga, D.; and Ageton, S. *Explaining Delinquency and Drug Abuse*. Beverly Hills, CA: Sage, 1985.

Fagan, J.; Weis, J.G.; and Cheng, Y.T. Delinquency and substance use among inner-city students. *J Drug Issues* (Summer):351-402, 1990.

Falck, R.; Siegel, H.A.; Forney, M.A.; Wang, J.; and Carlson, R.G. The validity of injection drug users' self-reported use of opiates and cocaine. *J Drug Issues* 22(4):823-832, 1992.

Fendrich, M., and Vaughn, C.M. Diminished lifetime substance use over time: An inquiry into differential underreporting. *Public Opin Q* 58:96-123, 1994.

Feucht, T.E.; Stephens, R.C.; and Walker, M.L. Drug use among juvenile arrestees: A comparison of self-report, urinalysis and hair assay. *J Drug Issues* 24(1):99-116, 1994.

Hamid, A. The development cycle of a drug epidemic: The cocaine smoking epidemic of 1981-1991. *J Psychoactive Drugs* 24:337-348, 1992.

Hoffman, J.A.; Wish, E.D.; Koman, J.J.; and Flynn, P.M. "Self-Reported Drug Use Compared with Hair Analysis and Urinalysis." Paper presented at the College on Problems of Drug Dependence 56th Annual Scientific Meeting, Palm Beach, Florida, June 18-23, 1994.

Inciardi, J.A., and Pottieger, A.E. Kids, crack and crime. *J Drug Issues* 21(2):257-270, 1991.

Jessor, R., and Jessor, S.L. *Problem Behaviors and Psychosocial Development: A Longitudinal Study of Youth.* New York: Academic Press, 1977.

Magura, S.; Freeman, R.; Siddiqi, Q.; and Lipton, D.S. The validity of hair analysis for detecting cocaine and heroin use among addicts. *Int J Addict* 27(1):54-69, 1992.

Magura, S.; Goldsmith, D.; Casriel, C.; Goldstein, P.J.; and Lipton, D.S. The validity of methadone clients' self reported drug use. *Int J Addict* 22(8):727-749, 1987.

Magura, S.; Kang, S.Y.; and Shapiro, J.L. Measuring cocaine use by hair analysis among criminally-involved youth. *J Drug Issues* 25(4):683-701, 1995.

McGlothlin, W.H.; Anglin, M.D.; and Wilson, B.D. *An Evaluation of the California Civil Addict Program.* Rockville, MD: National Institute on Drug Abuse, 1977.

Mieczkowski, T.; Barzelay, D.; Gropper, B.; and Wish, E. Concordance of three measures of cocaine use in an arrestee population: Hair, urine and self-report. *J Psychoactive Drugs* 23(3):241-246, 1991.

Mieczkowski, T.; Landress, H.J.; Newel, R.; and Coletti, S.D. Testing hair for illicit drug use. In: *Research in Brief.* Washington, DC: National Institute of Justice, U.S. Dept. of Justice, January, 1993.

National Institute of Justice. *Drugs and Crime in America. 1988 Drug Use Forecasting Annual Report.* Washington, DC: National Institute of Justice, 1990.

Page, W.F.; Davies, J.E.; Ladner, R.A.; Alfasso, J.; and Tennis, H. Urinalysis screened versus verbally reported drug use: The identification of discrepant groups. *Int J Addict* 12(4):439-450, 1977.

Wasserman, D.A.; Havassy, B.E.; Weinstein, M.A.; and Hall, S.M. "Validity of Self-Reports of Heroin and Cocaine Use in Methadone Maintenance Patients: Data from Repeated Assessments." Paper presented at the College on Problems of Drug Dependence 55th Annual Scientific Meeting, Toronto, Canada, June 1993.

Zanis, D.A.; McLellan, A.T.; and Randall, M. Can you trust patient self-reports of drug use during treatment? *Drug Alcohol Depend* 35:127-132, 1994.

ACKNOWLEDGMENTS

This work was supported by National Institute on Drug Abuse grants 1 R01 DA-05942 and 1 R01 DA-03991. The data were collected through the cooperation of the New York City Department of Correction, Beth Israel Medical Center-Methadone Maintenance Treatment Program (St. Luke's Clinic), and Albert Einstein College of Medicine-Van Etten Drug Treatment Program.

AUTHORS

Stephen Magura, Ph.D., C.S.W.
Director
Institute for Treatment and Services Research

Sung-Yeon Kang, Ph.D.
Principal Investigator

National Development and Research Institutes, Inc.
11 Beach Street
New York, NY 10013

Assessing Drug Use in the Workplace: A Comparison of Self-Report, Urinalysis, and Hair Analysis

Royer F. Cook, Alan D. Bernstein, and Christine M. Andrews

ABSTRACT

A random sample of 1,200 employees of a steel plant in the western United States was randomly assigned to four different self-report methods of assessing illicit drug use: individual interview in the workplace, group-administered questionnaire in the workplace, telephone interview, and individual interview off the worksite. Urine specimens were collected and analyzed on all 928 subjects participating in the study, and hair analysis was conducted on 307 of the subjects. Although self-reports produced higher prevalence rates than the chemical tests, analyses combining the results of the three assessment methods showed that the actual prevalence rate was approximately 50 percent higher than the estimate produced by self-reports alone. The group-administered questionnaire method produced prevalence rates that were roughly half those of the other self-report methods. The findings cast doubt on the validity of self-reports as means of estimating drug use prevalence and suggest the need for multiple assessment methods.

INTRODUCTION

Working adults constitute a large proportion of the users of illicit drugs, particularly workers between 18 and 34 years of age. In the most recent National Household Survey on Drug Abuse for which employment data are available, 13.1 percent of full-time employees reported illicit drug

---

A version of this article was originally published in the International Journal of the Addictions, *Marcel Dekker, Inc., Publisher. Copyright is held by Marcel Dekker, Inc. The authors and the National Institute on Drug Abuse would like to express their appreciation to Marcel Dekker, Inc., for granting permission to publish this material.*

use in the past year (National Institute on Drug Abuse (NIDA) 1993). Within the 18- to 25-year-old group, 26.9 percent of those employed full-time reported illicit drug use in the past year, and among 26- to 34-year-olds, the prevalence rate was 17.7 percent. Drug use among workers has been linked to increased absenteeism (Normand et al. 1990), higher accident rates (CONSAD 1989; Crouch et al. 1989), more costly use of medical benefits (Winkler and Sheridan 1989), and job withdrawal (Lehman and Simpson 1992).

Researchers with interests in exploring issues of illicit drug use in the workplace have long been concerned about the validity of self-reported drug use. Two decades ago, research was conducted on drug use prevalence assessment methods in organizational settings by comparing self-reports to urinalysis data in the military (Cook et al. 1976; Hurst et al. 1975). Although those early studies generally supported the validity of self-reports, more recent research has cast considerable doubt on a worker's willingness to disclose drug use, despite assurances of confidentiality and anonymity (Cook 1989). Chemical testing methods, particularly urinalysis, have also been used to estimate drug use prevalence (Anglin and Westland 1989). Self-reports and chemical testing methods would appear to offer contrasting strengths and weaknesses as prevalence assessment techniques. Self-reports offer the capability of producing data that are rich with information on frequencies, patterns, and consequences, but they are extremely susceptible to threats to validity. On the other hand, the basic validity of urinalysis is rarely disputed, despite continuing concerns about accuracy (Blanton et al. 1992). However, urinalysis typically provides only a single datum (i.e., whether the individual has recently used a drug). The vulnerability of self-reports to underreporting biases seems exacerbated in the workplace, where workers may fear that admission of illicit drug use could result in disciplinary actions or even job loss. However, as recently noted, despite continued research on workplace drug use, "very little data are currently available for assessing the validity of self-report substance use measures within organizations in populations not otherwise identified as drug users" (Lehman and Simpson 1992, p. 310).

On a broader level, new concerns about the general validity of self-reports of drug use have recently been voiced. Both the National Household Survey on Drug Abuse and the Monitoring the Future survey—perhaps the foremost national indicators of drug use—have been criticized by the General Accounting Office (GAO) for their reliance on self-reports of drug use, and the GAO has recommended the

use of hair analysis in a limited field trial to study "the general level of agreement between self-reports and hair analysis in anonymous survey situations" (GAO 1993, p. 59).

Although the technology of hair analysis is still in its relative infancy, it offers the prospect of a biological indicator that is potentially as accurate as urinalysis, but that also provides a wider detection period, one that is limited only by the length of the hair sample (Baumgartner et al. 1989). An inch of hair typically contains a record of approximately 2 months of potential drug use. Although a variety of criticisms have been leveled at hair analysis, recent tests of its validity with addicts and arrestees have resulted in qualified support for the validity and utility of the technique (Magura et al. 1992; Mieczkowski et al. 1993).[1] To date, there has been no research on the use of hair analysis as a method for assessing drug use in the workplace.

The currrent study had multiple objectives. Its original purpose was to compare different techniques of self-report to each other and to urinalysis as methods for assessing illicit drug use in the workplace. Workers were randomly assigned to four different modes of self-report, and were also assessed by urinalysis. In a second phase, an additional sample of workers was assigned to two of the self-report conditions, and both urinalysis and hair analysis were conducted on all subjects. Preliminary findings from the first phase were previously published as a research note by Cook and Bernstein (1994). This chapter presents results for both phases of the research.

METHODS

Design

The study was conducted in two phases. In the first phase, 800 employees of a large steel plant were randomly selected (using simple random sampling) from a workforce of approximately 2,400 total employees and randomly assigned to one of four conditions of self-report: (1) individual interviews in the workplace, (2) questionnaire administration in small groups, (3) telephone interviews, and (4) individual interviews off the worksite. Urine specimens were collected and analyzed on all subjects. In the second phase, another 400 employees were randomly selected and randomly assigned to two conditions of self-report: (5) individual interview in the workplace, and (6) questionnaire administration in small

groups. In these two conditions, both urine specimens and hair samples were collected on all subjects.

Pilot tests of the data-collection procedures were conducted in the fall of 1990; the first phase was conducted in 1991 and the second phase in 1992. This steel plant was selected for study mainly because its workforce was sufficiently large and varied, and also had a considerable proportion of young, blue-collar male empolyees, among whom the use of alcohol and illicit drugs is especially concentrated (Cook 1989).

The results from hair analysis, urinalysis, and four different self-report techniques were compared to each other. Preliminary findings from the first phase were reported previously by Cook and Bernstein (1994). Of the 1,058 employees available for participation, a total of 928 agreed to participate.

Subjects

All eligible subjects were asked to report information about their age, gender, ethnicity, and other demographic variables after responding to questions measuring their drug use. As shown in table 1, the vast majority of the subjects were white males, most of whom were married and between the ages of 18 and 54. Nearly all subjects were high school graduates, and about half reported some amount of college education. Most subjects reported annual salaries between $30,000 and $50,000.

Procedures

Generous incentives were offered to bolster participation rates. The selected employees were notified that they would be paid $5 just to attend a recruitment session, and $15 if they agreed to participate in the research. By participating, they would also be eligible for a raffle cash prize of $1,000. The interviewers emphasized that the data collection was anonymous and confidential. Matching code numbers (no names) were placed on the questionnaires and specimen containers. The interviewers explained that the research was being conducted by a private research firm; that no one but the research team would know their answers; and that no one would be informed if there were a positive result of the chemical tests. The fact that pilot tests involving approximately 25 employees were conducted several months before main data collection without any negative effects to participants probably enhanced the credibility of the confidentiality assurances.

TABLE 1. *Background characteristics of subjects[1] (N=928).*

Characteristics	Percent of sample
Ethnicity	
White	96.6
Hispanic	1.6
Other	1.2
Asian	0.2
Black	0.1
Gender	
Male	93.0
Female	7.0
Marital status	
Married	85.0
Unmarried	12.0
Age	
18-34	34.2
35-44	37.5
45-54	20.6
55-64	7.3
65 and older	0.2
Education	
Some high school	4.0
High school diploma	34.0
Some college	52.3
College graduate	9.8
Annual salary	
less than $12,000	1.6
$12,000 to 19,999	3.3
$20,000 to 29,999	23.1
$30,000 to 39,999	49.9
$40,000 to 49,999	15.6
$50,000 and over	6.1

KEY: 1 = Percentages may not total 100 percent due to rounding and/or missing data.

Urine samples were collected from subjects in the telephone interview self-report condition after the initial recruitment interview. All other urine samples were collected from subjects at the time of self-report data collection. Hair samples were also taken from groups 5 and 6 at the time of self-report data collection. Analysis of the urine specimens, conducted by the Center for Human Toxicology at the University of Utah, was performed in three stages: an initial test of the urine for suitability for further testing (pH and specific gravity), an initial radioimmunoassay screen, and confirmational analysis using gas chromatography/mass spectrometry (GC/MS) for any specimens testing presumptively positive by the screen. Cutoff concentrations for specific drug groups are shown in table 2a, along with the specific analyte for which the specimens were tested. Most of the cutoff concentrations were considerably lower than Department of Health and Human Services (DHHS) recommended levels, as the analyses were being conducted for research purposes only.

Hair samples were collected by cutting small locks of hair just above the scalp from the back of the subject's head. The samples were sent to a commercial laboratory for analysis, where an initial radioimmunoassay screen was performed to determine the presence of marijuana, cocaine, opiates, phencyclidine (PCP), and methamphetamines. Unlike the urinalysis, the hair analysis did not include testing for barbiturates or benzodiazepines. Once collected, the hair samples were sectioned, washed four times to remove any external contamination, and then subjected to wash kinetic analysis.[1] The samples were then assayed using radioimmunoassay of hair (RIAH) Standard Screen B for cocaine, methamphetamines, opiates, PCP, and marijuana (Psychemedics 1991).

Positive RIAH screening results for cocaine, methamphetamines, and PCP were reassayed and followed by GC/MS confirmation. In addition, the results of all washes (including the final wash) were assayed for evaluation of three wash kinetic criteria. If the wash criteria did not eliminate the probability of external contamination, additional work was performed (referred to as abnormal wash kinetic or AWK) to further examine the possibility of contamination (Psychemedics 1991).

Because marijuana may not wash off hair in a manner similar to other drugs, wash kinetics are not useful in detecting external contamination.[1] Therefore, GC/MS confirmation for carboxy-THC (tetrahydrocannabinol) was conducted to reduce the probability of external contamination of hair by marijuana smoke (Psychemedics 1991). GC/MS confirmation was also conducted for presumptive positive results for marijuana,

TABLE 2a. *Urinalysis cutoffs.*

Drug group	Specific analyte	Screening cutoff	Confirmation cutoff
Cannabinoids	Delta-9-tetrahydrocannabinol-9-carboxlic acid (carboxy-THC)	20 ng/mL	10 ng/mL
Cocaine metabolite	Benzoylecognine	25 ng/mL	10 ng/mL
Opiates	Morphine/codeine	50 ng/mL	5 ng/mL
PCP	PCP	10 ng/mL	5 ng/mL
Amphetamine	Amphetamine	300 ng/mL	50ng/mL
Methamphetamine	Methamphetamine		
Benzodiazepines	Diazepam, nordiazepam, fluorazpam, N-desalkylfluorazpam, chlordiazepoxide	100 ng/mL	100 ng/mL
Barbiturates	Amobarbital, butalbital, pentobarbital, phenoabarbital, secobarbital		

methamphetamine, PCP, opiates, and cocaine. Cutoff levels for the drugs tested by standard RIAH screening are listed in table 2b.

TABLE 2b. *Hair analysis cutoff levels.*[1]

Drug group	Cutoff levels
Cocaine and benzoylecognine (metabolite)	5 ng/10 mg of hair
Methamphetamine	5 ng/10 mg of hair
Opiates (codeine and morphine)	5 ng/10 mg of hair
PCP	3 ng/10 mg of hair
Total THC (marijuana)	1 ng/10 mg of hair

KEY: 1 = Hair cutoff values cannot be compared to urinalysis cutoff values.

SOURCE: Psychemedics 1991.

Instruments

The self-report questionnaire/interview protocol contained items adapted from NIDA's National Household Survey (Turner et al. 1992). Subjects were asked about their frequency of use of alcohol and 10 major types of drugs in the past 6 months and in the past 30 days. The drug types included marijuana or hashish, cocaine or crack, inhalants, heroin, other opiates, hallucinogens, stimulants, tranquilizers, sedatives, and analgesics. Descriptions and examples of each type of drug were provided to all subjects. Only nonmedical use of drugs was categorized as illicit drug use. If the subject reported prescription drug use and tested positive for that drug, it was classified as a negative self-report and negative urinalysis or hair analysis (i.e., it was classified as medical use and not illicit use of drugs).

Except for the telephone interview condition, all drug use data were collected by means of self-administration of the questionnaire, the technique in which the subject marks on the questionnaire rather than telling the interviewer the answer. This technique has been found to yield higher rates of drug use disclosure than the orally administered interview method (Turner et al. 1992). Thus the individual interviews in the workplace and outside of the workplace were conditions in which one interviewer met in privacy with one subject, explained the study and the questionnaire, then provided the subject with a questionnaire and a pencil so that he or she could self-administer the questionnaire.

There were seven interviewers (four men and three women), all of whom were white and ranged in age from midtwenties to early forties. Four had masters degrees, three had bachelors degrees, and all had experience in both interviewing and in working with drug and alcohol users.

RESULTS

Participation Rates

In each of the six conditions, a small number of workers were unavailable due to vacation, termination, illness, or working at another location. The participation rates among the remaining eligible workers across the four self-report conditions are shown in table 3. The participation rates ranged from 81.1 percent in the offsite condition to 96.6 percent in the individual onsite interview condition.

TABLE 3. *Participation rates across six self-report conditions.*

	Self-Report Condition						
	#1 Workplace interview	#2 Group questionnaire	#3 Telephone interview	#4 Off-site interview	#5 Workplace interview	#6 Group questionnaire	Total
Initial sample	200	200	200	200	200	200	1200
Unavailable	12	17	26	20	21	34	130
dropped	9	3	0	0	0	0	12
Number eligible	179	180	174	180	179	166	1058
Number refusals	6	23	29	34	18	20	130
Number completed	173	157	145	146	161	146	928
	(96.6%)	(87.2%)	(83.3%)	(81.1%)	(89.9%)	(88.0%)	(87.7%)

Drug Use Prevalence Rates by Drug and by Assessment Method

Figure 1 displays prevalence rates for each drug as yielded by each assessment method. Because the hair analysis was conducted on only 307 subjects and for fewer drugs, the results across methods are not precisely comparable. Marijuana was clearly the most prevalent drug used in this sample: By all three methods, more workers were identified as marijuana users than users of all other drugs combined. Although there are some distinct differences by assessment technique, there is a general concordance among the methods, especially between the rates generated by urinalysis and hair analysis.

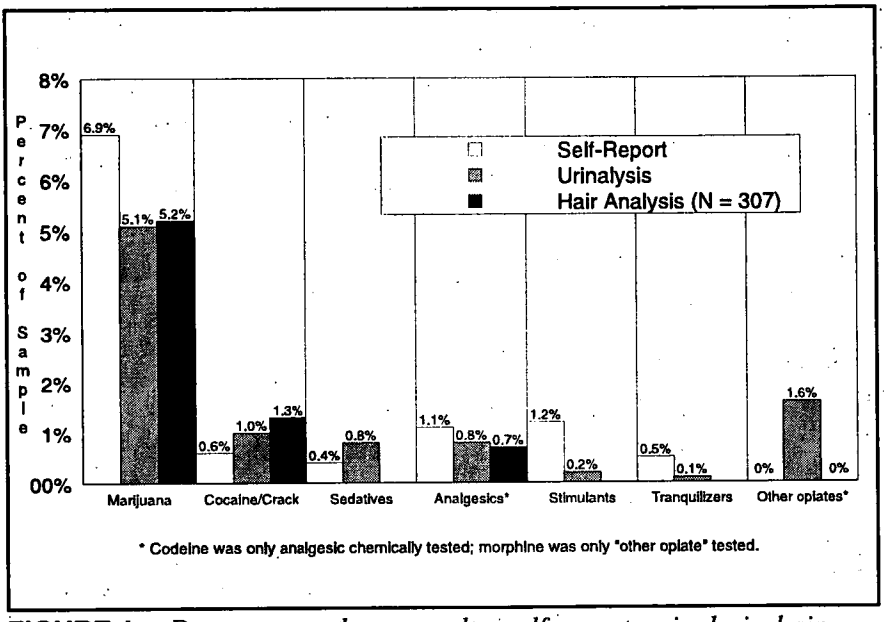


FIGURE 1. *Drug use prevlence results: self-report, urinalysis, hair analysis (N = 928).*

Comparisons of Self-Reports and Urinalysis

Conditions (1) and (5) employed the same self-reporting technique, an individual interview in the workplace. The overall results (any drug) from self-reports and urinalysis for these two conditions are shown in table 4. Included among the 283 subjects who reported no drug use and

TABLE 4. *Comparisons of urinalysis and self-report results for conditions 1 and 5, workplace interview.*

Self-report result	Urinalysis Result		
	Negative	Positive	Total
Negative	283	14	297
Positive	27	10	37
Total	310	24	334

tested negative are 17 subjects who reported legal use of prescription drugs and tested positive for those drugs.

The combined agreement rate (the percentage of subjects classified the same—positive or negative—by both techniques) for the first and fifth self-report conditions was 87.7 percent. Among the discrepancies, 14 subjects tested positive but did not admit to using drugs. Thirty-seven subjects (11.1 percent) self-reported illicit drug use, 27 of whom were found negative by urinalysis. A total of 24 subjects was found positive by urinalysis (7.2 percent), only 10 of whom self-reported illicit drug use. Comparisons of overall results (any drug) from self-reports and urinalysis for the two group questionnaire administration conditions (2 and 6) are shown in table 5. The agreement rate for these conditions was 91.4 percent. In this condition, 16 subjects tested positive but did not admit using drugs, and 10 subjects admitted drug use but tested negative. Seven subjects self-reported drug use and were also found positive by urinalysis.

Comparative results for the third condition (telephone interview) are shown in table 6. The agreement rate between self-report and urinalysis for the telephone interview was 91.0 percent. In this condition, 16 subjects self-reported drug use and 13 subjects tested positive. However, eight of the self-reported drug users tested negative, and five of those testing positive did not report any drug use.

The comparative results for the fourth condition (individual interview off the worksite) are shown in table 7. The agreement rate between urinalysis and self-reports in this condition was 91.1 percent. Seventeen subjects self-reported drug use and 12 subjects tested positive. Nine of the

TABLE 5. *Comparisons of urinalysis and self-report results for conditions 2 and 6, group questionnaire.*

Self-report result	Urinalysis Result		
	Negative	Positive	Total
Negative	270	16	286
Positive	10	7	17
Total	280	23	303

TABLE 6. *Comparisons of urinalysis and self-report results for condition 3, telephone interview.*

Self-report result	Urinalysis Result		
	Negative	Positive	Total
Negative	124	5	129
Positive	8	8	16
Total	132	13	145

17 self-reported drug users tested negative, and 4 of those testing positive did not report any drug use.

The comparative results from self-reports and urinalysis for all conditions combined are shown in table 8. The overall agreement rate across these 928 subjects was 90.0 percent, with 72 subjects testing positive and 87 self-reporting drug use. However, 39 subjects tested positive but did not admit any drug use, and 54 subjects who reported drug use tested negative.

Among the 39 subjects reporting no drug use but testing positive, 8 tested positive for morphine/codeine combined while 7 were positive for morphine alone. Because morphine often appears as a metabolite of codeine, it is likely that many of these subjects may simply have failed to report prescription use of a codeine-based medication. Similarly, the

TABLE 7. *Comparisons of urinalysis and self-report results for condition 4, offsite interview.*

Self-report result	Urinalysis Result: Negative	Positive	Total
Negative	125	4	129
Positive	9	8	17
Total	134	12	146

TABLE 8. *Comparisons of urinalysis and self-report results for all conditions.*

Self-report result	Urinalysis Result: Negative	Positive	Total
Negative	802	39	841
Positive	54	33	87
Total	856	72	928

seven subjects who reported no drug use but tested positive for sedatives may also have simply failed to report prescription use. Unfortunately, there is no way of knowing which of these 20 subjects (2 subjects tested positive for both morphine/codeine and sedatives) simply failed to report prescription use and which were using the drugs illegally.

A total of 54 subjects across all conditions admitted drug use but tested negative by urinalysis. The central reason for discrepancies in this direction is that of the 48 subjects who responded, all but 2 reported a frequency of use—only 1 or 2 days in the past month (or less)—that would place them beyond the range of detection by urinalysis.

Comparisons of Self-Reports and Hair Analysis

Comparisons of overall results (any drug) from self-reports and hair analysis for the individual onsite interview (condition 5) are shown in table 9.

TABLE 9. *Comparisons of hair analysis and self-report results for condition 5, workplace interview.*

	Hair Analysis Result		
Self-report result	Negative	Positive	Total
Negative	141	4	145
Positive	10	6	16
Total	151	10	161

The agreement rate for this condition was 91.3 percent. However, four subjects tested positive but did not admit to using drugs. Sixteen subjects (9.9 percent) self-reported illicit drug use, 10 of whom were found negative by hair analysis. A total of 10 subjects in this condition were found positive by hair analysis (6.2 percent), 4 of whom reported no illicit drug use.

Comparisons of overall results (any drug) from self-reports and hair analysis for the group questionnaire administration (condition 6) are shown in table 10. The overall agreement rate for this condition was 92.5 percent. In this condition, seven subjects tested positive but did not admit using drugs, and four subjects admitted drug use but tested negative. Three subjects self-reported drug use and were also found positive by urinalysis.

The comparative results from self-reports and hair analysis for both conditions combined are shown in table 11. The overall agreement rate across these 307 subjects was 91.9 percent, with 20 subjects testing positive and 23 self-reporting drug use. However, 11 subjects tested positive but did not admit any drug use, and 9 subjects who reported drug use tested positive.

TABLE 10. *Comparisons of hair analysis and self-report results for condition 6, group questionnaire.*

Self-report result	Hair Analysis Result		
	Negative	Positive	Total
Negative	132	7	139
Positive	4	3	7
Total	136	10	146

Among the 11 subjects reporting no drug use but testing positive, 3 tested positive for codeine alone. As previously mentioned, this may result from subjects' failure to report prescription use of a codeine-based medication.

TABLE 11. *Comparisons of hair analysis and self-report results for both conditions.*

Self-report result	Hair Analysis Result		
	Negative	Positive	Total
Negative	273	11	284
Positive	14	9	23
Total	287	20	307

As shown in table 11, a total of 14 subjects across both conditions admitted drug use but tested negative by hair analysis. Of these, five admitted use of tranquilizers, analgesics (other than codeine), or sedatives, drugs that were not screened by hair analysis. Of the remaining nine drug users who tested negative by hair analysis, only one marijuana user reported using the drug three to six times per week. The other drug users reported using the drug twice a month or less, with the last use occurring more than 1 week before testing.

Comparisons of Urinalysis and Hair Analysis

The comparative results from hair analysis and urinalysis for conditions 5 and 6 combined are shown in table 12. The overall agreement rate across these 307 subjects was 94.8 percent, with 20 subjects testing positive by hair analysis and 22 testing positive by urinalysis. There were few discrepancies, with seven subjects testing positive by hair analysis but not by urinalysis, and nine subjects testing positive by urinalysis but not by hair analysis.

Of the seven subjects testing positive by urinalysis and negative by hair analysis, two tested positive for use of a morphine/codeine combination (counted as four positives), three tested positive for morphine alone, and two were positive for marijuana use. The presence of morphine combined with codeine possibly suggests the use and subsequent metabolization of codeine, which was screened by both urinalysis and hair analysis. The remaining two subjects tested positive by urinalysis and

TABLE 12. *Comparisons of hair analysis and urinalysis results for conditions 5 and 6.*

	Hair Analysis Result		
Urinalysis result	Negative	Positive	Total
Negative	278	7	285
Positive	9	13	22
Total	287	20	307

negative by hair analysis for marijuana use. Although the hair analysis procedure did detect several marijuana users, the laboratory has indicated that the detection of marijuana is the most problematic of the drugs for which hair analysis is conducted.

Among the seven subjects testing positive by hair analysis and negative by urinalysis, three tested positive for marijuana use, three tested positive for cocaine use, and one tested positive for codeine. Of these subjects, only one reported use of any illicit drugs. This subject reported cocaine

TABLE 13. *Overall drug use prevalence rates by assessment method and condition.*

	Self-Report Condition						
Assessment method	#1 Workplace interview	#2 Group questionnaire	#3 Telephone interview	#4 Off-site interview	#5 Interview phase 2	#6 Quest. phase 2	Total
Self-report	12.1%	6.4%	11.0%	11.6%	9.9%	4.8%	9.4%
Urinalysis	6.9%	8.3%	9.0%	8.2%	7.5%	6.8%	7.8%
Hair analysis					6.2%	6.8%	6.5%

use to be 1 or 2 days within the past month. The subjects last use was reported to be more than 1 month ago, which could explain the lack of detection by urinalysis.

As mentioned above, 14 subjects in conditions 5 and 6 yielded conflicting chemical test results; however, 13 of these 14 subjects reported no illicit drug use. Had these subjects reported use of these drugs, more information would be available to explain the possible causes of discrepancies between the chemical analysis techniques.

Calculation of Drug Use Prevalence Rates

Drug use prevalence rates can be calculated for this workforce based on the specific testing methods employed. As shown in table 13, the drug use prevalence rates based on self-reports are generally around 11 percent, except for the group administration condition, which generated a prevalence rate less than half that of the other conditions. The aggregate prevalence rate for urinalysis was 7.8 percent across the entire sample of 928, while the self-report method produced a prevalence rate of 9.4 percent. Across the sample of 307 for conditions 5 and 6, the hair analysis prevalence rate was 6.5 percent and the urinalysis prevalence rate was 7.2 percent.

However, the actual prevalence rate is clearly higher than indicated by any of these methods used alone. A better estimate of drug use prevalence is obtained by combining the number of employees self-reporting illicit drug use with those testing positive by either the urinalysis or hair analysis but not admitting drug use. Using this estimation, 87 workers self-reported illicit drug use, another 39 not admitting use were found positive by urinalysis, and 6 who did not report drug use were found positive by hair analysis but negative by urinalysis. Therefore at least 132 workers, or 14.2 percent of the workforce, may be classified as drug users.

DISCUSSION

The Prevalence of Illicit Drug Use in the Workforce Sample

The rate of illicit drug use found in this study (14.2 percent) was perhaps somewhat lower than might have been expected, as the National Household Survey on Drug Abuse reported a rate of 13.1 percent among employed adults, a rate based solely on self-reports (NIDA 1993). In this

study, the prevalence rate produced by self-reports alone was only 9.4 percent, a rate that was clearly suppressed by the group administration conditions. Indeed, the rate produced by the individual interview conditions (a method very similar to that used in the National Household Survey) ranged between 9.9 percent and 12.1 percent. In addition, the workforce in this study was located in a medium-sized western city, away from any of the major urban areas where drug use is relatively high. Therefore, although the prevalence rate may be considered low in comparison to other populations and regions, it is quite comparable to the rates reported by other investigators during the past few years (e.g., Lehman and Simpson 1992).

The Validity of Self-Reported Drug Use in the Workplace

This study may be viewed as a classic criterion validity design in which the chemical tests (urinalysis and hair analysis) are the objective criteria against which the self-report is compared. Although the chemical tests are susceptible to error, the urinalysis techniques are generally considered quite accurate, particularly when initial positives are confirmed by GC/MS. Questions remain about the accuracy of hair analysis, especially with respect to environmental contamination (Harkey and Henderson 1988). In this sample, the rates of false negatives and false positives for hair analysis appear quite low, and many of the false positives are probably attributable to the wider window of detection in comparison to urinalysis, the typical criterion measure used (Magura et al. 1992; Mieczkowski et al. 1993). This is not to suggest that the chemical tests are perfect criterion measures. The three methods are measuring constructs of drug use that overlap yet are distinctly different; therefore, one would not expect complete congruence among the three methods. Indeed, when subjects disclosed their drug use but produced a negative (i.e., drug-free) urinalysis result, the discrepancies were shown to be almost entirely a function of the subject's low frequency of drug use. However, when the discrepancy lies in the other direction (self-reports of no drug use accompanied by a positive urinalysis), there is little doubt that the urinalysis result is correct and the self-report is not. Thus, the urinalysis serves as a partial, but effective, validity criterion. In this study, hair analysis serves a similar criterion function. Because of its putatively longer period of detection, hair analysis should provide results that are temporally more isomorphic to self-reports than are those of urinalysis. However, the technology of hair analysis often (as in the current instance) does not provide tests for as many drugs as urinalysis.

The comparisons of self-report and chemical testing raise serious questions about the validity of self-reports of illicit drug use in the workplace. Of the 72 subjects whose urinalysis showed them to have recently ingested an illicit drug, less than half admitted any drug use in the past 6 months. Mitigating this effect somewhat is the likelihood that some fraction of these nondisclosers may have used prescribed codeine. Yet it is also likely that given the limited detection period of urinalysis, there were additional subjects who were nondisclosing drug users but whose last use was sufficiently in the past that they were beyond the detection range of urinalysis. The comparison of hair analysis results with self-reports produced similar findings. Of the 20 subjects whose hair analysis showed them to have used an illicit drug, less than half (i.e., 9 subjects) admitted any drug use in the past 6 months.

Stated differently, these comparisons show that the drug use prevalence rate in a workplace is likely to be approximately 50 percent higher than the estimate based on self-reports. When the subjects who refused to participate are taken into account, the actual rate might be higher still—although probably not substantially higher. The prevalence rate in the first condition (individual interview in the workplace), where the refusal rate was only 3.4 percent, was virtually the same as the fourth condition (offsite interview), where the refusal rate was 18.9 percent. If the refusal group was heavily laden with drug users, it is likely (though by no means necessary) that the fourth condition, with its high refusal rate, would produce a prevalence rate considerably lower than the first condition. Moreover, the detected nondisclosers are current (and perhaps frequent) drug users—the people in whom one would be most interested if one were studying the effects of worker drug use.

These findings have significant implications for studies that are attempting to determine relationships between illicit drug use and any number of job performance issues and are relying on self-reports as the primary measure of drug use. Based on these data, it appears that such studies will be missing a sizable, important group of drug-using workers. The findings also cast considerable doubt on the accuracy of workforce prevalence estimates based solely on self-reports. However, these results do not necessarily invalidate studies of drug use in the workforce that have relied heavily on self-reports. If one is not developing prevalence estimates but rather conducting research on general issues of drug use in the workforce, the problem of underreporting is less consequential.

Although these results are most relevant to studies of drug use in the workplace, they may also have implications that reach beyond the workplace to the general question of the validity of drug use self-reports. For several years, Wish (1990) has contended that prevalence estimates based on self-reports (including the National Household Survey) underestimate the rates of drug use, a contention based mainly on the lack of self-disclosing drug use among arrestees tested in the Drug Use Forecasting system. This study provides one of the few comparisons of self-reports and chemical tests in a normal (i.e., nonarrestee, nonaddict) population. One might expect a great deal of denial of drug use among arrestees questioned by law enforcement authorities in a jail. Less expected was the considerable denial of drug use among employed adults when assessed by a research team under conditions of anonymity and confidentiality. Although the setting is different, the data-collection procedures and the population were quite similar to those used in the National Household Survey (NIDA 1993). The underreporting found in this study also lends support to the position taken by GAO in a recent report expressing concern that the two major prevalence assessment activities of the Federal Government—the National Household Survey and the Monitoring the Future Survey—rely solely on self-reports (GAO 1993). Both that report and a recent NIDA publication on drug use survey methodologies discuss the need for "direct assessment of the *validity* of the measurements themselves" (Turner et al. 1992, p. 305).

Caution must be exercised, however, in the interpretation of these particular results, as the sample was drawn from only one company's workforce and did not contain a large number of drug users. Moreover, with the exception of marijuana, no specific type of drug was reported or detected with high frequency.[1]

Comparisons of Different Modes of Self-Report

Because the subjects were randomly assigned to the four different self-report conditions, one would expect the samples to be roughly equivalent in composition and in druguse prevalence rates. In fact, three of the four conditions produced drug use rates remarkably similar to each other, between 9.9 percent and 12.1 percent across the three conditions and four groups. It seems to matter little whether the mode of self-report is an individual interview/questionnaire in the workplace, a telephone interview in the worker's home, or an individual interview/questionnaire outside of the workplace. However, the group questionnaire method produced self-report drug use rates that were roughly half those of the

other conditions. This lower rate was produced by the group method in the first phase of the research, and was essentially replicated in the second phase. In the first phase, the group rate was 53 percent of the rate produced by the workplace interview method; in the second phase, it was 48 percent of the workplace interview method. It seems clear that this difference is not a function of there being fewer actual drug users in the group condition. In the first phase, the rate of urinalysis positives in the group condition was 8.3 percent, compared to an average of 8.0 percent in the other three conditions. In the second phase, the urinalysis rates across the two conditions were similar. There seems to be little doubt, therefore, that in this workplace, the group situation greatly suppressed self-reports of illicit drug use.

The fact that the telephone interview produced drug use rates that were comparable to the in-person individual interview was unexpected and stands in some contrast to the findings of Gfroerer and Hughes (1992), who found that surveys conducted by telephone tend to produce underestimates of drug use prevalence compared to in-person interviews. The higher disclosure rates found in the current study probably occurred, at least in part, because the telephone interview subjects in this study were first recruited through individual in-person sessions; the actual interview was later conducted by telephone. This initial, in-person recruitment session doubtless helped to engender trust and rapport that would otherwise not be gained in a telephone interview.

These data indicate that the general underreporting of drug use noted above is greatly exacerbated when the self-reports are collected from groups in the workplace. This group suppressor effect may also be present in other studies of drug use, both in and outside the workplace, where data are collected in groups. For example, it is noted that as the Monitoring the Future survey (Johnston et al. 1993) is conducted in classrooms, the self-reporting of illicit drug use may be further suppressed—although students are quite accustomed to providing a variety of information in group conditions.

The Uses of Urinalysis and Hair Analysis in Drug Use Prevalence Assessment

By themselves, urinalysis and hair analysis typically provided estimates of drug use prevalence that were substantially lower than those produced by self-reports. Only in the group administration condition did the urinalysis and the hair analysis generate higher prevalence rates than self-reports. Of

the 87 subjects who self-reported drug use, a sizable majority (54) produced a negative urinalysis result, mainly due to the constricted detection window of urinalysis. Similarly, of the 23 subjects who self-reported drug use in the last two groups (from whom hair samples were taken), a comparable majority (14) tested negative on hair analysis. The latter finding was somewhat unexpected, as hair analysis is reputed to provide a wider period of detection. Although 6 of the 14 subjects were using drugs not screened by hair analyses in this study, 7 of the remaining 8 subjects reported marijuana use. It appears that the hair analysis procedures are especially prone to false negatives in cases of marijuana use, particularly if the use is infrequent.

In short, as prevalence assessment methods, the chemical tests—when used alone—perform even more poorly than the self-report methods. It should be pointed out, however, that this investigation into hair analysis was more exploratory than definitive; future research should test for more drugs on larger samples.

On the other hand, when the chemical tests are used in combination with self-reports, they become a powerful addition to the prevalence assessment methods, doubtless providing a drug use prevalence rate that is much closer to the true rate. Thus, when the urinalysis and hair analysis results are combined with self-report, the resultant prevalence rate (14.2 percent) was 51 percent higher than the rate based on self-report alone. Indeed, given these findings, it would seem evident that the best strategy would be to combine self-report with chemical testing—not only for the workplace, but for surveys of the general population as well, and not only for validation purposes, but for prevalence assessment purposes. In response to a GAO recommendation that the National Household Survey include hair testing (on a limited test basis), NIDA officials expressed concern that response rates might be depressed as a result (GAO 1993). This study showed that with adequate incentives and confidentiality assurances, response rates equivalent to those currently achieved by the National Household Survey (80 to 85 percent) are possible even when biological specimens are obtained from respondents (GAO 1993).

ENDNOTE

1. Refer to the Technical Note at the end of the Introduction (p. 13).

REFERENCES

Anglin, M.D., and Westland, C.A. Drug monitoring in the workplace: Results from the California commercial laboratory drug testing project. In: Gust, S.W., and Walsh, J.M., eds. *Drugs in the Workplace: Research and Evaluation Data.* National Institute on Drug Abuse Research Monograph 91. DHHS Pub. No. (ADM) 89-1612. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1989.

Baumgartner, W.A.; Baer, J.D.; Hill, V.A.; and Blahd, W.A. *Hair Analysis for the Detection of Substance Abuse in Pretrial/Probation/ Parole Populations.* Final report of activities under National Institute of Justice (NIJ) grant no. 86-IJ-CX-0029. Washington, DC: NIJ, 1989.

Blanton, A.E.; Kidwell, R.E.; and Bennett, N. Application of performance tests to identify workplace drug users: A panacea or a familiar set of problems? *J Employee Assist Res* 1:350-361, 1992.

CONSAD Corporation. *Analysis of Occupational Substance Use and Workplace Safety: Final Report.* Pittsburgh, PA: CONSAD Research Corporation, 1989.

Cook, R.F. Drug use among working adults: Prevalence rates and estimation methods. In: Gust, S.W., and Walsh, J.M., eds. *Drugs in the Workplace: Research and Evaluation Data.* National Institute on Drug Abuse Research Monograph 91. DHHS Pub. No. (ADM)89-1612. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1989.

Cook, R.F., and Bernstein, A. Assessing drug use prevalence in the workplace: A comparison of self-report methods and urinalysis. *Int J Addict* 29(8):1057-1068, 1994.

Cook, R.F.; Walizer, D.; and Mace, D. Illicit drug use in the Army: A social-organizational analysis. *J Appl Psychol* 6(3):262-272, 1976.

Crouch, D.J.; Webb, D.O.; Peterson, L.V.; Buller, P.F.; and Rollins, D.W. A critical evaluation of the Utah Power and Light Company's substance abuse management program: Absenteeism, accidents and costs. In: Gust, S.W., and Walsh, J.M., eds. *Drugs in the Workplace: Research and Evaluation Data.* National Institute on Drug Abuse Research Monograph 91. DHHS Pub. No. (ADM)89-1612. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1989.

General Accounting Office (GAO). *Drug Use Measurement: Strengths, Limitations and Recommendations for Improvements.* Washington, DC: U.S. General Accounting Office, 1993.

Gfroerer, J., and Hughes, A. Collecting data on illicit drug use by phone. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* Rockville, MD: National Institute on Drug Abuse, 1992.

Harkey, M.R., and Henderson, G.L. *Hair Analysis for Drugs of Abuse: A Critical Review of the Technology.* Sacramento, CA: Department of Alcohol and Drug Programs, 1988.

Hurst, P.; Cook, R.F.; and Ramsay, D. *Assessing the Prevalence of Illicit Drug Use in the Army.* Arlington, VA.: Army Research Institute for the Behavioral and Social Sciences, 1975.

Johnston, L.D.; O'Malley, P.M.; and Bachman, J.G. *National Survey Results on Drug Use from Monitoring the Future Study, 1975-1992.* Rockville, MD: National Institute on Drug Abuse, 1993.

Lehman, W., and Simpson, D. Employee substance use and on-the-job behaviors. *J Appl Psychol* 77:309-321, 1992.

Magura, S.; Freeman, R.C.; Siddigi, Q.; and Lipton, D. The validity of hair analysis for detecting cocaine and heroin use among addicts. *Int J Addict* 27(1):51-69, 1992.

Mieczkowski, T.; Landress, H.J.; Newel, R.; and Coletti, S. Testing hair for illicit drug use. *Research in Brief.* Washington, DC: National Institute of Justice, 1993.

National Institute on Drug Abuse. *National Household Survey on Drug Abuse: Highlights 1991.* Rockville, MD: NIDA, 1993.

Normand, J.; Salyards, S.; and Mahoney, J.J. An evaluation of pre-employment drug testing. *J Appl Psychol* 75:629-639, 1990.

Psychemedics Corporation. *Policies and Procedures Manual.* Santa Monica, CA: Psychemedics Corp., 1991.

Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C. Future directions for research and practice. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* Rockville, MD: National Institute on Drug Abuse, 1992.

Winkler, H., and Sheridan, J. "An Examination of Behavior Related to Drug Use at Georgia Power Company." Paper presented at the National Institute on Drug Abuse Conference on Drugs in the Workplace: Research and Evaluation Data. Bethesda, MD, September 1989.

Wish, E. U.S. drug policy in the 1990's: Insights from new data from arrestees. *Int J Addict* 25:377-409, 1990.

ACKNOWLEDGMENTS

This study was supported by grant 5-R01-DA05691 from the National Institute on Drug Abuse. The authors wish to thank Thadeus Arrington of the Institute for Social Analysis and Gordon Marshall of CONSAD for their contributions to this research.

AUTHORS

Royer Cook, Ph.D.
President

Christine Andrews, M.A.
Research Associate

Institute for Social Analysis
201 North Union Street
Alexandria, VA 22314

Alan Bernstein, M.S.
CONSAD Research Corporation
121 North Highland Avenue
Pittsburgh, PA 15206

Studies of Nonresponse and Measurement Error in the National Household Survey on Drug Abuse

Joseph Gfroerer, Judith Lessler, and Teresa Parsley

ABSTRACT

A summary of the results of a series of studies of nonresponse and measurement error in the National Household Survey on Drug Abuse (NHSDA) is given in this chapter. Two studies not previously reported, the Skip Pattern Experiment and the Census Match Study, are the primary focus of the chapter. The Skip Pattern Experiment involved a test of a modified NHSDA questionnaire that made extensive use of skip patterns in drug use questions. Compared to the standard NHSDA method, which avoids skip patterns, the modified questionnaire tended to produce lower rates of reported drug use. The Census Match Study involved linking 1990 NHSDA nonrespondent cases with data from the 1990 Decennial Census. Household and individual data for NHSDA nonrespondents were obtained from the Census and used to characterize NHSDA nonresponse patterns in detail. A multilevel logistic model of response propensity identified the important predictors of nonresponse, including characteristics of the sampled person, the selected household, the neighborhood, and the interviewer.

INTRODUCTION

This chapter reports on a series of methodological studies conducted in conjunction with the NHSDA. These studies, sponsored primarily by the National Institute on Drug Abuse (NIDA) (sponsorship of the NHSDA and related methodological studies was given to the Substance Abuse and Mental Health Services Administration (SAMHSA) in 1992), were intended to evaluate NHSDA methodologies and test new ones; their focus was primarily on survey errors resulting from nonresponse and measurement error. Many of the results of these studies can be found elsewhere (Turner et al. 1992). This chapter briefly summarizes the results of the previously published studies and focuses on two more

recent studies, which are described in more detail in unpublished reports (Gfroerer, unpublished data). The first study, the Skip Pattern Experiment, assessed the potential measurement error that would result by introducing skip patterns into the NHSDA self-administered answer sheets (Lessler and Durante 1992). The second study, the Census Match Study, involved an analysis of nonresponse in the NHSDA and its potential for causing errors in estimation (Parsley 1993).

SOURCES OF ERROR IN SURVEYS

To put these studies in context, it is useful to summarize all the sources of error that occur in the NHSDA and surveys in general. Efforts to improve the quality of survey estimates should always focus on total survey error. Such a discussion may also serve to clarify terminology and, it is hoped, contribute to an improvement in communication between survey researchers, drug abuse researchers, and drug abuse policymakers. A commonly used term such as "nonresponse," for example, could easily be misinterpreted by some to mean the denial of drug use by survey participants who have used drugs (i.e., incorrect response), when in fact it refers to the failure to obtain data from some sampled units (i.e., no response).

Survey errors can be classified into four types: coverage, sampling, nonresponse, and measurement (Groves 1987). Coverage error results from using a sampling frame that does not include all of the target population. In establishment surveys (surveys of schools or businesses, for example), coverage error often results when eligible units are not included in lists of establishments from which the sample is drawn. In household surveys, undercoverage primarily occurs because members of the target population are not reported during screening as being members of any household.

Sampling error results when data are intentionally collected from only a portion of the sample frame. Methods of estimating the magnitude of sampling error are available when probability-based sampling is used.

Nonresponse errors result from the failure to obtain data from units that are selected to be in the sample. This can occur because potential respondents cannot be located or because they refuse to participate in the survey. The magnitude of this error depends on both the response rate (the percent of the sample from which data are obtained) and the

difference between respondents and nonrespondents in the attribute (e.g., use of drugs) being measured. In household surveys, nonresponse can occur at the household level, person level, and questionnaire item level.

Finally, the most often studied type of error is measurement error, which can be defined as the discrepancy between respondents' true attributes and the data obtained in the survey about their attributes. Response error has many sources, including the mode of interview, wording of questions, interviewer behavior, sensitivity of information requested, respondents' recall, and coding errors.

All four types of error are of concern in estimating drug use with household surveys. In measuring hardcore drug use, coverage error could be significant because many drug users may be transient or not permanently attached to one particular household. Sampling error is a problem in household surveys of drug use because many behaviors being measured have very low prevalence in the general population. Many users of NHSDA data, when told that the data are from a survey asking people to report on their drug use, assume that most drug users refuse to participate in such a survey (i.e., nonresponse error) or, if they do participate, they will lie about their drug use (i.e., measurement error). While all of these types of errors are undoubtedly present in NHSDA data, no study has comprehensively evaluated the relative contribution of each type.

DESIGN OF THE NHSDA

A description of the NHSDA sample design and estimation methodology can be found in published reports from the survey (SAMHSA 1993, 1994). For this chapter, a summary of the NHSDA data-collection method is given.

In-person interviews were conducted with sample persons, incorporating procedures that would be likely to enhance respondents' cooperation and willingness to report honestly about their illicit drug use behavior. Introductory letters were sent to sampled addresses, followed by an interviewer visit. A 5-minute screening procedure involved listing all household members along with their basic demographic data and a selection of sample person(s) based on the household composition. Zero, 1, or 2 persons could be selected. Interviewers attempted to conduct interviews in a private place, away from other household

members. The interview averaged about an hour, and included a combination of interviewer-administered and self-administered questions. With this procedure, the answers to sensitive questions (such as those on illicit drug use) were recorded by the respondent and not seen or reviewed by the interviewer. After these answer sheets were completed, they were placed by the respondent in an envelope, which was sealed and mailed back to the contractor. The self-administered answer sheets are also designed to conceal responses from interviewers by avoiding the use of skip patterns that could allow nondrug users to skip questions on drug use that did not pertain to them, thus identifying drug users as those who take longer to complete the answers. Skip patterns might also induce some drug users to deny their use as a way of avoiding answering more detailed questions that might follow a positive response.

MEASUREMENT ERROR IN THE NHSDA

Studies of measurement error in the NHSDA have focused on errors resulting from the questionnaire design and the mode of data collection. The basic issues of interest are how questions should be asked to maximize validity and reliability, and under what survey conditions are respondents most likely to provide accurate data. Results of some of this research can be found elsewhere (Turner et al. 1992); a few highlights are described below.

- Cognitive evaluations and analyses of inconsistent reporting patterns identified questionnaire items that needed revision to reduce response error (Cox et al. 1992; Forsyth et al. 1992).

- A comparison of NHSDA data to a national telephone survey found that respondents are less likely to report drug use by telephone than in person (Gfroerer and Hughes 1992).

- A methodological field test found that respondents are more likely to report drug use using self-administered answer sheets than with interviewer-administered questionnaires (Turner et al. 1992).

- Underreporting of drug use was found to be significant among a sample of former treatment clients who were selected to participate in the NHSDA at home (Harrell et al. 1986; Harrell, this volume).

- An analysis of NHSDA data found that youth were more likely to report drug use when interviews were conducted in private (Gfroerer 1985).

- A comparisons of NHSDA and Monitoring the Future data found that among young teenagers, reporting of drug use was more likely in a school setting than at home (Gfroerer 1992).

SKIP PATTERN EXPERIMENT

Design

Throughout its history the NHSDA has generally avoided the use of skip patterns because of the fear that respondents will realize that their use or nonuse of a drug will be revealed to interviewers (and others present during the interview) based on the length of time needed to complete answer sheets, thus diminishing confidentiality. Also, drug users may deny their use if they realize that a "never used" response will allow them to avoid answering a series of questions, thus saving time. Another concern is that respondents may not be able to follow skip patterns correctly in a self-administered answer sheet.

However, substantial benefits in terms of reduced respondent burden and expanded questionnaire content would occur with the introduction of skip patterns into the NHSDA questionnaire. Therefore it is important to know whether skip patterns can be implemented without seriously affecting data quality. Furthermore, if data quality is affected, it is also important to understand the mechanisms involved if there is to be a successful movement toward computer-assisted data collection.

In 1990, NIDA conducted a large methodological field test, primarily to evaluate the effect of using interviewer-administered versus self-administered questionnaires. In this study, some questionnaire answer sheets incorporated skip patterns. It was found that, in general, respondents were able to follow skip patterns that were not too complex, particularly if skips were always to the top of a page. Generally, when errors

occurred, they resulted in respondents answering additional questions unnecessarily, so there was no loss of data. However, the design of the field test did not allow a determination of the effect of skip patterns on reporting of drug use. In an attempt to address this, NIDA conducted an experiment during the first 3 months of 1992 to test a new questionnaire that incorporated skip patterns into the drug use answer sheets.

An experimental questionnaire was developed that included a number of skip patterns that allowed respondents to skip out of entire sections of questions if they responded "no" to an initial question on whether they had used a drug. This questionnaire was called the skip version. The regular NHSDA questionnaire was called the nonskip version. Differences between the two questionnaires varied by section in the questionnaire. The main differences between the two versions are summarized below.

- Cigarettes—Questions are identical. Both use interviewer-administered mode with skip patterns.

- Alcohol—In the skip version, respondents are told, "If you've never had an alcohol drink, just circle the 991 in the box after A-1 and tell me that you are finished with this answer sheet." In the nonskip version, respondents are told to answer all questions.

- Sedatives, tranquilizers, stimulants, analgesics—Questions are identical. Both use answer sheets with a skip pattern.

- Marijuana, inhalants, cocaine, crack, hallucinogens, heroin—In the skip version, respondents were told after an early question that if they never used the drug, they could skip the remainder of this answer sheet. In the nonskip version, all questions had to be answered.

Another difference between the two versions that could potentially affect the reporting of drug use was the technique used by interviewers to administer the answer sheets. For certain answer sheets, interviewers were instructed to read the questions aloud to respondents while respondents filled in the answers. This procedure is used in the NHSDA for the first several answer sheets to help respondents understand the questions. In the nonskip version, interviewers read the alcohol, marijuana, inhalants (for 12- to 17-year-olds only), cocaine, and crack answer sheets. In the skip version, reading questions aloud would have made it obvious to interviewers what the responses were to initial drug use questions. Thus,

to enhance respondent privacy, only the questions on the first answer sheet (alcohol) were read aloud to respondents.

The Skip Pattern Experiment was embedded in the first quarter 1992 NHSDA sample. One-eighth of the first quarter sample was randomly assigned to receive the skip questionnaire, while the other seven-eighths received the nonskip version. Assignment of questionnaire versions to sampled dwelling units was done in advance of any contact by field staff. Allocation of the skip version was done within sample segments to maximize the power of statistical comparisons between the two groups. Interviewers were trained to administer both versions of the questionnaire. Overall, the nonskip version was administered to 7,149 respondents and the skip version was administered to 974 respondents.

Results

In the first stage of analysis, unweighted, unedited estimates of lifetime prevalence from the two questionnaires were compared to determine whether there were any indications that the skip questionnaire resulted in lower prevalence rates. A one-sided test of the hypothesis that the nonskip version produced lower prevalence was employed, using Fisher's exact test, assuming a simple random sample, and with a level of significance of 0.15. If this hypothesis could not be rejected (for answer sheets that had skips in the skip version and no skips in the nonskip version), it would suggest that using skips was a viable option for the NHSDA. The results of this analysis are shown in table 1. Of the six illicit drug categories that had different questionnaire versions, the hypothesis was rejected in three cases. For marijuana, cocaine, and hallucinogens, rates were significantly lower with the skip version. For legal drugs and for drugs that used skip patterns in both questionnaires, the hypothesis could not be rejected in any case.

To evaluate the impact of using skips on actual NHSDA prevalence estimates, a comparison was done of prevalence rates in the two versions based on weighted and edited data. This analysis revealed that, due to variations across subgroups in how the skip questionnaire affected response, weighted comparisons showed larger differences than unweighted comparisons. Analysis of these differences showed that this was primarily because the lower reports with the skip questionnaire tended to be concentrated among more educated groups, which are generally sampled at lower rates in the NHSDA. For example, the skip questionnaire produced a weighted estimate of lifetime marijuana use that was 11 percent lower

TABLE 1. *Lifetime prevalence of drug use in two questionnaire versions used in first quarter 1992 NHSDA skip pattern experiment; unweighted and unedited, all ages 12+.*

Drug type	Nonskip (N = 7,149) %	Skip (N = 974) %	P-value[1]
Identical questions in two versions			
Cigarettes	58.9	59.0	NA
Sedatives	3.5	3.9	0.681
Tranquilizers	5.1	5.9	0.836
Stimulants	6.4	6.9	0.733
Analgesics	6.6	6.6	0.497
Different questions in two versions			
Alcohol	74.4	77.7	0.988
Marijuana	35.2	33.3	0.107
Inhalants	6.8	6.1	0.200
Cocaine	13.2	11.1	0.025
Hallucinogens	8.8	7.0	0.020
Heroin	2.2	2.1	0.394

KEY: 1 = Based on a one-sided test of the hypothesis that the nonskip version produced a lower prevalence, using Fisher's exact test.
NA = not available.

than the nonskip questionnaire among respondents with less than a high school education. Among respondents with a college degree, the skip estimate was 27 percent lower than the nonskip estimate. Table 2 shows the comparison of weighted, edited prevalence rates for the two questionnaires. The nonskip version produced higher rates of lifetime drug use prevalence for all five illicit drugs that did not have skip patterns. Estimates of past month use indicate an even larger effect of using skip patterns for marijuana and cocaine.

Overall, this methodological study indicates that using skip patterns tends to reduce the prevalence of illicit drug use. There is also an indication that the bias due to using skips would not be uniform across different

TABLE 2. *Lifetime and past month prevalence of drug use in two questionnaire versions used in first quarter 1992 NHSDA skip pattern experiment; weighted and edited, all ages 12+.*

Drug type	Nonskip (N = 7,149) %	Skip (N = 974) %	Relative difference %
Lifetime prevalence			
Identical questions in two versions			
Cigarettes	70.9	67.7	4
Sedatives	4.2	4.8	-14
Tranquilizers	5.6	5.0	11
Stimulants	6.8	6.7	1
Analgesics	6.1	6.0	2
Different questions in two versions			
Alcohol	82.1	80.6	2
Marijuana	32.8	27.9	15
Inhalants	5.1	4.5	11
Cocaine	12.0	7.5	38
Hallucinogens	8.0	6.1	24
Heroin	0.9	0.7	16
Past month prevalence			
Identical questions in two versions			
Cigarettes	27.7	26.3	5
Difference questions in two versions			
Alcohol	49.3	51.5	-4
Marijuana	5.0	2.9	42
Cocaine	0.8	0.3	56

populations, as it seemed to be more pronounced among respondents with more education. It was not possible to conclude whether the lack of privacy or the desire to avoid additional questions was operating. However, it is interesting to note that the skip in the alcohol questions had no apparent impact. This would suggest that it is the sensitivity of the illicit drug questions and the loss of privacy in the skip version that is most important.

NONRESPONSE ERROR IN THE NHSDA

Nonresponse error in the NHSDA is of particular concern because of a recent decrease in response rates. In the 1992 NHSDA, the screening and interview response rates were 95.0 percent and 82.5 percent, respectively. In 1993, the respective rates had dropped to 93.9 percent and 79.2 percent. While data are available on the demographic characteristics of NHSDA nonrespondents that allow inferences of whether nonresponse bias is likely, there is very little information on the drug use patterns of nonrespondents. In the only study to attempt to determine the drug use of NHSDA nonrespondents, a followup of 1990 NHSDA nonrespondents in the Washington, DC, area resulted in completed interviews with 38 percent with a shortened questionnaire and monetary incentives (Caspar 1992). This study found that nonrespondents did have higher rates of cocaine use than respondents. Because of the difficulties involved in obtaining data on nonrespondents' drug use, studies of nonresponse bias often rely on obtaining measurements of known correlates of drug use. This was the intent of the Census Match Study conducted in 1991.

CENSUS MATCH STUDY

Design

In 1991, NIDA and the Research Triangle Institute cooperated with the U.S. Bureau of the Census on a study of nonresponse in surveys. This effort was part of the multiagency work conducted by Groves and Couper at the Census Bureau to study nonresponse in seven major Federal Government surveys, including the NHSDA. The study involved linking data from a sample of respondents and nonrespondents to the 1990 NHSDA with their 1990 Decennial Census data to provide descriptive information about NHSDA nonrespondents. NHSDA and census records were linked by Census Bureau clerks using primarily address and other location information.

A total of 5,030 NHSDA households were selected to be matched to 1990 census records. All 860 screener nonresponse households were selected. In addition, all 1,821 households with at least one interview nonrespondent were selected for matching. To allow comparisons between respondents and nonrespondents on the census items, a random systematic sample of 1,938 households was selected in which all sample persons completed the interview. Finally, to assess the accuracy of

interviewers' classifications, a sample of 411 cases classified as vacant, temporary dwellings, or nonhousing units by the NHSDA interviewer was selected.

Excluding the 411 noneligible cases, the NHSDA sample had a very high matching success rate of 97.2 percent, compared with rates ranging from 93.4 percent to 97.5 percent in the other Government surveys in the study (Groves and Couper 1992). It should be noted that this matching was based primarily on address information, and that in some cases the household members were different in the NHSDA and census data for matched addresses.

In households with completed NHSDA screeners, a second match procedure was performed to identify the individuals selected for the survey interview. This person match used the gender, race, Hispanic origin, and age information from the NHSDA screener to match with the census data on household residents. Criteria were set up to define what was considered a successful match. Person-level matching was attempted for 3,793 persons for whom census data were available. Of these, 3,392 (89.4 percent) were successfully matched. Matching success rates were generally higher for persons in households in which the date of NHSDA screening was close to April 1, the census date. Matching rates were lowest in large apartment buildings and highest in single-family units. These patterns suggest that the nonmatches are largely due to true changes in the household composition at the address over time rather than to errors in matching at the household level.

Special weights appropriate for this analysis were created, taking into account the probabilities of selection for households in the NHSDA and the subselection of cases to be matched.

Another component of this study was a questionnaire completed by NHSDA interviewers before and after NHSDA data collection. These questionnaires obtained information about expectations before data collection and reports of behaviors during data collection. Interviewer characteristics, including their expectations, behaviors, demographics, and experience, may affect the success rate in obtaining competed interviews.

Results

Survey response rates (both screening and interview) by type of nonresponse were tabulated by variables characterizing the location of the interview, the interviewer characteristics, the respondents' characteristics, and the household characteristics. The two types of nonresponse were noncontacts and refusals. Noncontacts at the screening level are cases where the interviewer was never able to contact anyone in the household, could not find an eligible screening respondent, or was denied access to the entire structure. Noncontacts at the interview level are cases where the interviewer was never able to reach the individual once he/she was selected for interview. It should be mentioned that these noncontacts at the screening and interview level may sometimes actually be passive refusals in which potential respondents are available but deliberately avoid interviewers because they do not want to participate.

These detailed tabulations, including standard errors, can be found elsewhere (Parsely 1993). Some examples are shown in tables 3 to 5. For each variable, these tables show the p value for a chi-square test of the significance of the relationship between the components of the response rate and the variable.

Households in structures containing 50 or more units (i.e., large apartment buildings or condominiums) had the highest screening noncontact rates. This reflects the commonly encountered high-security or controlled-access buildings where guards or lock systems prevent interviewers from accessing whole segments of housing units. Other household types with high screening noncontact rates were single adult households with children under 18 and households with never-married heads. The screening refusal rate in general is more constant across all groups than the noncontact rate. This is perhaps because the screening task requires minimal time and effort from households once contacted. Overall, the high screening response rate and the small differences among demographic groups suggest that screening nonresponse is not likely to result in biased drug use estimates in the NHSDA.

Because it is more frequent than screening nonresponse, interview nonresponse has greater potential for causing bias. Tables 4 and 5 show interview noncontact and refusal rates for selected household and person-level characteristics. Age was related to interview nonresponse, with refusal rates increasing as age increased. An exception to this trend was the higher rate among 12- to 17-year-olds, reflecting the fact that parents

TABLE 3. *Screening completion, noncontact, and refusal rates in 1990 NHSDA from Census Match Study.*

	N	% complete	% noncontact	% refusal
Number of units in structure (χ^2 = 35.05, P = 0.000)				
Mobile home	169	99.31	0.32	0.29
Single family	2,914	95.46	0.53	0.93
2 - 9 apartments	553	97.56	1.24	0.81
10 - 49 apartments	400	95.62	2.43	1.41
50+ apartments	339	93.93	4.98	0.87
Tenure (χ^2 = 29.15, P = 0.000)				
Owner	2,586	98.53	0.54	0.86
Renter	1,642	96.90	1.80	0.97
Urban/rural status (χ^2 = 38.84, P = 0.000)				
Large urban, including suburbs, pop. > 50,000)	3,651	97.25	1.38	1.17
Small urban (pop. 2,500-50,000)	268	99.14	0.16	0.54
Rural (pop. < 2,500)	555	99.26	0.27	0.43
HH composition (χ^2 = 76.51, P = 0.000)				
Single person	726	97.19	1.77	0.77
Single adult, child under 18	259	95.94	2.78	1.17
Couple, child under 18	1,062	98.63	0.35	0.90
Couple, no children	526	98.23	0.80	0.90
Couple or single, child over 18	489	98.84	0.22	0.94
At least one non-nuclear	502	98.74	0.46	0.62
At least one non-relative	502	97.84	0.92	1.09
Race/ethnicity of HH head (χ^2 = 10.47, P = 0.334)				
Hispanic	488	97.66	1.18	0.97
Non-Hispanic, white	2,583	98.25	0.78	0.88
Non-Hispanic, black	956	97.22	1.70	0.97
Non-Hispanic, other	307	97.52	0.82	0.46

TABLE 3. *Screening completion, noncontact, and refusal rates in 1990 NHSDA from Census Match Study (continued).*

	N	% complete	% noncontact	% refusal
Education of HH head (χ^2 = 27.38, P = 0.004)				
Less than high school	142	98.72	0.69	0.28
High school graduate	159	98.54	0.46	0.93
Some college	133	98.85	0.48	0.67
College degree and higher	175	97.61	1.35	1.05
Marital status of HH head (χ^2 = 22.95, P = 0.003)				
Married	2,446	98.52	0.47	0.91
Widowed/divorced/separated	1,037	97.98	1.01	0.83
Never married	772	96.50	2.50	0.75

KEY: HH = household.

sometimes refuse to allow their teenager to participate. The lowest refusal rate was among 18- to 24-year-olds, who generally have the highest rates of drug use. Refusal rates varied by race, with whites being the most likely to refuse. There was little variation in noncontact rates among racial groups. Divorced/separated people had the lowest response rate. There was no significant relationship between education and interview response rates.

Nonresponse rates were significantly related with housing value and rents, with noncontact rates generally increasing with higher values and rents. Noncontact and refusal rates were highest in large urban areas (including suburbs).

By linking all of the data from the various components of the Census Match Study, it was possible to describe the characteristics of sampled persons, households, neighborhoods, and interviewers for NHSDA cases, including both nonrespondents and respondents. This provided the opportunity to develop an overall model of nonresponse that could take into account all of these factors. The final logistic regression model, including only significant (at the 0.1 level) predictors of propensity of a sampled person to participate in the NHSDA, is shown in table 6. The

TABLE 4. *Interview completion, noncontact, and refusal rates in 1990 NHSDA from Census Match Study: person-level variables.*

	N	% complete	% noncontact	% refusal
Age ($\chi^2 = 81.60$, P = 0.000)				
12 - 17	579	88.24	2.24	7.96
18 - 24	594	88.64	4.89	4.87
25 - 34	903	86.15	3.81	8.03
35 - 50	609	83.59	4.91	10.16
Over 50	536	84.15	2.17	9.77
Sex ($\chi^2 = 7.95$, P = 0.056)				
Male	1,573	83.44	4.75	9.84
Female	1,774	86.38	3.11	8.00
Race ($\chi^2 = 20.92$, P = 0.005)				
White	2,427	84.69	3.94	9.35
Black	684	87.36	3.49	6.68
Other	139	83.78	3.84	3.69
Hispanic origin ($\chi^2 = 5.27$, P = 0.164)				
No	2,526	85.25	3.52	9.23
Yes	546	86.21	4.16	6.57
Marital status ($\chi^2 = 28.31$, P = 0.003)				
Married	1,180	85.67	3.24	9.66
Widowed	150	83.98	2.59	5.84
Divorced/separated	327	82.54	5.43	10.08
Never married	1,649	85.10	4.49	7.84
Education ($\chi^2 = 12.45$, P = 0.427)				
8th grade or less	82	89.87	1.45	7.07
9th - 12th grade, no diploma	106	86.05	3.64	6.80
High school graduate	118	89.96	0.24	8.70
Some college	99	89.26	2.26	6.75
College degree and higher	66	90.69	2.14	6.28

TABLE 4. *Interview completion, noncontact, and refusal rates in 1990 NHSDA from Census Match Study: person-level variables (continued).*

	N	% complete	% noncontact	% refusal
Worked last week ($\chi^2 = 7.06$, P = 0.080)				
Yes	236	87.92	2.03	9.04
No	178	90.16	1.39	5.56

TABLE 5. *Interview completion, noncontact, and refusal rates in 1990 NHSDA from Census Match Study: household-level variables.*

	N	% complete	% noncontact	% refusal
House value (owners) ($\chi^2 = 100.24$, P = 0.000)				
Less than $20,000	120	90.72	1.70	6.35
$20,000 - 39,999	179	86.21	3.37	8.55
$40,000 - 59,999	272	87.56	2.50	8.09
$60,000 - 79,999	265	84.95	2.16	11.57
$80,000 - 99,999	239	82.77	3.29	11.73
$100,000 - 149,999	497	84.26	3.67	9.20
House value (owners)				
$150,000 - 199,999	306	81.17	6.03	9.31
$200,000 - 299,999	243	77.56	8.18	12.02
$300,000+	132	80.57	7.88	10.88
Monthly rent (renters) ($\chi^2 = 47.61$, P = 0.002)				
Less than $200	138	87.23	2.80	6.87
$200 - 299	163	88.36	2.98	6.12
$300 - 399	225	87.84	2.01	6.41
$400 - 499	205	86.20	3.77	7.23
$500 - 599	185	87.43	2.47	7.04
$600 - 699	141	80.14	10.61	5.67
$700+	268	76.77	6.50	10.92

TABLE 5. *Interview completion, noncontact, and refusal rates in 1990 NHSDA from Census Match Study: household-level variables (continued).*

	N	% complete	% noncontact	% refusal
Urban/rural status ($\chi^2 = 50.29$, P = 0.000)				
Large urban, including suburbs (pop. > 50,000)	3,103	80.28	5.41	10.79
Small urban (pop. 2,500-50,000)	158	88.60	1.51	7.76
Rural (pop. < 2,500)	346	90.32	2.86	5.89

person-level factors indicate that males are less likely to participate, while Hispanics are more likely to participate in the NHSDA. Owner-occupied households, homes with greater value or rent, and households consisting of single persons over age 65 were less likely to participate. Households with children under 5 years of age were more likely to participate. Block-level characteristics associated with response propensity were urbanicity (persons living in urban areas were less likely to participate) and the percent of housing units boarded up (lower response rate when more units were boarded up). Several of these person-, household-, and block-level factors are included in statistical models used to adjust for nonresponse in NHSDA analysis weights. Because these factors are significantly associated with nonresponse, they are likely to be effective in reducing nonresponse bias in NHSDA estimates.

Several interviewer characteristics were significantly associated with interview completion rates. Hispanic interviewers and interviewers with a household income of less than $30,000 were less likely to obtain completed interviews. (The positive effect of Hispanic respondents and negative effect of Hispanic interviewers reflects the poor completion rates by Hispanic interviewers at non-Hispanic households.) Interviewers who had worked on any household surveys in the past 3 years were more likely to obtain completed interviews. Interviewers' attitudes and behavior were found to be associated with success in the field as well. Interviewers who always left a copy of the NHSDA advance letter when visiting a household and finding no one at home were more likely to complete interviews. Interviewers who felt that "with enough effort, I can convince even the most reluctant respondent to participate" were more

TABLE 6. *Final multilevel logistic model of response propensity in 1990 NHSDA from Census Match Study (N = 3,201).*

Variable	Coefficient	Std. error	P	Adjusted odds ratio
Intercept	2.653271	0.275089	0.000	---
Person level				
Male	-0.317930	0.124322	0.013	0.7277
Hispanic	0.622700	0.221535	0.007	1.8640
Household level				
Owner	-0.531859	0.197145	0.009	0.5875
House value[1]	-0.000001	0.000001	0.035	0.9900
Monthly rent[2]	-0.001001	0.000333	0.004	0.9512
Children under 5 years	0.394853	0.163665	0.019	1.4842
Single person over 65	-0.771967	0.309151	0.015	0.4621
Block level				
Urban	-0.819406	0.176381	0.000	0.4407
Percent boarded up[4]	-0.062186	0.033672	0.069	0.9876
Interviewer level				
Hispanic	-0.746810	0.282162	0.010	0.4739
Recent experience[5]	0.323332	0.110695	0.005	1.3817
Advance letter[6]	0.348507	0.122957	0.006	1.4170
Income < $30,000[7]	0.298085	0.113771	0.011	0.7422
Can convince anyone[8]	0.242386	0.117690	0.044	1.2743
Personally affect[9]	0.449398	0.163389	0.008	1.5674

KEY: 1 = Based on the census question for owners, "What is the value of this property?" 2 = Based on the census question for renters, "What is the monthly rent?" 3 = Includes large or small urban. 4 = Percent of boarded up housing units in block. 5 = Worked on other household surveys in past 3 months. 6 = "Always" leave copy of advance letter at HH when no one is home. 7 = Family income. 8 = Agreed that "With enough effort, I can convince even the most reluctant respondent to participate." 9 = "Always" informed respondents how NHSDA results could affect them personally.

successful in completing interviews, as were interviewers who reported that they always informed the respondent of how the survey results could affect them personally. The significance of these interviewer-level variables in the model suggests potential areas for interviewer training that could lead to improved response rates in other surveys.

DISCUSSION

Drug abuse surveys are particularly vulnerable to nonresponse and measurement error because of the difficulties in accessing heavy drug users and the likelihood that the illegal and stigmatized nature of drug abuse may lead to underreporting. The Skip Pattern Experiment confirms once again that respondents' reporting of their drug use behavior is highly sensitive to the conditions under which they report. This conclusion makes clear the need to proceed with great caution in interpreting differences in drug use rates obtained in different surveys. It also suggests caution in the implementation of new technologies such as computer-assisted data collection that undoubtedly will have some as yet unknown effect on respondents' willingness to report their drug use. The Skip Pattern Experiment may have implications in the introduction of these new technologies, as one of the advantages of computer-assisted interviewing is the ease with which skips can be implemented.

The Skip Pattern Experiment and previous studies of measurement error are useful for indicating differences in measurement error caused by different data-collection methods. However, none of these studies provides what is most needed: a measure of the overall underreporting level. There is also a critical need for a well-designed study of the validity of self-reported drug use in the household population. New technologies for obtaining criterion measures of drug use, such as hair testing, may provide the means for conducting such a study.

The Census Match Study demonstrates that response rates are not constant across various interviewer, respondent, household, and neighborhood characteristics. To the extent that rates of drug use vary by these same characteristics, bias due to nonresponse may be a problem. However, it is not always the case that low response rates occur in conjunction with high drug use prevalence (table 7). Some populations with low response rates (e.g., older adults and high-income populations) tend to have low rates of drug use. On the other hand, some populations

TABLE 7. *Relationship between interview response rate and illicit drug use rate.*

		Interview response rate	
		High	Low
Illicit drug use rate	High	Age 18-34 Low income Black	Male Large metro
	Low	Female Married College degree Rural	Age 35+ High income White

(e.g., large metro residents and men) have low response rates and high drug use rates. In estimating overall prevalence, many of these potential sources of bias would be in opposite directions and would therefore tend to cancel each other.

The potential biases suggested by the Census Match analysis are also reduced somewhat by nonresponse adjustments built into drug use prevalence estimates. The extensive knowledge gained from studies of nonresponse patterns has led to improvements in these statistical adjustments. The NHSDA now utilizes a sophisticated propensity of response logistic regression model to adjust estimates (Folsom 1991). This model incorporates many of the variables found by the Census Match Study to be associated with survey participation, such as block-level rent and housing values. Since screening response rates in the NHSDA are high, the response propensity model can also take advantage of the screening data on the characteristics of most interview nonrespondents. This basic data collected on the 1995 NHSDA screening form includes age, gender, race/ethnicity, marital status, and current smoking status.

Nonresponse adjustments, however, are only a partial solution to the problem of reducing nonresponse bias. By taking advantage of available auxiliary data known to be related to drug use rates and to response propensity, statistical adjustments undoubtedly improve the accuracy of NHSDA drug use prevalence estimates. But auxiliary data and statistical correlations are only proxies for the true drug use data for nonrespondents.

Achievement of a high response rate remains the most important goal and the most effective method of reducing nonresponse bias.

REFERENCES

Caspar, R. Followup of nonrespondents in 1990. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Rockville, MD: National Institute on Drug Abuse, 1992.

Cox, B.; Witt, M.; Traccarella, M.; and Perez-Michael, A. Inconsistent reporting of drug use in 1988. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Rockville, MD: National Institute on Drug Abuse, 1992.

Folsom, R.E. Exponential and logistic weight adjustments for sampling and nonresponse error reduction. In: *Proceedings of the American Statistical Association, Survey Research Methods Section*. Washington, DC: American Statistical Association, 1991.

Forsyth, B.; Lessler, J.; and Hubbard, M. Cognitive evaluation of the questionnaire. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Rockville, MD: National Institute on Drug Abuse, 1992.

Gfroerer, J. Influence of privacy on self-reported drug use by youths. In: Rouse, B.; Kozel, N.; and Richards, L., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity*. National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985.

Gfroerer, J. "An Overview of the National Household Survey on Drug Abuse and Related Methodological Research." Paper presented at Joint Statistical Meetings, Boston, August 1992.

Gfroerer, J., and Hughes, A. Collecting data on illicit drug use by phone. In: Turner, C.F.; Lessler, J.T.; and Gfrorer, J., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Rockville, MD: National Institute on Drug Abuse, 1992.

Groves, R.M. Research on survey data quality: Survey research as a methodology without a unifying theory. *Pub Opin Q* 51:S156-S172, 1987.

Groves, R.M., and Couper, M.P. "Correlates of Nonresponse in Personal Visit Surveys." Paper presented at the 1992 Joint Statistical Meetings, Boston, August 1992.

Harrell, A.V.; Kapsak, K.A.; Cisin, I.H.; and Wirtz, P.W. "The Validity of Self-Reported Drug Use Data: The Accuracy of Responses on Confidential Self-Administered Answer Sheets." Report of activities under contract no. 271-85-8305 prepared for the National Institute on Drug Abuse, 1986.

Lessler, J.T., and Durante, R.C. "1992 NHSDA, Findings of First Quarter Skip Test." Final report of activities under contract number 271-91-5402 prepared for the Substance Abuse and Mental Health Services Administration, 1992.

Parsley, T.L. "Report on 1990 NHSDA-Census Match." Final report of activities under contract no. 271-89-8333 prepared for the Substance Abuse and Mental Health Services Administration, 1993.

Substance Abuse and Mental Health Services Administration. *National Household Survey on Drug Abuse: Main Findings 1991*. DHHS Pub. No. (SMA)93-1980. Rockville, MD: Substance Abuse and Mental Health Services Administration, 1993.

Substance Abuse and Mental Health Services Administration (SAMHSA). *Preliminary Results from the 1993 National Household Survey on Drug Abuse*. Advance Report No. 7. Rockville, MD: SAMHSA, 1994.

Turner, C.; Lessler, J.; and Devore, J. Effects of mode of administration and wording on reporting of drug use. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Rockville, MD: National Institute on Drug Abuse, 1992.

AUTHORS

Joseph Gfroerer
Chief, Prevalence Branch
Office of Applied Studies
Substance Abuse and Mental Health Services Administration
Room 16C-06
5600 Fishers Lane
Rockville, MD 20857

Judith Lessler, Ph.D.
Research Director, Statistics Research Division
Research Triangle Institute
Ragland Building 020
P.O. Box 12194
Research Triangle Park, NC 27709

Teresa Parsley
Research Triangle Institute
Hill Building 202
P.O. Box 12194
Research Triangle Park, NC 27709

167352

Adaptive Sampling in Behavioral Surveys

Steven K. Thompson

ABSTRACT

Studies of populations such as drug users encounter difficulties because the members of the populations are rare, hidden, or hard to reach. Conventionally designed large-scale surveys detect relatively few members of the populations so that estimates of population characteristics have high uncertainty. Ethnographic studies, on the other hand, reach suitable numbers of individuals only through the use of link-tracing, chain referral, or snowball sampling procedures that often leave the investigators unable to make inferences from their sample to the hidden population as a whole. In adaptive sampling, the procedure for selecting people or other units to be in the sample depends on variables of interest observed during the survey, so the design adapts to the population as encountered. For example, when self-reported drug use is found among members of the sample, sampling effort may be increased in nearby areas. Types of adaptive sampling designs include ordinary sequential sampling, adaptive allocation in stratified sampling, adaptive cluster sampling, and optimal model-based designs. Graph sampling refers to situations with nodes (for example, people) connected by edges (such as social links or geographic proximity). An initial sample of nodes or edges is selected and edges are subsequently followed to bring other nodes into the sample. Graph sampling designs include network sampling, snowball sampling, link-tracing, chain referral, and adaptive cluster sampling. A graph sampling design is adaptive if the decision to include linked nodes depends on variables of interest observed on nodes already in the sample. Adjustment methods for nonsampling errors such as imperfect detection of drug users in the sample apply to adaptive as well as conventional designs.

INTRODUCTION

Surveys to estimate human behavioral characteristics such as drug use encounter a number of inherently difficult sampling and estimation problems. Among the factors making sampling and estimation difficult

for such populations are the rarity and geographic unevenness of some of the populations of interest, the elusiveness or hiddenness of individuals in the population, and the variability of behaviors within and between subpopulations. In addition, the sensitive nature of the behaviors of interest gives rise to nonsampling errors including nonresponse and underreporting of stigmatized behaviors. Each of these issues arises in design and implementation of a large survey on drug use such as the National Household Survey on Drug Abuse as well as in ethnographic studies focusing on specific drug-using populations (cf., Lambert 1990; Turner et al. 1992*c*; Weppner 1977). Similar sampling and estimation problems arise in surveys of persons infected with rare diseases, populations defined by sexual orientation or behavior, sex workers and others involved in underground economic activities, homeless people, and other underrepresented groups. Some of the statistical problems arising with human populations have arisen also in environmental and biological surveys so that methods first developed for one area have subsequently been applied to another (Freedman 1991; Kalton and Anderson 1986; Thompson 1992; Wolter 1986, 1991).

Conventionally designed large-scale surveys detect relatively few members of rare or hidden populations so that estimates of population characteristics have high uncertainty. For example, the original impetus for the national Health and Social Life Survey (Laumann et al. 1994), a national probability sample survey of sexual behavior, was in large part provided by concern regarding the acquired immunodeficiency syndrome (AIDS) epidemic. Funding constraints limited the sample size to 3,432 people, and the number of people in the sample who reported having tested positive for the human immunodeficiency virus (HIV) was only 6.

Ethnographic studies and other studies focusing on the behaviors of people in rare and hidden populations find suitable numbers of individuals for their samples only through the use of link-tracing, chain referral, or snowball sampling procedures that usually leave the investigators unable to make inferences from their sample to the hidden population as a whole. For example, to investigate sex-for-crack exchanges in Philadelphia and Newark (French 1993), the sample of 100 crack cocaine users was obtained by the investigators going to known drug-dealing areas and talking with persons they believed were users or dealers, who then referred the investigators to other users. To analyze how some opiate addicts had overcome their addiction on their own, Biernacki (1986) used referral chains. Finding starting points for

the chains was difficult, however, for the people of interest had little contact with treatment centers or social service agencies.

One approach to increasing the sample representation of members of a rare or hidden population while still obtaining unbiased estimates of population characteristics is through the use of adaptive sampling procedures. In adaptive sampling, the selection of people or other units to include in the sample adapts to observations made during the survey. In particular, whenever "interesting" values are observed, sampling intensity may be adaptively increased for neighboring or linked units. Examples of interesting values that could be specified include reported drug use, involvement in underground economic activities, high-risk sexual behaviors, or a positive HIV test result. In a spatial context, additional units may be added to the sample from the geographic vicinity of any unit in which an interesting value is observed. Linkages such as social contact or kinship could be used in place of spatial proximity. The condition for adaptively adding units can be based either on the variable of interest, such as self-reported drug use, or on an auxiliary variable such as tobacco use. For some populations, adaptive designs can produce gains in efficiency, relative to conventional designs, for estimating the population mean or total. In addition, adaptive sampling designs can substantially increase the yield of interesting units in the sample.

Graph sampling refers to situations with nodes connected by edges. In studies of hidden populations, the nodes may represent individual people and the edges represent social links or geographic proximity between people. An initial sample of nodes or edges is selected and edges are subsequently followed to bring other nodes into the sample. Graph sampling designs include network sampling, snowball sampling, link-tracing, chain referral, and adaptive cluster sampling. A graph sampling design is adaptive if the decision to include linked nodes depends on variables of interest observed in nodes already in the sample.

In this chapter, the use of adaptive sampling and graph sampling methods for studies of behavioral characteristics in rare and hidden populations is examined. Methods of adjusting for the nonsampling errors that arise in such studies are discussed for both adaptive and conventional designs.

ADAPTIVE SAMPLING STRATEGIES

For populations that are rare, unevenly distributed, hidden, or hard to reach, conventional sampling designs such as simple random sampling lead to estimates with high variances and potential biases. With sufficient previous knowledge of the population, precision can be increased through such devices as stratification, systematic designs, and use of auxiliary information in the design and estimation stages (Cochran 1977; Thompson 1992).

Often, however, the uneven patterns in the populations cannot be predicted before the survey. For example, patterns of drug use may change over time, epidemics progress through cycles, neighborhoods may change their compositions, and economic changes occur; similarly, natural populations of animals or fish may change unpredictably in spatial pattern. For such populations, adaptive sampling strategies can be useful.

Adaptive sampling designs are those in which the procedure for selecting units to include in the sample may depend on values of the variable of interest observed during the survey. For spatially clustered populations, additional observations may be added in the neighboring vicinity whenever high abundance is encountered. Whenever an infected person appears in a survey sample of a rare, contagious disease, close contacts of that person might be added to the sample. In a drug use study, sampling intensity could be adaptively increased in the neighborhood of respondents with self-reported use.

Types of Adaptive Designs

Adaptive sampling designs include sequential stopping designs, adaptive allocation designs, optimal model-based strategies, and adaptive cluster sampling.

With sequential stopping designs, sampling continues until a given criterion, based for example on observed incidence or sample variance, is attained. The procedure may be based on sequentially observing the variable of interest and evaluating the criterion as each unit is selected, or it may be based on batches of units. Much of the statistical literature in sequential analysis (Chernoff 1972; Siegmund 1985; Wald 1947; Woodroofe 1982) concerns sequential stopping problems, in which the size of a random sample is determined sequentially from the observed values.

With adaptive allocation designs, an initial stratified sample is selected. Based on the observed values for the initially selected units, an additional stratified sample is selected with allocation of sample sizes to strata based on the initial observations. For example, after an initial stratified sample is obtained in a drug use epidemiological survey, the remaining sample sizes could be allocated to give larger sample sizes in strata showing high reported drug use in the initial sample. For adaptive allocation designs, the strata may be defined either spatially or through sociological variables. With adaptive allocation, the usual stratified estimate is not unbiased for the population total. Unbiased adaptive allocation strategies are described in Kremers (1987), Thompson and associates (1992), and Thompson and Seber (1996). Other adaptive allocation strategies are described in Solomon and Zacks (1970), Francis (1984), Gasaway and colleagues (1986), and Jolly and Hampton (1990).

Optimal model-based sampling strategies are often adaptive ones. Much of survey sampling practice is design based. That is, no assumptions are made about the population itself, and properties such as unbiasedness of estimates are calculated over all possible samples that might have been selected. In the model-based approach to sampling, a probabilistic model is assumed for the population. For example, the values of the variable of interest may be assumed to have a multivariate lognormal distribution, with units that are close to each other either geographically or socially having positive correlation. For many assumed population models, the theoretically optimal sampling strategy can be shown to be an adaptive one (Thompson 1988; Thompson and Seber 1996; Zacks 1969). However, the theoretically optimal strategies are not necessarily the most practical because they may require an unattainable amount of previous information about the population and tend to be computationally and implementationally complex (Solomon and Zacks 1970). Simpler designs that are model based and have some features of optimality have been applied to environmental sampling problems by Englund and Herari (in press) and by Geiger (1994).

In adaptive cluster sampling, an initial sample is selected based on a conventional sampling design such as simple random sampling, systematic sampling, cluster sampling, or stratified sampling. Whenever the variable of interest of a unit in the sample satisfies a specified condition, units in the neighborhood of that unit are added to the sample. If in turn any of the added units satisfies the condition, still more units are added, and so on. For example, neighboring units may be added whenever the variable of interest has a high or interesting value, such as reported drug use,

incidence of a disease, sexual behaviors of interest to the study, high economic activity, high observed occurrence of homelessness, high abundance of an animal species, or high concentration of a pollutant. A variety of adaptive cluster sampling designs are described in Thompson (1990, 1991*a*, 1991*b*, 1992, 1993, 1994*a*, in press), Seber and Thompson (1993), Thompson and Seber (1996), and Munholland and Borkowski (1993). The use of adaptive cluster sampling to estimate the prevalence of insect infestation in forest trees is described by Roesch (1993). Adaptive cluster sampling of winter waterfowl populations is described by Smith and colleagues (1995). Adaptive cluster sampling for rare household characteristics is described by Danaher and King (1994).

The neighborhoods of adaptive cluster sampling may be defined spatially, as with the geographically neighboring sample sites of environmental and ecological surveys and with urban blocks and larger geographic regions in human surveys, but they also may be defined by social or institutional connections. For example, in a survey of a rare disease, the neighborhood of a person in the sample could be defined to include that person's siblings or close social contacts. Determining key social links of persons in the sample can present many methodological challenges. Wiebel (1990) reports good success in obtaining the identities of sexual partners of active intravenous drug users once the trust between outreach workers and subjects solidified. Additional effective methodologies for obtaining sensitive information on hidden populations are described in Adler (1990) and Goldstein and associates (1990). However, as pointed out below in the section on graph sampling, estimation in adaptive cluster sampling is based on the empirical or observable links, not on hidden or underlying links, so that unbiased estimation of a population total is possible even though some of the underlying links between sample respondents may remain hidden.

Typically, a survey is used to obtain estimates of more than one characteristic of interest. Generalizations of adaptive cluster sampling results to the multivariate case produce unbiased estimators of the population mean and total for each variable as well as unbiased estimators of variances and covariances (details are found in Thompson 1993). The results hold whether the condition for additional sampling depends on just one of the variables or on a function of all of them. The result giving conditions under which adaptive sampling produces more precise estimates than conventional sampling is also generalized to the multivariate case (Thompson 1993). Further, adaptive addition of units can be based on

an easier-to-measure auxiliary variable or one that is less sensitive, such as tobacco use, rather than the variable of interest, such as illegal drug use.

One problem with conventional survey designs when applied to populations with rare characteristics of interest, such as heroin use or HIV infection, is that typically few cases with the characteristic show up in the sample. Adaptive designs, in addition to potentially increasing the precision of survey estimates, have the potential to increase the yield of the sample in terms of the characteristic of interest. For example, the number of self-reported drug users or people who have tested positive for HIV in the sample may be increased by adaptively either increasing the allocation to strata in which such people are encountered or following social links from such people as in adaptive cluster sampling. The objective is to obtain data on more individuals of the rare population in order to more effectively carry out analytic studies such as evaluations of drug use outcomes, outcomes of treatment programs, and identification of risk factors. Indeed, without adaptively following leads and links from initially encountered individuals, it may not be possible to penetrate a hidden population sufficiently for study (Adler 1990; Frank and Snijders 1994). Importantly, unbiased estimates of population totals and other parameters are still possible with such surveys even though the sample contains members of the target population in higher proportion than the population as a whole.

Unbiased Estimation With Adaptive Designs

Estimators that are unbiased with conventional sampling designs may be no longer unbiased with adaptive designs. For example, adaptive cluster sampling typically produces a sample with a higher than representative yield of the variable of interest—more birds or whales sighted, more persons reporting drug use, or more persons infected with the disease—than would occur with a random sample. With such a sample, the conventional expanded sample mean would tend to overestimate the total in the population. Fortunately, unbiased estimators are available for use with adaptive designs. The simplest of these estimators are design unbiased, meaning that the unbiasedness is based on the way the sample is selected and does not depend on any assumptions about the population itself.

Suppose the initial sample consists of a simple random sample of n units. For a unit in the initial sample whose y-value is observed to satisfy the specified condition, the units in its neighborhood are added to the sample. For any of those added units satisfying the condition, the neighboring

units are added, and so on. For any of the added units not satisfying the condition, on the other hand, no neighboring units are added. Thus, in the final sample associated with each initial unit that satisfied the condition, there is a network of units satisfying the condition and a number of added edge units that do not satisfy the condition. Unbiased estimation in adaptive cluster sampling must deal with the fact that selection or inclusion probabilities cannot be determined from the sample data for every unit in the sample. Even so, simple unbiased estimates can be computed.

The simplest of the unbiased estimates is obtained by averaging the y-values within networks (but excluding edge units). Let w_i be the average of the y-values for the network associated with the i-th unit of the initial sample. Any unit not satisfying the condition is considered a network of size one. An unbiased estimate of the population mean is given by

$$\hat{\mu} = (w_1 + w_2 + \ldots + w_n) / n$$

An unbiased estimator that is only slightly more complicated to compute but that in practice tends to be more efficient than that shown above is obtained by computing for each network intersected by the initial sample the probability a_κ of that network being intersected. Edge units are again ignored. Suppose that κ networks have been intersected by the initial sample and let y^*_κ denote the total of the y-values in the κ-th network. The unbiased estimator is where N is the number of units in the population.

$$\hat{\mu} = \left(\frac{y^*_1}{\alpha_1} + \frac{y^*_2}{\alpha_2} + \ldots + \frac{y^*_\kappa}{\alpha_\kappa} \right) / N$$

An unbiased estimate of the population total is obtained by multiplying the estimate of the mean by N. Unbiased estimates are also available for adaptive cluster sampling with other initial designs such as cluster, systematic, and stratified sampling. Unbiased estimates of variances are also readily computed. The efficiency of the above estimators can be

improved using the Rao-Blackwell method, so that edge units receive some weight in the estimates, but the improved estimates are more complicated to compute. Full details on estimation with adaptive cluster sampling are given in Thompson (1992) and Thompson and Seber (1996).

Efficiency of Adaptive Sampling

For some populations, particularly those that are rare and clustered, adaptive sampling strategies have been found to produce remarkable increases in precision or efficiency compared to conventional sampling designs of equivalent sample size. In addition, adaptive designs can significantly increase the yield of interesting observations in the sample. Efficiency comparisons for specific populations are given in Thompson (1990, 1991*a*, 1991*b*, 1992, 1994*a*), Roesch (1993), Thompson and associates (1992), Francis (1984), and Smith and colleagues (1995). Factors influencing the relative efficiency of adaptive cluster sampling to simple random sampling are described in Thompson (1994*a*) and summarized below. The efficiency of an adaptive cluster sampling design compared to a conventional design for household surveys of rare characteristics was estimated using a trial survey by Danaher and King (1994).

The relative efficiency of adaptive cluster sampling to simple random sampling depends on characteristics of the population, the design, and the cost of sampling. Any of the following characteristics tend to increase the efficiency of adaptive cluster sampling relative to conventional random sampling: (1) the within-network variance is a high proportion of the total population variance (i.e., the population is clustered or aggregated with high variability within aggregations); (2) the population is rare; (3) the expected final sample size with adaptive cluster sampling is not much larger than the initial sample size; (4) the cost of observing units in clusters or networks is less than the cost of observing the same number selected at random; (5) the cost of observing units not satisfying the condition is less than the cost of observing units satisfying the condition; (6) the condition for extra sampling may be based on an auxiliary variable that is easy to measure; and (7) an efficient estimator or Rao-Blackwell improved estimator is used with the adaptive cluster sampling.

Because the final sample size depends on what is observed during the survey, practical measures are needed to ensure that the final sample size does not exceed the time or funding resources available for the survey. Ideally, a good choice of the criterion for extra sampling (as described

above) limits the adaptively added units to a relatively small proportion of the total. Further, if the population has been stratified, then the criterion in any stratum can be changed adaptively based on the time spent or observations made in previous strata, without affecting the unbiasedness of the estimates. Because of the design unbiased of the procedure within any stratum, the average value of the estimate over all possible samples equals the population mean for that stratum even though a previous stratum may have influenced the choice of the adaptive condition to be used. Thus, if time is running short halfway through the survey, the criterion can be made more stringent or adaptive sampling dispensed with completely for the remaining strata. If the stratification has not been done at the design stage, a pragmatic approach is to use poststratification at the estimation stage to approximate the same result. Thus, if adaptive sampling is discontinued part way through the survey, estimates can be poststratified with adaptive cluster sampling estimators used for that portion of the population in which the sampling was adaptive and the conventional estimator used in that portion in which the sampling was conventional. Other methods for limiting sample size include adding the extra units only for the top few values of the initial sample (Thompson, in press) and stopping sampling as soon as a specified total sample size has been reached (Brown 1994).

GRAPH-SAMPLING METHODS

Sampling methods such as network sampling, snowball sampling, chain referral sampling, adaptive cluster sampling, and other link-tracing designs in which investigators use links between people to find other people to include in the sample are examples of survey sampling in graphs. A directed graph consists of a set of nodes such as people or other units, and a set of edges linking some nodes to others. For two people (nodes), the links (edges) could be provided by physical proximity such as living on the same block, by hereditary relationship such as siblinghood, or by a social relationship. The edges can be directional so that two nodes i and j can be linked from i to j, from j to i, in both directions, or in neither direction. For example, individual i might provide investigators with the name of individual j, while user j either did not know or would not reveal individual j.

Associated with the i-th node is a variable of interest, y_i. For example, with nodes representing individual people, the variable of interest could be an indicator of cocaine use or dollar amount spent on heroin. The

basic problem in graph sampling is to select a sample of nodes or edges by some means and then estimate some population quantity, such as the total of the y-values, of the nodes or edges. The population quantities of interest could be number of cocaine users in the population, dollar amount spent on heroin, or average number of partners with whom needles are shared.

A graph sampling design is adaptive if decisions on whether to follow links depend on the observed y-values in the sample. For example, if an individual in the sample is asked to name sexual partners only if the individual reports intravenous drug use, the survey is adaptive, whereas it is not adaptive if every person sampled is asked to name sexual partners. The inherent links in the population, such as the sexual partners each individual would name if asked (regardless of intravenous drug use status), correspond to the neighborhood connections of adaptive cluster sampling in the spatial setting. The links that are followed, connecting groups of intravenous drug users, determine the networks of units that satisfy the condition in adaptive cluster sampling. By the previously cited results on multivariate adaptive cluster sampling, in which the adaptive condition could be based on an auxiliary variable rather than the variable of interest, it might be sensible to base the links on information that is less sensitive than drug use or sexual partners. For example, instead of being asked to name other cocaine users, individuals could be asked to name the people with whom they spend the most time.

Network sampling was introduced by Birnbaum and Sirken (1965) to estimate the number of people with a rare disease when a random sample of medical centers was selected. A person with the disease who had been treated at more than one center would have a higher probability of being included in the sample than a person who had been treated at only one center, so the configuration of such linkages needed to be taken into account to allow for unbiased estimation of prevalence. Subsequent uses of network sampling included surveys in which each person in the sample would be asked to report not only on themselves but also on persons, such as siblings, linked to them. A variety of linking rules and sampling designs have been investigated (Czaja et al. 1986; Faulkenberry and Garoui 1991; Kalton and Anderson 1986; Levy 1977; Nathan 1976; Sirken 1970, 1972*a*, 1972*b*; Sirken and Levy 1974; Sudman et al. 1988; Thompson 1992).

The term "snowball sampling" has been applied to a variety of graph sampling procedures (cf., Thompson 1992). In one type (Kalton and

Anderson 1986), members of a rare population in an initial sample are asked to identify other members of the population, those so identified are asked to identify others, and so on, for the purpose of obtaining a nonprobability sample or constructing a frame from which to sample. In another type (Goodman 1961), individuals in the sample are asked to identify a fixed number of other individuals, who in turn are asked to identify other individuals, for a fixed number of stages, for the purpose of estimating the number of mutual relationships or social circles in the population. Uses of snowball sampling for surveying drug users and other hidden populations are reviewed by van Meter (1990), who notes the difficulty of putting estimation on a sound statistical basis with such surveys without either assuming a specific stochastic process giving the original sample or the prohibitive requirement of knowing the linkage structure for the entire population. However, recent approaches to estimation with snowball and other graph samples (Frank 1977, 1979; Frank and Snijders 1994; Snijders 1992; Snijders et al. 1995; Spreen 1992; Spreen and Zwaagstra 1994) appear to be very promising. In addition, the estimation methods of network sampling and adaptive cluster sampling apply to a number of graph sampling situations.

In network sampling, the links generally are symmetric, or at least the links between units in the sample and those outside the sample are known. Further, the addition of linked units to the sample does not depend on observed values of the variable of interest. Thus, with network sampling it is possible to calculate the selection or inclusion probability of any person in the sample from the sample data. With that information, unbiased estimates of the population total or mean can be obtained, including the Horvitz-Thompson estimator and the multiplicity estimator (Birnbaum and Sirken 1965).

With the snowball sampling procedures described by Frank (1977, 1979), Frank and Snijders (1994), and Snijders (1992), and with adaptive cluster sampling, estimation is complicated by the fact that the selection or inclusion probabilities cannot generally be calculated from the sample data for every unit in the sample. This results from the asymmetry of some of the directional links, so that for some units (people) in the sample the investigators do not know how many other units would potentially have directed investigators to that unit. In graph sampling terminology, the in-degree of that unit is unknown (Frank 1977). An even more fundamental estimation difficulty for many snowball samples as obtained in practice is the lack of a well-defined probability sampling procedure for obtaining the initial sample.

The snowball sampling procedure described by Frank and Snijders (1994) for estimating the number of people in a hidden population illustrates both the possibilities and the difficulties. The design was used to estimate the number of cocaine users in a town in The Netherlands. The estimates obtained were in fact consistent with current police and social agency estimates. The initial sample of cocaine users was obtained not from a designed probability sample but from police and social service encounters. In the first wave, users in the initial sample were questioned for the names of other cocaine users, and the names edited to eliminate duplicates. In general, the survey procedure involves additional names from a second wave provided by the first wave people and so on, but in this particular study only the first wave was carried out. Estimates were then obtained based on a variety of assumptions. Because the initial sample was not obtained from a deliberate probability sampling procedure, estimation was based on the assumption that it arose as a Bernoulli procedure; that is, it is assumed that each individual in the hidden population had the same unknown probability of being included in the initial sample and inclusion was independent between individuals. For the model-based estimators, the additional assumption was made that directional links between individuals were independent and identically distributed Bernoulli random variables, so that for example whether individual A knows individual B is independent of whether individual B knows individual A. A design-based (subject to the Bernoulli assumption) estimator was also used, based on expanding the number of individuals added in the first wave by dividing that number by the proportion of individuals in the initial sample who were linked to any other individuals in the initial sample. One potential difficulty with such an estimator is the possibility that the proportion of initial sample units so linked would be zero, in which case that estimator could not be calculated.

The design unbiased estimates of adaptive cluster sampling could be used with a graph sampling or snowball procedure such as the one described above provided that the probability basis of the initial sample could be determined and that the addition of linked units was completed through all waves. The ideal would be to have a probability survey sample at the first stage and then follow through using the links provided by self-reported users in the survey sample. With such a survey, nonusers would be sampled along with users. For example, the variable of interest could be cocaine use, with $y_i = 1$ if the i-th individual is a user and $y_i = 0$ otherwise. The total of the y-values is then the total number of users in the population. The estimate would also need to be adjusted for inaccurate reporting as described in the next section. For estimation

purposes, a network would consist of a connected set of people, so that for any two people in the network it is possible to get from one to the other through the directed links. Any unit connected to the initial sample only by an asymmetric link—individual j who was reported by person i, but who when questioned similarly does not reveal individual i—would be treated as edge units in the estimation.

Interestingly, since estimation deals only with empirical or observable links and not with any underlying or true links, the unbiasedness of the adaptive cluster sampling estimates of the population total or mean is not affected by misreporting of links. For example, suppose in a survey of intravenous drug users links from each respondent consist of the names of sexual partners given by the respondent to investigators. Suppose that individual i has actual sexual partners j and k but reveals only k and not j to the investigators. The investigators then add individual k to the sample. If individual k in turn gives the name of i to investigators, then i and k are in the same network. Individual j remains outside the sample unless independently selected as part of the initial sample or of another network. Even if j gives the name of i to investigators, the two will not be in the same network because of the asymmetric link. Thus, for the design and estimation, the links from an individual are defined to be the names that person would give if included in the sample and asked, not necessarily the actual estimators. Although misreporting of the links would not bias the estimate of the population total, it would bias estimates of interest to epidemiologists regarding the actual pattern of sexual contacts in the population.

The use of link-tracing, chain referral, snowball, networking, or other graph-related sampling methods pervades the field of behavioral and ethnographic studies. Examples include studies of cocaine use and associated sexual behaviors (French 1993; Inciardi 1993), marijuana and cocaine dealing (Adler 1985), marijuana use (True and True 1977), heroin use (Agar 1977; Soloway and Walters 1977), street drug culture (Preble and Miller 1977), opiate addiction in women (Rosenbaum and Murphy 1990), recovery from addiction (Biernacki 1986), prostitution and drug use (McNamara 1994), pickpockets (Inciardi 1977), sexual behavior of selected ethnic or age-defined groups (Sterk-Elifson 1994; Thompson 1994*b*), and sexually transmitted diseases (Bailey and Aunger 1995). Further, even behavioral surveys producing a standard probability sample of individuals often seek estimates of network or graph-related characteristics of the population, such as the sexual networks relevant to sexually transmitted infections (Laumann et al. 1994).

Because of the prevalence of graph-related methods, adaptive and otherwise, for obtaining samples from rare and hidden populations, improvements in design and estimation methodologies for such studies are highly desirable. Some investigators who use snowball and other link-tracing designs to obtain a suitably large sample for study do not try to make inference from the sample to the population from which it comes but instead include in their study description a disclaimer that statistical inference is impossible or questionable. The disclaimer misses the point, however, because characteristics of the sample are commonly summarized as means or percentages, and the interpretation of the meaning of such means and percentages requires consideration of how the sample was obtained. Quite possibly, a more meaningful sample mean or percentage could be obtained using a weighted average with the weights reflecting the distinct networks in the sample. When a random initial sample is not possible to obtain, Snijders (1992) suggests drawing respondents as much as possible from independent sources to satisfy the assumptions of estimation methods as closely as possible.

NONSAMPLING ERRORS IN CONVENTIONAL AND ADAPTIVE SAMPLING

Because of the sensitivity of issues related to behavioral characteristics such as drug use, such surveys potentially involve prominent nonsampling errors related to incomplete candor in self-reporting, imperfect detectability of persons in high-risk groups, and other special factors in addition to the usual nonsampling errors associated with frame development, nonresponse, measurement, and data recording (Biemer et al. 1991; Lessler and Kalsbeek 1992). Sources of nonsampling variability in surveys of drug use and other sensitive behaviors include untruthful or incorrect self-reporting (Gfroerer et al. 1992; Rouse et al. 1985), inconsistent answers to survey questions (Cox et al. 1992), misinterpretations of questions by respondents (Forsyth et al. 1992), and item and unit nonresponse (Rubin 1987; Witt et al. 1992; Caspar 1992, Graham et al. 1994). Underreporting, self-selection bias, and other sources of nonsampling errors in sexual behavior surveys are reviewed in Clement (1990) and Berk and colleagues (1995). The role of nonsampling errors in general in surveys of sensitive topics is discussed in Turner and associates (1992*a*). Obtaining the best possible estimates from surveys involves, in addition to using a good sampling design, developing methods for reducing nonsampling errors and using methods to assess and adjust for nonsampling errors that do occur.

Adjusting for nonsampling errors can be illustrated by modeling incomplete self-reporting in a drug use survey as a problem in detectability. Assume a sample of households and people within households is selected according to the survey design, but not all drug users in the sample are detected by the self-reporting. Suppose for illustration that independent studies comparing self-reported use to bioassay results indicate that only half of users report use, so that the rate of detection is 50 percent. Then a simple form of adjusted estimate, whether of prevalence of use or of another variable such as amount spent on drugs during a 2-week period, is obtained by taking the naive estimate from the survey and dividing by one-half, so that the adjusted estimate is twice the initial estimate. The effects of such adjustments on survey estimates are analyzed in Thompson and Seber (1994). When the detection rates differ for different subpopulations or for different kinds of individuals, separate adjustments can be made for each observation based on the individual detectability rate that applies (Thompson and Seber 1994). Imperfect detectability in surveys has been estimated with such techniques as double sampling, distance sampling, and capture-recapture methods, as well as studies comparing self-reported and bioassay results.

With any conventional sampling design, nonsampling errors affect only the values recorded for units in the sample, while with an adaptive sampling design nonsampling errors may additionally affect what sample is selected. Potentially, the problem looks much more complicated for the adaptive design, but a conditioning argument given every possible sequence shows that adjustment and analysis methods are as straightforward for adaptive as for conventional designs (Thompson and Seber 1994).

The adjustment methods for imperfect detectability are required to produce realistic estimates of population characteristics, while the analysis is required to evaluate the effect of each source of sampling and nonsampling error and to determine the most effective means of improving estimates and reducing overall mean square error. In surveys of self-reported drug use, estimates of prevalence can be adjusted to account for the estimated proportion of users reporting no use, but the variance of the resulting estimates includes the following three important terms. The first is sampling variance due to the difference of one sample of households from another under the design. The second is a detectability error due to drug users reporting that they do not use. The third term is associated with the uncertainty in estimating the proportion of users who report no use; such estimates typically come from comparative and

criterion studies. To reduce the uncertainty associated with the first component involves improving sampling design, increasing sample sizes, and using more efficient estimation methods. Reduction of the second term requires interview methods that increase the accuracy of self-reporting among users. Such methods include question wording and interview mode of administration, such as face-to-face, computer, or telephone (Gfroerer and Hughes 1992; Schober et al. 1992; Turner et al. 1992*a*, 1992*b*). Reduction of the third term is achieved with larger, more specific, and more effective comparative and criterion studies, such as comparisons between self-reporting and the results of bioassays such as hair or urine tests.

ACKNOWLEDGMENT

Research on which this article is based was supported by the National Science Foundation.

REFERENCES

Adler, P. Ethnographic research on hidden populations: penetrating the drug world. In: Lambert, E.Y., ed. *The Collection and Interpretation of Data from Hidden Populations*. National Institute on Drug Abuse Research Monograph 98. DHHS Pub. No. (ADM)90-1678. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Adler, P. *Wheeling and Dealing, an Ethnography of an Upper-Level Drug Dealing and Smuggling Community*. New York: Columbia University Press, 1985.

Agar, M.H. Ethnography in the streets and in the joint. In: Weppner, R.S., ed. *Street Ethnography: Selected Studies of Crime and Drug Use in Natural Settings*. Beverly Hills, CA: Sage, 1977. pp. 143-155.

Bailey, R.C., and Aunger, R.V. Sexuality, infertility, and sexually transmitted disease among farmers and foragers in Central Africa. In: Abramson, P.R., and Pinkerton, S.D., eds. *Sexual Nature, Sexual Culture*. Chicago: University of Chicago Press, 1995. pp. 195-222.

Berk, R.; Abramson, P.R.; and Okami, P. Sexual activities as told in surveys. In: Abramson, P.R., and Pinkerton, S.D., eds. *Sexual Nature, Sexual Culture*. Chicago: University of Chicago Press, 1995. pp. 371-386.

Biemer, P.P.; Groves, R.M.; Lyberg, L.E.; Mathiowetz, N.A.; and Sudman, S. *Measurement Errors in Surveys*. New York: Wiley, 1991.

Biernacki, P. *Pathways from Heroin Addiction*. Philadelphia: Temple University Press, 1986.

Birnbaum, Z.W., and Sirken, M.G. *Design of Sample Surveys To Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates*. National Center for Health Statistics, Vital and Health Statistics, Series 2, No. 11. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1965.

Brown, J.A. The application of adaptive cluster sampling to ecological studies. In: Fletcher, D.J., and Manly, B.F.J., eds. *Statistics in Ecology and Environmental Monitoring*. Otago Conference Series No. 2. Dunedin, New Zealand: University of Otago Press, 1994. pp. 86-97.

Caspar, R. Followup of nonrespondents in 1990. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 155-173.

Chernoff, H. *Sequential Analysis and Optimal Design*. Philadelphia: SIAM, 1972.

Clement, U. Surveys of heterosexual behavior. (Translated by U. Dous.) *Ann Rev Sex Res* 1:45-74, 1990.

Cochran, W.G. *Sampling Techniques*. 3d ed. New York: Wiley, 1977.

Cox, B.; Witt, M.; Traccarella, M.; and Perez-Michael, A. Inconsistent reporting of drug use in 1988. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 109-154.

Czaja, R.F.; Snowdon, C.B.; and Casady, R.J. Reporting bias and sampling errors in a survey of a rare population using multiplicity counting rules. *J Am Stat Assoc* 81:411-419, 1986.

Danaher, P.J., and King, M. Estimating rare household characteristics using adaptive sampling. *N Z Stat* 29:14-23, 1994.

Englund, E.J., and Herari, N. Phased sampling for soil remediation. *Environm Ecolog Stat*, in press.

Faulkenberry, G.D., and Garoui, A. Estimating a population total using an area frame. *J Am Stat Assoc* 86:445-449, 1991.

Forsyth, B.; Lessler, J.; and Hubbard, M. Cognitive evaluation of the questionnaire. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM) 92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 13-52.

Francis, R.I.C.C. An adaptive strategy for stratified random trawl surveys. *N Z J Marine Freshwater Res* 18:59-71, 1984.

Frank, O. Survey sampling in graphs. *J Stat Plan Inference* 1:235-264, 1977.

Frank, O. Estimation of population totals by use of snowball samples. In: Holland, P.W., and Leinhardt, S., eds. *Perspectives on Social Network Research.* New York: Academic Press, 1979.

Frank, O., and Snijders, T. Estimating the size of hidden populations using snowball sampling. *J Official Stat* 10:53-67, 1994.

Freedman, D.A. Adjusting the 1990 census. *Science* 252:1233-1236, 1991.

French, J.F. Pipe dreams: Crack and the life in Philadelphia and Newark. In: Ratner, M.S., ed. *Crack Pipe as Pimp.* New York: Lexington, 1993. pp. 205-232.

Gasaway, W.C.; DuBois, S.D.; Reed, D.J.; and Harbo, S.J. "Estimating Moose Population Parameters from Aerial Surveys." Biological Papers of the University of Alaska (Institute of Arctic Biology) Number 22. Fairbanks, AK: University of Alaska, 1986.

Geiger, H.J. A Bayesian approach for estimating hatchery contribution in a series of salmon fisheries. *Alaska Fishery Res Bull* 1:66-75, 1994.

Gfroerer, J., and Hughes, A. Collecting data on illicit drug use by phone. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 277-295.

Gfroerer, J.; Gustin, J.; and Turner, C. Introduction. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 3-10.

Goldstein, P.J.; Spunt, B.J.; Miller, T.; and Bellucci, P. Ethnographic field stations. In: Lambert, E.Y., ed. *The Collection and Interpretation of Data from Hidden Populations.* National Institute on Drug Abuse Research Monograph 98. DHHS Pub. No. (ADM)90-1678. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990. pp. 80-95.

Goodman, L.A. Snowball sampling. *Ann Math Stat* 32:148-170, 1961.

Graham, J.W.; Hofer, S.M.; and Piccinin, A.M. Analysis with missing data in drug prevention research. In: Collins, L.M., and Seitz, L.A., eds. *Advances in Data Analysis for Prevention Intervention Research.* National Institute of Drug Abuse Research Monograph 142. NIH Pub. No. 94-3599. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1994. pp. 13-63.

Inciardi, J.A. In search of the class cannon: A field study of professional pickpockets. In: Weppner, R.S., ed. *Street Ethnography: Selected Studies of Crime and Drug Use in Natural Settings.* Beverly Hills, CA: Sage, 1977. pp. 55-77.

Inciardi, J.A. Kingrats, chicken heads, slow necks, freaks, and blood suckers: A glimpse at the Miami sex-for-crack market. In: Ratner, M.S., ed. *Crack Pipe as Pimp.* New York: Lexington, 1993. pp. 37-67.

Jolly, G.M., and Hampton, I. A stratified random transect design for acoustic surveys of fish stocks. *Can J Fisheries Aquatic Sci* 47:1282-1291, 1990.

Kalton, G., and Anderson, D.W. Sampling rare populations. *J Royal Stat Soc Ser A* 149:65-82, 1986.

Kremers, W.K. "Adaptive Sampling to Account for Unknown Variability among Strata." Preprint No. 128. Institut für Mathematik, Universitat Augsburg, Federal Republic of Germany, 1987.

Lambert, E.Y., ed. *The Collection and Interpretation of Data from Hidden Populations.* National Institute on Drug Abuse Research Monograph 98. DHHS Pub. No. (ADM)90-1678. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Laumann, E.O.; Gagnon, J.H.; Michael, R.T.; and Michaels, S. *The Social Organization of Sexuality, Sexual Practices in the United States.* Chicago: University of Chicago Press, 1994.

Lessler, J.T., and Kalsbeek, W.D. *Nonsampling Error in Surveys.* New York: Wiley, 1992.

Levy, P.S. Optimum allocation in stratified random network sampling for estimating the prevalence of attributes in rare populations. *J Am Stat Assoc* 72:758-763, 1977.

McNamara, R.P. *The Times Square Hustler: Male Prostitution in New York City.* Westport, CT: Praeger, 1994.

Munholland, P.L., and Borkowski, J.J. "Adaptive Latin Square Sampling + 1 Designs." Technical Report No. 3-23-93. Bozeman, MT: Department of Mathematical Sciences, Montana State University, 1993.

Nathan, G. An empirical study of response and sampling errors for multiplicity estimates with different counting rules. *J Am Stat Assoc* 71:808-815, 1976.

Preble, E., and Miller, T. Methadone, wine, and welfare. In: Weppner, R.S., ed. *Street Ethnography: Selected Studies of Crime and Drug Use in Natural Settings.* Beverly Hills, CA: Sage, 1977. pp. 229-248.

Roesch, F.A., Jr. Adaptive cluster sampling for forest inventories. *Forest Sci* 39:655-669, 1993.

Rosenbaum, M., and Murphy, S. Women and addiction: Process, treatment, and outcome. In: Lambert, E.Y., ed. *The Collection and Interpretation of Data from Hidden Populations*. National Institute on Drug Abuse Research Monograph 98. DHHS Pub. No. (ADM)90-1678. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Rouse, B.A.; Kozel, N.J.; Louise G.; and Richards, L.G., eds. *Self-Report Methods of Estimating Drug Use, Meeting Current Challenges to Validity*. National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985.

Rubin, D.B. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley, 1987.

Schober, S.; Fe Caces, M.; Pergamit, M.; and Branden, L. Effect of mode of administration on reporting of drug use in the national longitudinal survey. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 267-276.

Seber, G.A.F., and Thompson, S.K. Environmental adaptive sampling. In: Patil, G.P., and Rao, C.R., eds. *Handbook of Statistics*. Vol. 12, *Environmental Statistics*. New York: North Holland/Elsevier Science Publishers, 1993. pp. 201-220.

Siegmund, D. *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer, 1985.

Sirken, M.G. Household surveys with multiplicity. *J Am Stat Assoc* 63:257-266, 1970.

Sirken, M.G. Stratified sample surveys with multiplicity. *J Am Stat Assoc* 67:224-227, 1972*a*.

Sirken, M.G. Variance components of multiplicity estimators. *Biometrics* 28:869-873, 1972*b*.

Sirken, M.G., and Levy, P.S. Multiplicity estimation of proportions based on ratios of random variables. *J Am Stat Assoc* 69:68-73, 1974.

Smith, D.R.; Conroy, M.J.; and Brakhage, D.H. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics* 51:777-788, 1995.

Snijders, T.A.B. Estimation on the basis of snowball samples: How to weight. *Bull Methodologie Sociologique* 36:59-70, 1992.

Snijders, T.; Spreen, M.; and Zwaagstra, R. The use of multilevel modeling for analysing personal networks: Networks of cocaine users in an urban area. *J Quant Anthropol* 5:85-105, 1995.

Solomon H., and Zacks, S. Optimal design of sampling from finite populations: A critical review and indication of new research areas. *J Am Stat Assoc* 65:653-677, 1970.

Soloway, I., and Walters, J. Workin' the corner: The ethics and legality of ethnographic fieldwork among active heroin addicts. In: Weppner, R.S., ed. *Street Ethnography: Selected Studies of Crime and Drug Use in Natural Settings*. Beverly Hills, CA: Sage, 1977. pp. 159-178.

Spreen, M. Rare populations, hidden populations, and link-tracing designs; what and why? *Bull Methodologie Sociologique* 36:34-58, 1992.

Spreen, M., and Zwaagstra, R. Personal network sampling, outdegree analysis and multilevel analysis: Introducing the network concept in studies of hidden populations. *Int Sociol* 9:475-491, 1994.

Sterk-Elifson, C. Sexuality among African-American women. In: Rossi, A.S., ed. *Sexuality Across the Life Course*. Chicago: University of Chicago Press, 1994. pp. 99-126.

Sudman, S.; Sirken, M.G.; and Cowan, C.D. Sampling rare and elusive populations. *Science* 240:991-996, 1988.

Thompson, S. Changing lives, changing genres: Teenage girls' narratives about sex and romance, 1978-1986. In: Rossi, A.S., ed. *Sexuality Across the Life Course*. Chicago: University of Chicago Press, 1994*b*. pp. 209-232.

Thompson, S.K. Adaptive sampling. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*. Arlington, VA: American Statistical Association, 1988. pp. 784-786.

Thompson, S.K. Adaptive cluster sampling. *J Am Stat Assoc* 85:1050-1059, 1990.

Thompson, S.K. Adaptive cluster sampling: Designs with primary and secondary units. *Biometrics* 47:1103-1115, 1991*a*.

Thompson, S.K. Stratified adaptive cluster sampling. *Biometrika* 78:389-397, 1991*b*.

Thompson, S.K. *Sampling*. New York: John Wiley and Sons, 1992.

Thompson, S.K. Multivariate aspects of adaptive cluster sampling. In: Patil, G.P., and Rao, C.R., eds. *Multivariate Environmental Statistics*. New York: North Holland/Elsevier Science Publishers, 1993. pp. 561-572.

Thompson, S.K. "Factors Influencing the Efficiency of Adaptive Cluster Sampling." Technical Report 94-0301. University Park, PA: Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, 1994*a*.

Thompson, S.K. Adaptive cluster sampling based on order statistics. *Environmetrics*, in press.

Thompson, S.K., and Seber, G.A.F. Detectability in conventional and adaptive sampling. *Biometrics* 50:712-724, 1994.

Thompson, S.K., and Seber, G.A.F. *Adaptive Sampling*. New York: Wiley, 1996.

Thompson, S.K.; Ramsey, F.L.; and Seber, G.A.F. An adaptive procedure for sampling animal populations. *Biometrics* 48:1195-1199, 1992.

True, W.R., and True, J.H. Network analysis as a methodological approach to the study of drug use in a Latin city. In: Weppner, R.S., ed. *Street Ethnography: Selected Studies of Crime and Drug Use in Natural Settings*. Beverly Hills, CA: Sage, 1977. pp. 125-141.

Turner, C.F.; Lessler, J.T.; and Devore, J. Effects of mode of administration and wording on reporting of drug use. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992*a*. pp. 177-220.

Turner, C.F.; Lessler, J.T.; George, B.; Hubbard, M.; and Witt, M. Effects of mode of administration and wording on data quality. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992*b*. pp. 221-244.

Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C. Future directions for research and practice. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. National Institute of Drug Abuse. DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992*c*. pp. 299-306.

Van Meter, K.M. Methodological and design issues: Techniques for assessing the representatives of snowball samples. In: Lambert, E.Y., ed. *The Collection and Interpretation of Data from Hidden Populations*. National Institute on Drug Abuse Research Monograph 98. DHHS Pub. No. (ADM)90-1678. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Wald, A. *Sequential Analysis*. New York: Wiley, 1947.

Weppner, R.S. *Street Ethnography: Selected Studies of Crime and Drug Use in Natural Settings*. Beverly Hills, CA: Sage, 1977.

Wiebel, W.W. Identifying and gaining access to hidden populations. In: Lambert, E.Y., ed. *The Collection and Interpretation of Data from Hidden Populations*. National Institute on Drug Abuse Research Monograph 98. DHHS Pub. No. (ADM)90-1678. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Witt, M.; Pantula, J.; Folsom, R.; and Cox, B. Item nonresponse in 1988. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 85-108.

Wolter, K.M. Some coverage error models for census data. *J Am Stat Assoc* 81:338-346, 1986.

Wolter, K.M. Accounting for America's uncounted and miscounted. *Science* 253:12-15, 1991.

Woodroofe, M. *Nonlinear Renewal Theory in Sequential Analysis*. Philadelphia: SIAM, 1982.

Zacks, S. Bayes sequential designs of fixed size samples from finite populations. *J Am Stat Assoc* 64:1342-1349, 1969.

AUTHOR

Steven K. Thompson, Ph.D.
Associate Professor
Department of Statistics
Pennsylvania State University
326 Classroom Building
University Park, PA 16802

167353

Self-Reported Drug Use: Results of Selected Empirical Investigations of Validity

Yih-Ing Hser

ABSTRACT

This chapter reviews the literature on factors related to quality of self-report data on drug use and discusses two series of empirical studies investigating the quality of those data. One set of analyses examined the quality of the longitudinal retrospective self-report from narcotics addicts, including validity of recent narcotics use, reliability of various measures, stability of relationships among these measures, and pattern reliability among latent constructs. Results contribute strongly to confidence in the validity of the relationships among these data derived from addicts' self-report. The second set of analyses focused on validity of self-reported drug use among high-risk groups, including samples from sexually transmitted disease (STD) clinics, hospital emergency rooms (ERs), and jails. Results suggest that the accuracy of self-report of recent drug use varies by the sample sources, drug types, and subject characteristics. Targeting these high-risk groups may improve prevalence estimation. The chapter concludes that empirical validation of self-report is always necessary to enhance the utility of collected self-report data and provide means of controlling for potential biases.

INTRODUCTION

Surveys on drug use usually are conducted to establish estimates of prevalence rates or to improve the understanding of the relationships between drug use and related measures (e.g., antecedents, consequences, or intervention effects). Most general surveys rely on self-report or self-rating by participating subjects. But serious doubts have been cast on the truthfulness of data collected by self-report on sensitive topics related to stigmatized behaviors such as use of illicit drugs. Confidence in these data depends on their demonstrated validity and reliability, which must be empirically established. A related question pertains to the generalizability of results of general surveys, particularly when participation bias

(e.g., nonresponse or noncoverage) is known to be serious among groups at high risk for drug use. Therefore, reporting accuracy and sampling adequacy are among the most important concerns in the investigation of the quality of data on self-reported drug use.

This chapter presents several studies that address some of the issues involved in the investigation of reliability and validity of self-reported drug use. The first series of investigations illustrate several analytic approaches to examining the quality of longitudinal, retrospective self-reports from narcotics addicts, particularly when longer term external objective criteria are not available, as is often the case in such studies. Also described is an investigation that illustrates how prevalence estimation of drug use can be improved by targeting high-risk populations and examining the validity of their self-report to identify adjustment factors. Although these studies are of distinct natures and purposes, each offers some methodological approaches to improve the accuracy of data based on self-report. To provide background for the two studies, a brief literature review including substantive findings is presented, followed by a general discussion of analytical approaches that have been used in the empirical testing of validity of self-report data.

BACKGROUND

Findings from the Literature

Overall, the literature suggests that there is a high degree of variability in the validity of self-reported data according to differences in methodological and research context variables (Magura et al. 1987; Maisto et al. 1990). The validity of self-reported drug use may vary widely as a result of survey conditions, types of drug used, types of measure (e.g., frequency or amount), and characteristics of the sample population. There is an extensive body of research on the effects of data-collection methods (modes, interviewers) on respondent cooperation (Bradburn and Sudman 1988), but only recently has this research begun to focus on the assessment of drug use or other highly sensitive behaviors (Turner et al. 1992; Harrison 1995). This chapter, however, focuses on the types of error or bias that are attributable mainly to respondents themselves as opposed to external factors such as questionnaire construction or interview setting.

Generally speaking, respondent-based reporting errors may include memory failures, concealment of the less desirable aspects of one's life,

and overreporting or exaggeration (Cooper et al. 1980; Sobell and Sobell 1981). Memory failures are usually considered unintentional errors and can be a cause of underreporting or overreporting. Such errors are less serious with salient or frequently recurring events (Linton 1986; Loftus and Marburger 1983; Tversky and Kahneman 1974). Concealment and overreporting are often related to the social desirability of the recalled behavior (Edwards 1957; Harrell 1985).

Errors attributed to memory failures, concealment, or exaggeration may also be time related. The degree of these errors, when considered together, may depend on the nature of recalled events and on temporal proximity to the time of reporting (Garrison et al. 1987; Hser et al. 1992*b*). Recall failures become more likely to occur as the event becomes distant in time. A concealment (or, occasionally, an exaggeration) of less desirable behaviors is more likely when the event is closer in time to the interview (Hser et al. 1992*b*; O'Malley et al. 1983).

Most studies on self-report of drug use have focused primarily on reliability (e.g., test-retest or internal consistency), especially when external criteria were absent. In general, most studies showed a relatively high level of reporting reliability regarding drug use (e.g., coefficients ranged between 0.80 and 0.95) (Hser et al. 1992*b*). Several methods of objective corroboration (e.g., comparison of data with official records, peer reports) have been applied to assess the validity of self-report. Urinalysis has been the most common method for validating self-reported recent use of drugs. Among the studies that examined validity using urinalysis, 25 percent to 72 percent of subjects whose urine tested positive for drugs denied current or recent use (Maisto et al. 1990; McNagny and Parker 1992).

Few studies have investigated sample characteristics that are correlated with the degree of validity, and results are generally inconsistent from one study to another. For example, some studies (McElrath 1994) found self-reports of drug use to be more valid with samples drawn from community settings than with samples of arrestees; others found that subjects recruited from treatment samples would overreport or underreport drug use depending on the perceived consequences as to whether and how reporting of use might affect their treatment status (e.g., Sherman and Bigelow 1992).

In terms of types of drugs, several studies based on treatment samples have found that the most accurate self-reports were for use of heroin and

other opiates (Magura et al. 1987). However, a review of literature conducted by Maisto and colleagues (1990) concluded that no drug or drug class emerged as associated overall with self-reports of higher reliability and accuracy.

In summary, the literature shows some evidence that drug abusers' self-reports are generally reliable and accurate, but the studies are more strikingly marked by findings of wide variations in accuracy and in the samples and procedures used to obtain them. The empirical evaluation of drug abusers' self-reports of drug use is still in its beginning stages (Maisto et al. 1990) and needs several methodological improvements.

Analytical Approaches

There are several approaches to studying reliability and validity of self-reported data, and each has attendant criteria for empirical evaluation. This section provides a brief overview of common analytical approaches and others that have been used in studies reported later in this chapter. With data available at two points in time, two techniques used to assess reliability are measurements observing differences in means between data obtained at two points in time and test-retest correlations. Differences in means indicate a shift in the distribution of responses that is systematic across respondents. Test-retest correlations are determined from the relative position of a response by a given individual within the two distributions of responses. In this sense, such correlations measure reporting consistency between the two response distributions.

An alternate, complementary approach to the study of the reliability of self-report data extends the concept of test-retest correlation of individual measures to the level of the consistency of the relationships among multiple measures. For example, the stability of the relationship between level of narcotics use and level of drug dealing for a defined period reported at one interview can be compared with that obtained from a later interview. This concept, here termed "pattern reliability," examines the degree of association between two correlation matrices of a set of variables measured at different time points.

A simple test of pattern reliability among the variables examines the consistency in the correlational patterns observed at the two interviews. For example, the correlation coefficient of the two intervariable correlational patterns obtained at two time points provides an assessment of the stability of relationships among a set of common measures across

time. An analytically more sophisticated application would involve confirmatory factor analysis (CFA).

The application of CFA provides a more rigorous testing of the consistency of relationships among measures obtained by self-report. Besides test-retest correlations, when multiple measures of a similar construct are available, internal consistency among these measures can also be used as a reliability measure. The test-retest reliability for individual measures can be further extended to that for relationships among multiple measures obtained at different occasions. In this assessment of pattern reliability, theoretically based relationships, which may reflect internal consistency among multiple measures of a latent construct, are of particular interest.

Pattern reliability can be considered as the consistency of theoretically hypothesized relationships among variables measured at separate occasions. This conceptualization can then be empirically evaluated using CFA (Chou et al., in press). The CFA approach allows simultaneous consideration of internal consistency and test-retest reliability. Investigation of pattern reliability formulated in CFA models involves testing the consistency of measurement and structural models across occasions. Empirical evaluation of pattern reliability can be performed through testing of hypotheses on equality constraints in the specified models. Consequently, investigations of pattern reliability about data provide information for construct validity and construct consistency.

Regardless of the analytic approach, establishing reliability is only a prerequisite to the process of validating data derived from self-report. Reliability is considered a necessary condition of validity but is not sufficient by itself to establish validity. Establishing validity often requires objective information with which self-report data can be corroborated. In the study of drug use, objective data useful for corroboration may include urinalyses results, observational reports, and official records (e.g., earnings, treatment enrollment case files, and criminal justice system histories). The analytical approach of most early studies on validity pertains to percent agreement between self-report and criterion (urinalysis is the most commonly used corroboration for determining accuracy). The computations of kappa and intraclass correlation (ICC) take chance agreement into account. However, these statistics may have biased findings when base rates of an event are extremely low (Spitznagel and Helzer 1985).

The studies described in this chapter illustrate analytical approaches that address some of the empirical issues involved in assessing drug use based on self-report data. Because the primary goal of these original data collections was to obtain accurate self-report of behavior, the studies incorporated procedures that have been suggested to improve the quality of data. For example, confidentiality, anonymity, and privacy during data collection by trained interviewers, with subjects informed in advance of the interview that researchers had access to corroborative information, have all been adopted as effective strategies to improve the quality of self-report. In this last respect, official criminal records were used as memory aids to help respondents recall other life events; therefore, these criminal records cannot be used as independent criteria to validate the self-report data.

ASSESSING THE QUALITY OF SELF-REPORT DATA FROM RETROSPECTIVE LONGITUDINAL SURVEYS

Examinations of the initiation, progression, and course of addiction history, often termed "natural history studies," typically rely on self-report surveys as the primary source of data. In such studies as in other surveys, it is difficult if not impossible to obtain adequate and objective information or to identify criteria that establish validity. For example, surveys often query respondents regarding the frequency of a behavior or the quantity of a substance they have consumed, but the accuracy of responses to such quantitative questions can be affected by many psychological processes or interviewing factors (Bradburn et al. 1987). Validation of responses is particularly difficult if the relevant topic is personal and sensitive or if the recalled events happened in the distant past, as in the case of recall of illicit drug use or criminal involvement during a person's life. Careful scrutiny of the reliability and validity of data obtained from drug-using populations is needed to support the utility of the collected data as well as to understand contributing factors that may affect the quality of data. A series of studies (Anglin et al. 1993; Chou et al., in press; Hser et al. 1992*b*) was conducted to examine the reliability of behaviors reported by a sample of narcotics addicts for the same period of time, but recalled at two widely separated interviews.

Methods

The data used were collected at two face-to-face interviews conducted 10 years apart with the sample group of narcotics addicts (N = 323). The

first interview was conducted during 1974-75, over 10 years after the sample had been admitted to treatment. The interview collected information retrospectively on the individual addiction career starting from 1 year before first narcotics use until the time of interview (a period of approximately 15 years on average). The second interview was completed in 1985-86 and obtained the same self-report data for the period from January 1, 1970, to the time of the second interview. There is an overlapping period of 4 to 5 years in both interviews, from January 1, 1970, to the first interview date in 1974 or 1975. Similar interview instruments and procedures were used on both occasions, and recalled information was elicited on the same set of multiple measures. Urine specimens were also collected at both interviews and used to validate self-report of recent (past 7 days) drug use. (See Hser et al. 1992*b* for a detailed description of subject characteristics, interview procedures, and instruments.)

Analyses and Results

Four sets of analyses were conducted. The first examined the congruence between urinalysis results and self-reported current drug use at each interview point. The second investigated item reliability (reliability of individual variables) measured by test-retest correlations (consistency) and mean level differences (discrepancy). The third set examined the pattern reliability, or consistency of relationship patterns, among all 46 of the selected self-report variables. The fourth examined pattern reliability between narcotics use and property crime using CFA.

Validity of Recent Drug Use. Urinalyses conducted at both interviews provided a limited validity check on recent self-reported narcotics use. At the first interview, among the 97 subjects who tested positive for opiates, 38 (or 39.2 percent) failed to report recent use. At the second interview, 14 (13 percent) of the 105 subjects who tested positive failed to report recent use. The rates of congruence between self-reported current opiate use and urinalyses results among those who provided a urine specimen was 73.6 percent at the first interview and 85.8 percent at the second.

At least two factors may have contributed to the marked difference. A higher proportion (54.9 percent) of the respondents were under some type of legal supervision at the first interview than at the second (28.9 percent), and the perception of possible adverse consequences resulting from divulging recent use may have contributed to under-

reporting. Also, subjects were more confident in the study's intent and staff by the time of the second interview and were more likely to be truthful.

Item Reliability. Overall, behaviors such as narcotics use (including narcotics abstinence and daily use) and employment were recalled with test-retest correlations of at least 0.6. However, less-than-daily narcotics use was recalled less consistently (0.27). Differences between the means of the reported levels at the two interviews were significant for all measures of narcotics use and in general, use levels were reported at a lower level in the first interview (Hser et al. 1992*b*).

Stability of Relationships Among Measures. A correlation matrix containing the intervariable correlation coefficients among 46 variables[1] was constructed for each interview. A single correlation coefficient between the two matrices can be calculated using all the corresponding elements in the lower triangle of the two intervariable correlation matrices (N = 46 * 45/2 or 1,035). Correlation coefficients were obtained from the total overlap period and for each of its constituent 4 years.

For the total of 46 variables for the total overlap period, the absolute difference of the correlation coefficients ranged from 0.06 to 0.07, and the correlation coefficients ranged between 0.84 and 0.90. The absolute difference between the two within-interview correlation coefficients indicated that they remained similar across the 4 years. The correlation coefficients between the two interviews, on the other hand, increased with the reliability of the constituent variables. These results imply, as would be expected, that the correlational pattern among variables becomes more stable when the constituent variables are more reliable. In addition, as opposed to test-retest of individual items, pattern reliability did not decrease with proximity to first interview. Apparently, despite underreporting tendencies, the subjects maintained internal consistency each time they reported on their behaviors (Anglin et al. 1993).

Pattern Reliability: A Confirmatory Assessment of Construct Validity and Consistency. An example of a pattern reliability assessment has been conducted to study the relationships of narcotics use and property crime behaviors (Chou et al., in press). Figure 1 presents a model supporting the pattern reliability and related coefficients. The findings can be summarized as follows. First, the measures used for constructs of narcotics use or property crime yielded substantial factor

loadings on their respective constructs in all models evaluated. This internal consistency among measures is evidence of construct validity through the application of CFA techniques. Second, evaluation of a series of models and comparisons among them showed that constructs reflected by repeatedly measured variables were time dependent. Third, although relevant measures taken at separate occasions cannot support one single construct, construct consistency was demonstrated by the invariance of measurement and structural models as supported by the model presented in figure 1.

Taken together, these findings suggest that the correlational relationships within each of the two sets of repeated measures can be adequately represented by similar models using CFA. It is important for researchers who are interested in substantive issues of behavior to be assured that structural patterns among self-report data such as narcotics use and property crime measures (e.g., factor loadings) and their relationships (e.g., factor correlations) are consistent when measurement errors are separated from the true measures, regardless of when the measures were taken. Results of these analyses contribute strongly to confidence in the validity of the relationships among these self-reported data.

Discussion

Considering the 10-year separation of the two interviews, the test-retest reliability of many drug use variables was reasonably good. The absolute difference level showed systematic discrepancies, however, increasing with proximity to the interview. The distortions seem to have less impact on the reliability of the relational patterns among sets of variables. The pattern reliability of self-reported data from these narcotics addicts is actually quite impressive. The correlation coefficients of the intervariable relationships ranged as high as 0.86 and 0.90. These analyses suggest that although the absolute levels or rates obtained through retrospective self-report may not be as accurate as one would desire, their relative levels (e.g., the relationships among variables) are quite valid, if appropriate interview procedures are conducted. Furthermore, the results of the CFAs based on selected measures of theoretical interests and reasonable reliability (e.g., narcotics use and crime) further demonstrate the utility of the data and the appropriateness of the analytical approaches. Using these model-testing procedures, examination of pattern reliability offers an alternative means of assessing validity of self-report data.

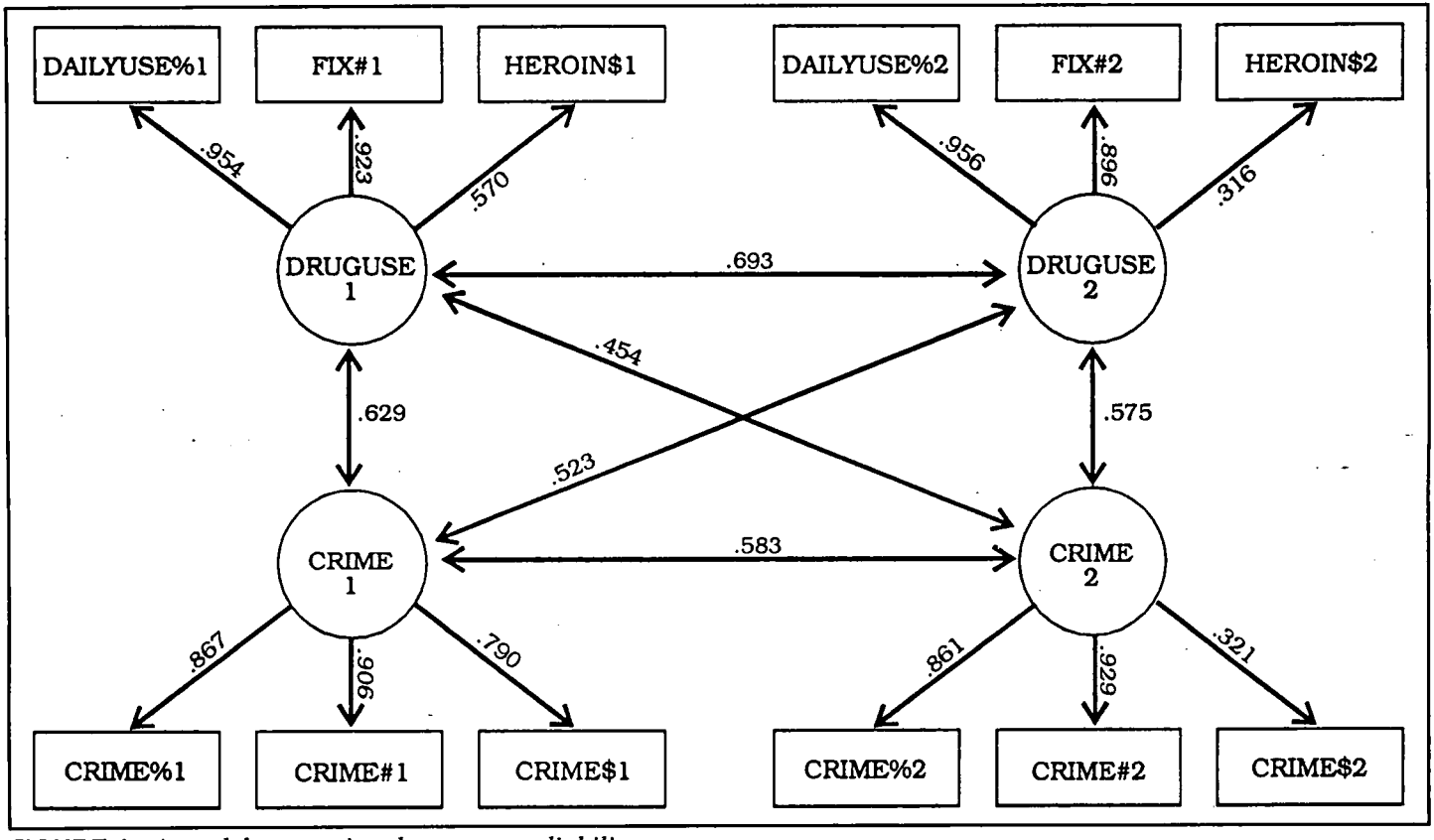


FIGURE 1. *A model supporting the pattern reliability.*

TARGETING HIGH-RISK GROUPS TO IMPROVE PREVALENCE ESTIMATION

Several issues arise when self-reported drug use data collected by general surveys are used as a sole basis for prevalence estimation. First, many general population surveys on drug use do not adequately cover certain populations (e.g., homeless, institutionalized). These populations are likely to have a high level of drug use. Second, data from self-report on such a sensitive topic are often inaccurate due to reporting bias and error. The cost to improve sampling design and reporting accuracy in these large-scale studies can be prohibitively high. Alternatively, researchers have developed statistical models that use complementary data focusing on high-risk populations that are not adequately surveyed to improve the accuracy of estimates for overall drug-using populations (Hser and Anglin 1993; Hser et al. 1992*a*). In addition, smaller scale studies that can provide adjustment factors and suggest ways to improve accuracy of self-report may prove to be more cost efficient in providing improved prevalence estimation results.

Based on practical application and through review of previous efforts, Hser and Anglin (1993) identified several prominent quantitative procedures for making prevalence estimations, including synthetic estimation and capture-recapture models. In addition, there is a need for improved data, particularly regarding the undersurveyed populations at high risk of being drug users as well as the linkages among these populations. A study is currently being conducted by the author to survey drug use among several important high-risk populations. Some preliminary results are reported in this section to illustrate the investigation of validity of self-reported drug use in these population samples. Findings suggest ways to improve accuracy of estimates produced by surveys based on self-report.

Methods

Face-to-face interviews were conducted with a total of 3,493 subjects screened from patients at STD clinics (N = 1,134), patients visiting ERs (N = 680), and arrestees (N = 1,679). (Table 1 describes background characteristics of the sample.) Interview procedures and questionnaires were similar across sources, which are all located in Los Angeles County. Recruitment procedures did not follow a random sampling procedure and were slightly modified to accommodate constraints of a particular setting or subjects' clinical needs. Subjects were surveyed

regarding recent use of illicit drugs and drug use history, along with many other health questions. To assess the overlap of populations from the three different sources, all subjects were also asked whether they had been arrested or had visited STD clinics or emergency rooms in the past year. Urinalysis results were used to corroborate the validity of self-report of recent drug use. (See Hser et al. (submitted) for a detailed description of subject recruitment and interview procedures.)

TABLE 1. *Background characteristics.*

	STD (N = 1,134)	ER (N = 680)	Jail (N = 1,679)
Age (%)			
18-14	40.4	20.0	32.4
25-39	48.0	41.8	55.9
40+	11.6	38.2	11.7
Mean	28.9	36.8	29.6
Standard deviation	(9.2)	(13.1)	(8.3)
Female (%)	39.9	35.0	35.1
Ethnicity (%)			
White	6.3	13.4	20.4
Hispanic	33.1	51.2	41.3
African American	58.6	30.6	36.1
Other	2.0	4.9	2.3

Analyses and Results

Three sets of analyses were conducted. The first compared self-report data on use of several drugs with results of urinalyses. Using urine-testing results as the accuracy criteria, two rates of inaccurate self-report (denial among users and denial among self-reported nonusers) were calculated. The second analysis examined correlates of these measures of inaccuracy using logistic regression analyses, and the third investigated the degree of overlap among the three study samples.

Validity of Recent Drug Use. Urinalyses were conducted and results compared to self-report among those who provided urine specimens. Table 2 shows results of recent drug use in several categories by self-report

and by urinalysis. As expected, illicit drug use is quite high in the study samples, most notably among the arrestees (52.5 percent positive for cocaine, 68.9 percent positive for any drug). In general, the rates of negative urinalyses among those reporting recent use are negligibly low.

For all types of drugs, self-reports are based on use in the 3 days before the interview. Because marijuana can stay in the body much longer, self-reported use in past 30 days is also included (Ellis et al. 1985; Mieczkowski 1990). Except for marijuana use in the past 30 days, use levels for all drugs were higher by urinalysis than by self-report.

TABLE 2a. *Percents self-report versus urinalyses.*

STD clients (N = 1,061)				
	Self-reported use in past 3 days	Positive	Denial among users*	Denial among negative self-report**
Marijuana (3 days)	17.8	18.9	33.0	7.6
Marijuana (30 days)	30.6	18.9	13.0	3.5
Cocaine	3.4	10.0	68.9	7.1
Opiates	0.5	0.9	60.0	0.6
Amphetamines	0.4	0.3	66.7	0.2
PCP	0.5	1.4	73.3	1.0
Benzodiazepine	0.7	1.3	71.4	0.9
Any drug less marijuana	5.2	13.7	71.7	10.3
Any drug	20.7	28.1	21.8	9.5

KEY: * = percent reported no use among urine positive; ** = percent of positive urine among those reported no use of the respective drug.

Two discrepancy measures can be calculated by contrasting the same group of subjects who provided discrepant reporting (those whose

TABLE 2b. *Percents self-report versus urinalyses.*

	ER clients (N = 482)			
	Self-reported use in past 3 days	Positive	Denial among users*	Denial among negative self-report**
Marijuana (3 days)	8.7	8.3	35.0	3.2
Marijuana (30 days)	15.3	8.3	22.5	2.2
Cocaine	6.8	12.2	50.8	6.7
Opiates	1.9	8.9	83.7	7.6
Amphetamines	0.6	1.2	100.0	1.3
PCP	0.6	2.3	81.8	1.9
Benzodiazepine	0.8	3.7	88.9	3.3
Any drug less marijuana	9.3	27.6	68.4	20.8
Any drug	15.8	32.4	42.3	18.4

KEY: * = percent reported no use among urine positive; ** = percent of positive urine among those reported no use of the respective drug.

self-reported drug use disagreed with urinalysis results) with the other two groups whose urine results were either positive or negative but were consistent with their self-reports. The most common discrepancy measure used in validity studies has been the rate of subjects who failed to admit their use but were tested positive. This denial among users is the percentage of those testing positive for the drug who claimed no recent usage. Relatively high rates of denial were observed in the study samples, although the appearance can be exaggerated in drugs (such as amphetamines) whose base use rate is low.

Another discrepancy measure that can be used as a correction of underreporting is denial among self-reported nonusers. This rate is calculated as the percentage of positive urine results among those who reported no recent use of the respective drug. For example, among arrestees, of those who did not admit recent use of cocaine, as many as

TABLE 2c. *Percents self-report versus urinalyses.*

Jail clients (N = 1,666)				
	Self-reported use in past 3 days	Positive	Denial among users*	Denial among negative self-report**
Marijuana (3 days)	19.0	19.4	43.0	10.3
Marijuana (30 days)	34.1	19.4	22.6	6.6
Cocaine	33.7	52.5	38.1	30.1
Opiates	10.2	11.1	25.9	3.2
Amphetamines	5.2	7.7	46.9	3.8
PCP	1.6	2.9	66.7	2.0
Benzodiazepine	1.6	6.2	86.4	5.4
Any drug less marijuana	41.6	62.1	35.0	37.2
Any drug	50.0	68.9	56.4	34.6

KEY: * = percent reported no use among urine positive; ** = percent of positive urine among those reported no use of the respective drug.

30.1 percent were using according to urinalysis. This denial rate, reflecting actual prevalence rates among those who deny use by self-report, can be extrapolated to adjust upward the survey results that are based entirely on self-report for similar populations.

It should be noted that the numerators for the two denial rates were the same groups of people, but the denominators were two different contrast groups. Both measures are important for improving the prevalence estimation of drug use based on self-report. The rate of denial among users indicates the likelihood of denying use among those users identified by urine testing. The rate of denial among self-reported nonusers suggests the degree of underreporting among respondents who did not admit drug use, which can be extrapolated to improve prevalence estimates based on self-reported data from similar populations where urine analysis is not available. In addition, correlates of these two

measures can be different and each may provide some useful information.

Correlates of Reporting Accuracy. Logistic regression was performed to examine factors related to inaccurate self-report measured as denial among users and denial among self-reported nonusers (tables 3 and 4 respectively). For each of these two outcome measures, separate regression analyses were conducted for three drugs (cocaine/crack, opiates, and marijuana) with high prevalence in the study samples. For the analysis on denial among users, each analysis included subjects who tested positive for the particular drug, with the dichotomous dependent variable coded 0 for accurate self-report of use and 1 for inaccurate self-report. Similarly, for the analysis of denial among self-reported nonusers, each analysis included subjects who reported no use of the particular drug, with the dichotomous dependent variable coded 0 for negative urine results (accurate self-report) and 1 for positive urine (inaccurate self-report).

The regression results on denial among users indicate that the type of interview site or sample source was an influence on validity of self-report. Compared to STD samples, subjects in jails were significantly less likely to lie about cocaine use and significantly more likely to lie about marijuana. Males were significantly less likely than females to lie about marijuana use, while persons 40 years of age or older were significantly more likely to lie regarding use of this drug than their younger counterparts. Subjects acknowledging past drug dependence were far less likely to be dishonest regarding current drug use. The stigma attached to use of cocaine/crack and opiates (the latter is not statistically significant) may lead to dishonesty in the more mainstream samples such as the STD clients. The social acceptability of marijuana use among young people may account for their more accurate reporting relative to older people.

The regression results on denial among self-reported nonusers indicate that, for example, the following subject characteristics were significantly correlated with positive urine results among subjects who reported no recent cocaine use: female, ethnic minority (African American and Hispanic), jail sample, older ages, multiple arrests in the past year, currently not in treatment, and past dependence. These factors should be considered to refine the adjustment factors of underreporting.

TABLE 3. *Logistic regression for denial (among users by urine testing).*

Predictors	Cocaine/crack N = 1,040		Opiates N = 238		Marijuana N = 563	
	β	Odds ratio	β	Odds ratio	β	Odds ratio
Gender						
Female						
Male	-0.1535	0.8577	0.2744	1.3158	-0.6989**	0.4971
Race						
White						
African American	0.1374	1.1473	0.8258	2.2836	0.4970	1.6437
Hispanic	0.1205	1.1281	-0.6810	0.5061	0.4586	1.5819
Other	0.0417	1.0426	1.5750	4.8305	0.2325	1.2618
Source						
STD (N = 1,061)						
ER (N = 482)	-0.7736*	0.4613	0.4813	1.6182	0.8164	2.2624
Jail (N = 1,666)	-1.5033**	0.2224	-1.7564	0.1727	0.8606*	2.3646
Age						
18-24						
25-39	-0.2809	0.7551	0.1163	1.1233	0.2628	1.3005
40+	-0.1043	0.9010	-0.2120	0.8090	1.6349**	5.1288
Sex partners/past year						
0						
1	0.2016	1.2234	-0.2676	0.7653	0.8538	2.3487
2	-0.2360	0.7898	0.3369	1.4006	0.4269	1.5326
3-10	-0.3350	0.7154	0.9982	2.7133	0.5039	1.6552
11+	-1.0667**	0.3442	0.4476	1.5645	0.5736	1.7747
Arrests/past year						
0						
1	0.4853	1.6247	0.3018	1.3523	0.1498	1.1616
2	0.5090	1.6636	-0.2078	0.8124	0.2911	1.3379
3+	0.1423	1.1529	-1.3089	0.2701	0.1587	1.1720
Currently in treatment						
No						
Yes	-0.8965	0.4808	-2.0502	0.1287	-0.0462	0.9549
Ever dependent						
No						
Yes	-1.4399**	0.2370	-3.2343**	0.0394	-1.4521**	0.2341

KEY: * = $p < 0.05$; ** = $p < 0.01$.

TABLE 4. *Logistic regression for denial (among self-reported nonusers).*

Predictors	Cocaine/crack N = 2,579		Opiates N = 3,025		Marijuana N = 2,224	
	β	Odds ratio	β	Odds ratio	β	Odds ratio
Gender						
Female						
Male	-0.3493*	0.7052	-0.2213	0.8015	0.1464	1.1577
Race						
White						
African American	1.6230**	5.0683	-0.6007	0.5484	0.4620	1.5872
Hispanic	0.9511**	2.5886	0.0543	1.0558	-0.5279	0.5899
Other	0.3747	1.4546	-0.6706	0.5114	-0.3639	0.6950
Source						
STD (N = 1,061)						
ER (N = 482)	0.1922	1.2120		11.6695	0.0127	1.0128
Jail (N = 1,666)	1.5598**	4.7581	1.0438*	2.8399	0.2036	1.2258
Age						
18-24						
25-39	1.0317**	2.8057	0.5034	1.6543		0.6177
40+	0.9313**	2.5377	0.8748	2.3985	-0.6233	0.5362
Sex partners/past year						
0						
1	0.0405	1.0413	-0.7054*	0.4939	1.1517	3.1637
2	0.0082	1.0082	-0.1096	0.8962	0.9338	2.5442
3-10	0.3603	1.4338	-0.0125	0.9876	1.0094	2.7439
11+	0.2563	1.2921	-0.3552	0.7010	0.9425	2.5663
Arrests/past year						
0						
1	0.4492	1.5670	0.7712*	2.1623	0.7717	2.1635
2		2.7426	0.8347	2.3040	1.1175*	3.0572
3+		2.8377	0.7038	2.0213	0.9416	2.5642
Currently in treatment						
No						
Yes	-1.6635*	0.1895	-1.4345	0.2382	-0.8047	0.4472
Ever dependent						
No						
Yes	0.5632**	1.7630	0.4232	1.5268	-0.7257*	0.4840

KEY: * = $p < 0.05$; ** = $p < 0.01$.

Overlap of Drug Users from High-Risk Groups. The degree of overlap of drug users identified from the three sources can be assessed by their reported appearance in the other two sources. With cocaine users as an example, the percentages of reported involvement with hospital ERs, STD clinics, and arrest in past year from each sample are presented in table 5. The results indicate that many cocaine users identified from jails represent significant proportions of users who also utilized hospital ERs and STD clinics. Likewise, a high number of cocaine-using patients who utilized an ER or STD clinic also reported at least one arrest in the past year (39.0 percent and 29.1 percent respectively). About 21.7 percent of STD patients reported at least one visit to an ER in the past year, while only 1.9 percent of ER patients reported an STD clinic visit.

TABLE 5. *Degree of overlap of drug users from the three sources (percents).*

	STD	ER	Jail
STD[1]	-----	1.9	50.8
ER[2]	21.7	-----	30.5
Jail[3]	29.1	39.0	-----
Other two	7.5	1.9	9.6

KEY: 1 = At least one visit to STD clinic in past year; 2 = at least one visit to hospital ER in past year; 3 = at least one arrest in past year.

Discussion

The subjects examined in these analyses are from source populations at high risk of being drug users. The collection of urine specimens provides a further opportunity for estimating the degree of underreporting.

The logistic regression analysis shed light on specific factors to be considered in improving the prevalence estimation. The analysis of denial among self-reported nonusers suggested that, using cocaine as an example, upward adjustments of prevalence estimates should be different

for gender (higher for female), race (higher for African Americans and Hispanics, as opposed to whites), recruitment source (higher for jail than other sources), arrests (higher for people with more arrests), treatment (lower for people currently in treatment), and past dependence. Groups identified to be associated with higher prevalence rates were generally consistent with those found in other surveys (with the exception of females). The rate of positive urine for cocaine was 34.8 percent among females and 31.0 percent among males. Similarly, urine-positive rate among self-reported nonusers for females was 17.8 percent and for males, 16.4 percent. On the other hand, contrary to findings from many other studies, the current analysis on denial among users indicates that drug users recruited from jails were not necessarily more likely than users recruited from STD clinics to lie about their use. For example, among those who had a positive urine result for cocaine, subjects from jails and ERs were less likely to lie about their cocaine use than those recruited from STD clinics; no differences were detected among recruitment sources for reports of opiate use. Curiously, compared to the STD sample, the jail sample showed a higher likelihood of denying marijuana use. This interaction between sample sources and drug type seems to suggest that while all samples underreport somewhat for all drugs, relative to the jail sample, underreporting among the STD sample is more serious for cocaine and less serious for marijuana.

The obvious limitation of the study is that the subjects were not obtained as a probability sample. However, the analyses are initial steps for the identification of empirical issues that need to be considered and analytic approaches that can be adopted to address issues of validity. Prevalence estimation based on self-report can be improved by making adjustments according to identified influencing factors. In addition, focusing on high-risk populations and taking into consideration the overlap and divergence of such groups can empirically improve the prevalence estimates that rely on single sources.

CONCLUDING COMMENTS

Despite advances in other ways to measure drug use, self-report remains the most efficient way to assess the various dimensions of drug use (e.g., quantity and frequency for a given substance over periods of time that can span from the past few days or weeks to a lifetime history of use). Therefore, rather than asking whether or not they are accurate, a more productive approach is to inquire about the determinants of the

accuracy of self-report data and then devise ways to improve accuracy of estimates based on self-report.

This chapter addressed two types of empirical questions often encountered when considering self-report data on drug use and suggested analytical approaches to address these issues. The first set of analytical approaches can be utilized when external objective criteria, particularly those of a longer term nature, are unavailable. In the absence of dependable objective measures, alternative methods of confirmation may have to be sought. Although purely statistical techniques cannot be taken as a fully satisfactory corroborative solution, the CFA procedures described have served as a valuable means of supporting confidence in the validity and reliability of the self-report data obtained. Given the cognitive limitation and other potential sources of response bias that are unlikely to be totally eliminated in self-report, methodologies such as CFA, which uses multiple measures to control measurement errors, should be considered for application in analyzing substantive issues of human behavior.

The second approach to improve prevalence estimation relies on enhanced data collection from targeted populations known to be at high risk of being drug users, then measuring their overlaps. Overall, compared to the survey results of general populations, the drug use prevalence rates are relatively high in all three samples. This type of small-scale study that targets high-risk populations and collects more objective measures can suggest appropriate rates of upward adjustments to be applied to estimates that are wholly reliant on self-report. Furthermore, findings of correlates of reporting accuracy can suggest control variables necessary for refining the adjustment rates.

Quality of self-report data can be a product of a variety of factors ranging from data collection procedures to subject characteristics. Although improvements have been made in methodologies for collecting self-reported drug use data over the past decade, more systematic methodological investigations in the context of measuring drug use are needed. Finally, even for survey studies that have followed all appropriate procedures to ensure the collection of the best possible quality of data, empirical validation of self-report data is always necessary to enhance the utility of these data and to suggest means of controlling for potential biases.

NOTE

1. These 46 variables include various measures of narcotics use (e.g., abstinence, daily use, number of fixes per month); nonnarcotics use; marijuana use; alcohol use; drug dealing; various property crimes (e.g., forgery, theft, robbery) in terms of numbers of crime days per month, percentage of time; legal supervision status (probation, parole, with or without urine testing); employment; welfare; and other similar variables.

REFERENCES

Anglin, M.D.; Hser, Y.; and Chou, C. Reliability and validity of retrospective behavioral self-report by narcotics addicts. *Eval Rev* 17(1):91-108, 1993.

Bradburn, N.M.; Rips, L.J.; and Shevell, S.K. Answering autobiographical questions: The impact of memory and inference on surveys. *Science* 236:157-161, 1987.

Bradburn, N.M., and Sudman, S. *Polls and Surveys: Understanding What They Tell Us.* San Francisco: Jossey Bass, Inc., 1988.

Chou, C.; Hser, Y.; and Anglin, M.D. Pattern reliability of narcotics addicts' self-reported data: A confirmatory assessment of construct validity and consistency. *Int J Addict*, in press.

Cooper, A.M.; Sobell, M.B.; Maisto, S.A.; and Sobell, L.C. Criterion intervals for pretreatment drinking measures in treatment evaluation. *J Stud Alcohol* 41(11):1186-1195, 1980.

Edwards, A.L. *The Social Desirability Variable in Personality Assessment and Research.* New York: Dryden, 1957.

Ellis, G.M., Jr.; Mann, M.A.; Judson, B.A.; Schramm, N.T.; and Tashchian, A. Excretion patterns of cannabinoid metabolites after last use in a groups of ethnic users. *Clin Pharmacol Ther* 38:572-578, 1985.

Garrison, C.Z.; Schoenbach, V.J.; Schluchter, M.D.; and Kaplan, B.H. Life events in early adolescence. *J Am Acad Child Adol Psychiatry* 26(6):865-872, 1987.

Harrell, A.V. Validation of self-report: The research record. In: Rouse, B.A.; Kozel, N.J.; and Richards L.G., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity.* DHEW Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985.

Harrison, L.D. The validity of self-reported data on drug use. *J Drug Issues* 25(1):91-111, 1995.

Hser, Y., and Anglin, M.D., eds. Prevalence estimation techniques for drug-using population. *J Drug Issues* 23(2), 1993.

Hser, Y.; Anglin, M.D.; and Boyle, K. "Drug Use and Correlates in Three High-Risk Samples." Manuscript submitted for publication.

Hser, Y.; Anglin, M.D.; and Chou, C. Reliability of retrospective self-report by narcotics addicts. *Psychol Assess* 4(2):207-213, 1992*b*.

Hser, Y.; Anglin, M.D.; Wickens, T.; Brecht, M.; and Homer, J. "Techniques for the Estimation of Illicit Drug-Use Prevalence: An Overview of Relevant Issues." Report prepared for the National Institute of Justice, 1992*a*.

Linton, M. Ways of searching and the contents of memory. In: Rubin, D.C., ed. *Autobiographical Memory*. New York: Cambridge University Press, 1986. pp. 50-67.

Loftus, E.F., and Marburger, W. Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Mem Cog* 11(2):114-120, 1983.

Magura, S.; Goldsmith, D.; Casriel, C.; Goldstein, P.J.; and Lipton, D.S. The validity of methadone clients' self-reported drug use. *Int J Addict* 22(8):727-749, 1987.

Maisto, S.A.; McKay, J.R.; and Connors, G.J. Self-report issues in substance abuse: State of the art and future directions. *Behav Assess* 12:117-134, 1990.

McElrath, K. A comparison of two methods for examining inmates' self-reported drug use. *Int J Addict* 29(4):517-524, 1994.

McNagny, S.E., and Parker, R.M. High prevalence of recent cocaine use and the unreliability of patient self-report in an inner-city walk-in clinic. *JAMA* 267(8):1106-1108, 1992.

Mieczkowski, T. The accuracy of self-reported drug use: An evaluation and analysis of new data. In: Weisheit, R.A., ed. *Drugs, Crime and the Criminal Justice System*. Cincinnati: Anderson Publishing Co., 1990. pp. 275-302.

O'Malley, P.M.; Bachman, J.G.; and Johnston, L.D. Reliability and consistency in self-reports of drug use. *Int J Addict* 18(6):805-824, 1983.

Sherman, M.F., and Bigelow, G.E. Validity of patients' self-reported drug use as a function of treatment status. *Drug Alcohol Depend* 30:1-11, 1992.

Sobell, L.C., and Sobell, M.B. Effects of three interview factors on the validity of alcohol abusers' self-reports. *Am J Drug Alcohol Abuse* 8:225-237, 1981.

Spitznagel, E.L., and Helzer, J.E. A proposed solution to the base rate problem in the kappa statistic. *Arch Gen Psychiatry* 42:725-728, 1985.

Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. Rockville, MD: National Institute on Drug Abuse, 1992.

Tversky, A., and Kahneman, D. Judgement under certainty: Heuristics and biases. *Science* 185:1124-1131, 1974.

ACKNOWLEDGMENTS

The author was supported by the Research Scientist Development Award (K02DA00139) from the National Institute on Drug Abuse. This chapter was prepared with additional support from NIDA grants P50DA07699 and DA07382, and from the California Department of Alcohol and Drug Program contract 92-00267. The author is grateful to Dr. M. Douglas Anglin for useful comments on an earlier draft, M. Maglione for assistance in data analysis, and to B. Perrochet and M. I. Mendez for their supportive work.

AUTHOR

Yih-Ing Hser, Ph.D.
Adjunct Associate Professor of Psychiatry
Associate Director
UCLA Drug Abuse Research Center
Suite 763
1100 Glendon Avenue
Los Angeles, CA 90024-3511

167354

Design and Results of the Women's Health Study

Roger Tourangeau, Jared B. Jobe, William F. Pratt, and Kenneth Rasinski

ABSTRACT

The Women's Health Study was a methodological experiment carried out in Chicago. More than 1,000 women took part; a comparison sample of 100 men was also included. The sample was selected from two sources. Most of the women and all of the men were selected from an area probability sample that had been screened to identify women in the eligible age range; the rest of the women were selected from rosters at cooperating abortion clinics and were known to have had an abortion. Questionnaires based on the one used in the National Survey of Family Growth were administered to the sample; the questionnaire included items on abortion, sexual behavior, and illicit drug use. The experiment examined five variables: whether the questionnaire began with a series of medical questions or with questions on pregnancy; whether the interview was conducted by a nurse or field interviewer; whether the interview was done at the respondent's home or outside the home; whether the interviewer or respondent administered the questions; and whether the data were collected on paper or via computer. Of the five experimental factors, the one with the most consistent effect was the method of administering the questions. Self-administration significantly increased the reported number of sexual partners, sexually transmitted diseases, and the level of condom use compared to administration by an interviewer. Computer assistance occasionally interacted with the site of the interview to effect reporting. The other two experimental variables—the version of the questionnaire and the data collection staff—had few discernible effects. None of the variables affected reported drug use over the lifetime.

INTRODUCTION

This study investigates some sources of error in surveys that collect information on sensitive topics, topics that involve illegal or embarrassing activities. More specifically, the study tested procedures to improve the

accuracy of data collected in the National Survey of Family Growth (NSFG). Since its beginning in 1971, NSFG has obtained detailed information on fertility and reproductive health. In each of its four cycles conducted to date, the survey has explored a broad range of sensitive questions that concern topics such as contraceptive practices, pregnancy histories (including fetal and infant deaths), unplanned and unwanted pregnancies, sexually transmitted diseases (STDs), and infertility (Judkins et al. 1991).

From the outset, there have been concerns about the sensitive nature of NSFG questions. For example, the first two cycles of NSFG (carried out in 1973 and 1976) generally excluded women who had never been married because it was believed that many unmarried women would not answer questions about pregnancy and contraceptive practice truthfully. Despite these concerns, the sample was expanded in the 1982 NSFG to represent all women regardless of marital status. Changes to the content of the NSFG questionnaires have also increased the sensitivity of the interview over time. For example, the 1988 NSFG added questions on risk factors for acquired immunodeficiency syndrome (AIDS). Despite the increasingly intimate information being sought in NSFG, response rates have remained high. Around 80 percent of the cases selected for NSFG complete the interview and only about one-third of the non-respondents are outright refusals (Judkins et al. 1991). Moreover, nonresponse to individual questions has generally been less than 1 percent. Of course, the fact that respondents answer the questions offers no assurance that their answers are truthful (Jones and Forrest 1992).

Sensitive Questions in Surveys

From the point of view of survey methodology, this study concerns a very general problem—how to collect data on topics that most people are likely to regard as private. Many surveys include questions about private or potentially embarrassing matters, asking respondents about their annual income, their employment status, and so on. Since the outbreak of the AIDS epidemic, the need for data on such sensitive topics as sexual behaviors and illicit drug use has dramatically increased. But though the need for such data is clear, it is not clear whether the data collected are accurate.

Findings from surveys on sexual behavior illustrate the problems in collecting sensitive data in surveys. Within a closed population (that is, a population with no sexual contacts with outsiders), equal numbers of

opposite-sex sexual partners should be reported by men and women; the total number should be the same because the same sexual pairings are being reported by respondents of both sexes. As Smith (1992) has demonstrated, however, men consistently report more opposite-sex sexual partners than do women, a difference that persists even when differences in the population sizes are taken into account. The most plausible account of the discrepancy is that men overstate their partners and that women overlook theirs. A recent review of the methodological problems in AIDS research described the situation this way: "Most sex research is based on self-reported sexual behavior of unknown validity" (Catania et al. 1990, p. 339). Much the same judgment would apply to research on illicit drug use, abortion, and other sensitive topics.

Improving Reporting on Sensitive Topics. A hypothesis guiding much of the survey literature on reports about sensitive topics is that a major source of error is more or less deliberate misreporting. Questions about sensitive topics create conflicts for the respondents, who generally want to cooperate by giving correct answers, but who also want to avoid embarrassment or, when the behavior in question is illegal, legal repercussions. Much of the methodological research designed to improve answers to sensitive questions has concentrated on techniques that reduce the perceived threat of the questions by increasing the privacy of data collection. More recently, methodological studies have begun to examine the effect of computer-assisted data-collection techniques on reporting of sensitive behaviors.

Increasing the privacy of data collection is widely believed to improve the accuracy of the answers. One of the most practical methods for increasing perceived privacy is to use self-administered questionnaires (SAQs) rather than face-to-face interviews to collect the data. In most surveys, the data are not entirely confidential because at least the interviewer is aware of the respondent's answers; further, when interviewers administer the questions and record the answers, it is possible for other household members to overhear what the respondent is saying. Surveys that employ SAQs (in which respondents record their answers without the mediation of an interviewer) overcome these threats to confidentiality. SAQs generally obtain higher levels of reporting of sensitive behaviors than do face-to-face interviews, with telephone interviews falling somewhere in between the other two modes in levels of reporting (see Bradburn 1983 for a review). The advantages of SAQs have been demonstrated for a number of sensitive topics, including sexual behavior (Boekeloo et al. 1994), illicit drug use (Aquilino and

LoSciuto 1990; Schober et al. 1992; Turner et al. 1992), alcohol consumption (Aquilino and LoSciuto 1990; Hochstim 1967), and abortion reporting (London and Williams 1990; Mosher and Duffer 1994; Mott 1985).

Several studies, for example, have shown that self-administration increases reporting of illicit drug use, alcohol consumption, or both. Aquilino and LoSciuto (1990) compared drug use data collected by interviewers over the telephone with data collected by self-administration as part of a personal interview. They found substantially higher reporting of both drinking and drug use with the self-administered questions. Two subsequent comparisons between face-to-face interviewing and SAQs also found greater reporting of cocaine and marijuana use in the self-administered condition (Schober et al. 1992; Turner et al. 1992). Finally, an early comparison of face-to-face data collection with data collection by mail and telephone revealed lower levels of reported alcohol consumption in face-to-face interviews (Hochstim 1967).

Self-administration also appears to reduce survey respondents' reluctance to admit that they have had an abortion. Mott (1985) reports evidence that self-administration greatly increased the number of abortions reported, and similar results have been obtained in studies of abortion reporting by London and Williams (1990; Mosher and Duffer 1994). Boekeloo and colleagues demonstrate that self-administration also increases reporting on other sexual topics; respondents were more likely to admit to unprotected sexual intercourse and a history of STDs in an SAQ than in a face-to-face interview (Boekeloo et al. 1994).

For all of their advantages, however, SAQs also have their limitations. Self-administration with a paper-and-pencil questionnaire requires that the respondents be able to read. In addition, the routing instructions have to be kept simple; the elaborate skip patterns used in many interviewer-administered surveys may be impossible to duplicate in an SAQ.

Another method that increases the apparent confidentiality of survey responses is the randomized response technique (Warner 1965). In this technique, a random device rather than the interviewer determines what question the respondent answers (e.g., the respondent spins a dial to determine which of two questions to answer); in this way, the interviewer cannot know for sure what the respondent's answer means. The randomized response technique method has been shown to increase the proportion of women reporting that they have had abortions (Abernathy

et al. 1970; I-Cheng et al. 1972; Shimizu and Bonham 1978). However, the procedure is difficult to use in a large survey and greatly complicates the analysis of the results.

OTHER VARIABLES AFFECTING LEVELS OF REPORTING

Answers to threatening questions also appear to be affected by the format and wording of the questions (Bradburn 1983), although the results for these variables are not so well documented as those for self-administration. Whether the items use an open or closed response format appears to have an effect on reporting of sensitive behaviors. For example, Bradburn and colleagues (1979) found that, compared to closed questions, open questions produced increases in reporting that ranged from 14 percent for frequency of sexual intercourse during the past month to 108 percent for frequency of masturbation; over a number of sensitive items, the average increase in the level of reporting was 52 percent for the open as compared to the closed versions of the questions. However, the format of the questions did not affect whether respondents report engaging in the behaviors at all.

Longer questions may also yield fuller reporting. Reports about the frequency of behavior and amount of consumption are subject to memory errors even when there are no motivational obstacles to truthful reporting (see Jobe et al. 1993 for a review). Particularly if the behavior is frequent and episodes are not highly differentiated (as with frequent use of illicit drugs), respondents may not remember how many times they have engaged in the behavior during a particular reference period. By giving respondents more cues and more time to search their memories, longer questions can produce more complete reporting (Marquis and Cannell 1971). In their study of sensitive behaviors, Bradburn and coworkers (1979) found that longer questions produced consistently higher levels of reported behavior.

A promising new technique that may increase perceived privacy and produce more accurate data on sensitive behavior is the computer-assisted self-administered interview (CASI). A study by Waterton and Duffy (1984) found that a computer-administered questionnaire produced greater reports of alcohol use than a conventional face-to-face interview. This study confounds the effects of computer assistance and those of self-administration, as do several other studies on CASI (Locke et al. 1992; Lucas et al. 1977; Robinson and West 1992). There is, however,

some evidence that computer assistance by itself can enhance the reporting of sensitive behaviors. In a test of computer-assisted personal interviewing (CAPI), Baker and Bradburn (1991) found that CAPI respondents were more likely than respondents to a paper-and-pencil interview to report having used birth control methods in the past month.

Variables for This Study

The Women's Health Study was sponsored by the National Center for Health Statistics, with additional support from the National Science Foundation. Its purpose was to test alternative methods for collecting sensitive information; the results were used in planning for Cycle V of the NSFG. Along with the experiment described here, the Women's Health Study included cognitive interviews and focus group discussions. The experiment was based on the assumption that survey reports about abortion and other sensitive topics might be improved through several means. The study examined three of these strategies for investigation. The first strategy—increasing the privacy of the data-collection process—was already well established in the survey methods literature. This strategy was chosen over some of the other possibilities in the literature because past results suggested that privacy was the single most powerful variable affecting reporting on sensitive topics and that manipulating this variable would have the largest impact on the survey estimates.

In examining the privacy variable, the experiment assessed the impact of self-administration rather than the major alternative, the randomized response technique, because of the practical and statistical difficulties associated with that procedure. However, in addition to the use of self-administered questions, the study tested the impact of moving the interview outside the respondent's home (and away from other family members). Relatively few studies have recorded whether face-to-face interviews involving sensitive topics were conducted in private or with other household members present or able to overhear the respondent's answers. As a result, the effects of the privacy of the setting in which the interview is carried out are unclear. It was hypothesized that moving the interview to a neutral site away from other family members might increase the respondent's sense of privacy and thus improve reporting.

The second approach examined by the study was that of placing the interview as a whole, and especially the questions on abortion, in a medical context. A medical context for the interview might reinforce the

need for accurate data for health planning purposes; in addition, it seemed possible that respondents might be more accustomed to providing candid answers in the setting of a medical interview than in the survey setting. Attempts to foster a medical context were done in two ways: First, interviewers who were themselves medical practitioners were used—that is, nurses and nursing assistants interviewed some of the respondents. Earlier work investigating this approach for collecting sensitive data is scarce and this project tests its effectiveness. Second, the interview began with a long series of questions about medical conditions and procedures.

A final strategy investigated in this project was the use of computer-assisted data collection. Early evaluations suggest that computer assistance may enhance either the apparent privacy of data collection or the perceived objectivity and importance of the study; either way, it was hypothesized that computerization of the data-collection process might increase respondents' willingness to report truthfully.

Methods

This large-scale field experiment was conducted in the city of Chicago. More than 1,000 women were interviewed, along with a small comparison sample of 100 men. The sample was selected from two sources. All of the men and most of the women respondents were selected from an area probability sample that had been screened to identify persons in the eligible age range (ages 15 through 35); the rest of the women were selected from rosters at cooperating health clinics and were known to have had abortions.

Questionnaires based on the one used in the NSFG were administered to the sample; the questionnaires included items on abortion, sexual behavior, and illicit drug use and took about an hour to complete (Rieger et al. 1991). The experiment examined five variables: whether the questionnaire began with a series of medical questions or with questions on pregnancy, whether the interview was conducted by a nurse or a regular field interviewer, whether the interview was done at the respondent's home or at a site outside the home, whether the interviewer or respondent administered the questions, and whether the data were collected via computer or on paper. The analysis examined a number of outcome variables, including the response rates under the various experimental conditions, the level and accuracy of abortion reporting, and the level of reporting on other sensitive topics (such as the number of sexual

partners). The focus here is on reports about sexual behavior (see Jobe et al., in press, for findings on the other topics).

Sample

Area Probability Sample. The area probability portion of the sample was a stratified, multistage sample of dwellings in the city of Chicago, selected using standard methods. In the first stage of selection, a sample of 85 area segments was drawn; each segment consisted of a single block or group of adjoining blocks, defined using data from the 1990 Census. After all of the blocks in the city of Chicago had been sorted by geographic area, a systematic sample of 85 of them was selected. Selection probabilities for each segment were proportional to the 1990 census count of the number of housing units it contained. This method of sample selection assured that each area in the city of Chicago would receive proportionate representation in the sample. Each segment included at least 40 housing units (according to the census data); blocks that did not meet this size standard were linked to adjacent blocks until the combined unit included 40 or more housing units.

In the 85 sample segments, a subsample of dwellings was designated to receive a short screening interview to identify persons eligible for the main experiment. In total, 6,325 occupied dwellings were selected for screening. Screening interviews were completed at 4,659 of these, for a response rate of 73.7 percent. Much of the nonresponse occurred in a few high-rise buildings where the interviewers were unable to gain entry. Since assignment to an experimental condition came after screening, the experimental groups remain comparable despite the relatively low response rate to the screening interview; however, the generalizability of the results may be limited. The screening interview gathered information on the race, gender, age, and Hispanic background of persons living at the dwelling. The screeners yielded information about 10,998 persons, of whom 3,141 were within the eligible age range (i.e., 15 to 35 years old at the beginning of the field period for the experiment).

Clinic Sample. Two Chicago health clinics agreed to cooperate in the study by providing the names of women who had had abortions during the preceding year or so. The time period was defined so that no one would be selected who had had an abortion during the 3-month period before the beginning of data collection for the experiment. (Because the field period for the experiment was delayed, this window of eligibility in fact ended more than 9 months before the experiment began.) The clinic sample was

also restricted to women who lived in the city of Chicago; the eligible age range, however, was expanded slightly relative to that for the area probability sample to include women between the ages of 15 through 40. The two clinics provided a total of 1,088 names. A systematic sample of 732 of these women was selected for the experiment (using procedures described below).

To protect the confidentiality of the women selected from the clinic sample, the first author carried out the selection of both the clinic and area probability samples and was the only person aware of the sample from which the individual cases had been selected. In addition, the interviewers carried out an after-the-fact permission form procedure in which they asked women who completed the interview to sign a release form giving the researchers access to their medical records at their sources of gynecological care. Women from the clinic sample who refused to sign the permission form were dropped from the analysis and their data were eliminated from the data files. A total of 48 members of the clinic sample were dropped for this reason.

Selection of Cases for the Experiment. Between the area probability and clinic samples, a total of more than 4,200 persons was available for the experiment. A subsample of 2,266 were randomly assigned to a treatment cell. Within the area probability sample, the selection of persons for the experiment required several steps. In the first step, each household with eligible members was placed in one of six strata that were defined by gender, age (15 through 19 versus 20 and over), and minority group membership. Households with members in more than one eligible group were randomly assigned to a single stratum. Because it was impractical to interview more than one person from the same household, only one eligible household member was retained for the main study. Then, after each household had been assigned to a single stratum, a systematic sample was selected; the use of a systematic procedure assured that the members of the sample were drawn from all of the area segments. Altogether, 1,564 cases were selected for the experiment from the area probability sample.

The selection process for the clinic cases was considerably simpler than that for the area probability cases. Once each woman on the clinic lists had been classified by age category and minority status, a systematic sample was selected from each group. In total, 732 women were selected from the clinic lists. Table 1 shows the number of cases selected for the experiment by source and stratum.

TABLE 1. *Initial sample sizes.*

Stratum	Source		
	Area	Clinic	Total
Younger minority women	237	82	319
Older minority women	549	398	947
Other younger women	52	22	74
Other older women	372	230	602
Younger men	29	--	29
Older men	325	--	325
Total	1,564	732	2,296

Response Rates. A few names provided by the clinics turned out to be duplicates; in addition, the screening data regarding a person's age were sometimes in error and some members of the sample had moved outside of Chicago before the field period began. After these losses, 1,914 women and 350 men remained eligible for the study. After the sample was fielded, it became necessary to subsample males as a cost-saving measure; ultimately, only 100 men were interviewed. Table 2 shows the response rates for the study; the overall response rate for women was 55.2 percent. More than two-thirds of the nonrespondents were cases who were never contacted (primarily because they could not be located during the field period); of the women who were contacted, about 85 percent completed an interview.

Of the 354 completed clinic cases, 48 refused to sign permission forms and the permission forms for another 6 women were lost; data for these cases were dropped, leaving 300 clinic cases for the analysis.

Although the response rates for the study seem relatively low, they are comparable to the rates obtained by NSFG in large metropolitan areas; in Cycle IV of NSFG, the response rate for the 10 largest metropolitan areas was below 60 percent (Rieger et al. 1991). It is difficult to predict the overall impact of nonresponse on the obtained levels of reporting. It seems likely that persons who are very active sexually are less willing to take part in surveys than persons with few partners; similarly, women who have had an abortion may be more reluctant to take part than other women. As table 2 indicates, the response rate was lower for the clinic sample than for the area sample. As a result of such differences in

TABLE 2. *Response rates.*

Group	Number eligible	Number complete	Response rate
Males	350	100	--
Females	1,914	1,059	55.2%
Black	1,053	652	61.7%
White	569	270	47.5%
Hispanic	191	88	46.1%
Other	101	49	48.5%
Area	1,191	705	59.1%
Clinic	723	354	48.8%

response rates, the figures presented here probably underestimate levels of sexual activity. Underreporting would, in addition, increase this bias.

Experimental Design

Five variables were manipulated in this experiment in a completely crossed design. Two of the variables, interviewing staff and version of the questionnaire, were attempts to enhance the medical context of the interview; it was thought that respondents might be more willing to discuss sensitive topics in a survey if the context reinforced the health-related purposes of the study and if medical practitioners administered the questions.

Accordingly, interviewing staff was varied, comparing nurses and nursing assistants with regular field interviewers. The hypothesis was that nurses would elicit more reports of sensitive behaviors than would regular field interviewers. The two versions of the questionnaire included the same items but varied the order in which two sets of abortion questions appeared. In one version, a series of pregnancy history questions came first; in the other, a set of questions about medical procedures was first. In the pregnancy first version, the topic of abortion was initially raised during a series of questions about the respondent's pregnancy history; in the medical conditions first version, abortion was first mentioned in connection with a series of medical procedures affecting reproduction. The study tested the hypotheses that more abortions would be reported by respondents receiving the medical procedures questions first and that

the combination of the two sets of abortion questions would yield more reported abortions than either set of questions alone.

The experiment varied the mode of data collection, comparing paper-and-pencil to computer-assisted interviews, and the method of administration, comparing interviewer-administered to self-administered interviews. Crossing the mode of data collection and method of administration resulted in four groups: interviewer-administered paper-and-pencil interviews (PAPI); computer-assisted personal interviews (CAPI); paper-and-pencil self-administered questionnaires (SAQ); and computer-assisted self-administered questionnaires (CASI). It was hypothesized that respondents in both self-administered conditions (those completing the SAQ or CASI questionnaires) would report higher levels of sensitive behaviors.

The site of data collection was varied so that interviews were conducted either in the respondent's home or at a neutral site. Levels of reporting were expected to be higher in the neutral site interviews, where other members of the household could not overhear the answers. A variety of sites were used for the interviews conducted outside the home, with the offices of the National Opinion Research Center and neighborhood restaurants being the most frequent.

Instruments

At the beginning of each interview, the respondent was asked to note three or four important personal events on a calendar to help date events later in the questionnaire. Both versions of the questionnaire began by asking demographic questions. These were followed by the medical procedures and pregnancy history questions in counterbalanced order; both of these series of questions included items on abortion. The pregnancy history questions were the questions usually used on NSFG, and asked the respondent to list all her pregnancies in order and to report certain data about each pregnancy, including its outcome (i.e., live birth, stillbirth, ectopic pregnancy, miscarriage, or abortion). The medical procedures questions were developed for this experiment and asked whether the respondent had had any of a number of medical procedures affecting reproductive health. Six of the procedures were methods for inducing an abortion: dilation and curettage (D and C) to end a pregnancy; dilation and evacuation (D and E) or suction curettage to end a pregnancy; injection of saline solution or prostaglandin to end a pregnancy;

hysterectomy to end a pregnancy; hysterectomy during a pregnancy; and abortion, type unknown.

For the remaining topics, the two versions of the questionnaire were identical. Both versions contained numerous questions about the respondent's sexual behavior. Items asked when and with whom the respondents first had sexual intercourse, and whether it was voluntary; other items asked about the number of sex partners during the previous year, the previous 5 years, and in total. The questionnaires also contained items on whether respondents had ever had an STD. In the section of questions on medical conditions, respondents were asked whether they had had chlamydia, gonorrhea, genital warts, genital herpes, or syphilis. Finally, there were items asking the respondents about their use of condoms in the last year and the last 30 days.

The questionnaires also included a series of items on illicit drug use. The initial drug question asked whether the respondent had ever used any illegal drug, and followup questions asked about their use of marijuana, amphetamines, barbiturates, tranquilizers, psychedelics, cocaine, crack, and heroin. Another series of questions, for users of injectable drugs, asked how they cleaned their needles and related drug paraphernalia, and how often they shared them with other users.

RESULTS

The discussion of the results focuses on sexual behaviors. More specifically, the research examined the average number of sexual partners reported as a function of the gender of the respondent and of the experimental variables; responses were then examined on the other sexual topics in the questionnaire, including STDs and condom use. The results bearing on abortion reporting are discussed elsewhere (Jobe et al., in press). Because so few men completed the interview, reports here are mainly the results for women. Because the emphasis is on comparisons between the different experimental groups, the results reported here are unweighted.

Reported Sexual Partners

The data on the number of reported sex partners are counts and, as is common with such data, the distribution of the responses is highly skewed. To compensate for this departure from normality, 0.5 was added to the values and a logarithmic transformation carried out before

the analyses of variance were performed. For ease of interpretation, untransformed values are reported in the group means. For respondents who had been sexually active for only 1 year, or for only 5 years, the number of sexual partners for that period was used as the value for questions about longer time spans.

Experimental Effects. For all three time periods, women who completed SAQs reported more sexual partners than women who responded to questions administered by an interviewer. There were significant effects for the method of administration for reported partners during the past year, the past 5 years, and the respondent's lifetime. For the past year, the women who answered self-administered questions reported a mean of 1.72 sexual partners versus 1.44 for those who answered questions administered by an interviewer ($F(1,39) = 9.30$, $p < 0.01$). For the 5-year period, women who completed SAQs reported a mean of 3.87 sexual partners versus 2.82 for those who answered interviewer-administered questions ($F(1,39) = 5.74$, $p < 0.05$). For the lifetime item, women who completed SAQs reported a mean of 6.51 sexual partners versus 5.43 for those who answered questions administered by an interviewer ($F(1,39) = 9.54$, $p < 0.01$). No other main effects were significant.

Computerization seemed to interact with the site of the interview to affect the number of sexual partners reported. During home interviews, more sexual partners were reported by women interviewed using computer-assisted questionnaires than by those responding to conventional, paper-and-pencil questionnaires; for women interviewed outside the home, more sexual partners were reported on the pencil-and-paper questionnaires. Table 3 displays the relevant means. For the previous year, women interviewed at home reported fewer sexual partners on the paper-and-pencil questionnaires than on the computer-assisted ones (1.36 versus 1.84), whereas the women interviewed outside the home reported more partners on the paper-and-pencil than on the computer-assisted questionnaires (1.68 versus 1.43; $F(1,39) = 7.72$, $p < 0.01$). Similarly, for the lifetime partners question, women interviewed at home reported fewer partners on the paper-and-pencil than on the computer-assisted questionnaires (5.06 versus 7.48), whereas those interviewed outside their homes showed the opposite pattern, reporting more partners on the paper-and-pencil than on the computer-assisted questionnaires (6.26 versus 5.08; $(F1,39) = 5.89$, $p < 0.05$). The pattern is in the same direction but not significant for the 5-year partners item. Overall levels of reporting are consistently higher using computer-assisted questionnaires, although not significantly so. Bringing computers into

TABLE 3. *Average number of reported sexual partners by mode and site.*

	At home		Outside the home	
	Paper	Computer	Paper	Computer
1 year	1.36	1.84	1.68	1.43
5 years	2.81	4.51	3.33	2.74
Lifetime	5.06	7.48	6.26	5.08

NOTE: Means based on untransformed counts.

the respondents' homes may have fostered a sense of the importance or objectivity of the survey, promoting fuller reporting of sexual partners. Outside the home, especially in public places, the computer may make respondents feel conspicuous, inhibiting reporting.

Males versus Females. As has been observed in earlier surveys on sexual behavior, the men reported more opposite-sex sexual partners than the women did. This was true for the past year (4.19 for the men versus 1.58 for the women), the past 5 years (12.47 versus 3.34), and lifetime (23.96 versus 5.97); all 3 differences are highly significant (F values all greater than 10; p values all less than 0.001). In the analyses that include the data for men, the main effect of self-administration remains significant and that variable does not interact with sex. (The sex of the respondent did occasionally enter into higher order interactions with the experimental variables, but none of these interactions was readily interpretable.)

Rounding of Values. Morris (1993) has argued that the discrepancy between men and women in the reported number of sexual partners largely reflects differences within the subgroup of respondents with a relatively large number of partners to report; within this subgroup, the differences between men and women may reflect differences in rounding behavior (with the women rounding their answers down and the men rounding theirs up). Figure 1 shows the distribution of the number of lifetime sexual partners reported by the men in the current authors' study. (The results for the women, which are not shown, are quite similar.) The preponderance of reported values that are exact multiples of 5 strongly suggests that respondents of both sexes are reporting their answers in round numbers. More than 57.2 percent of the 145

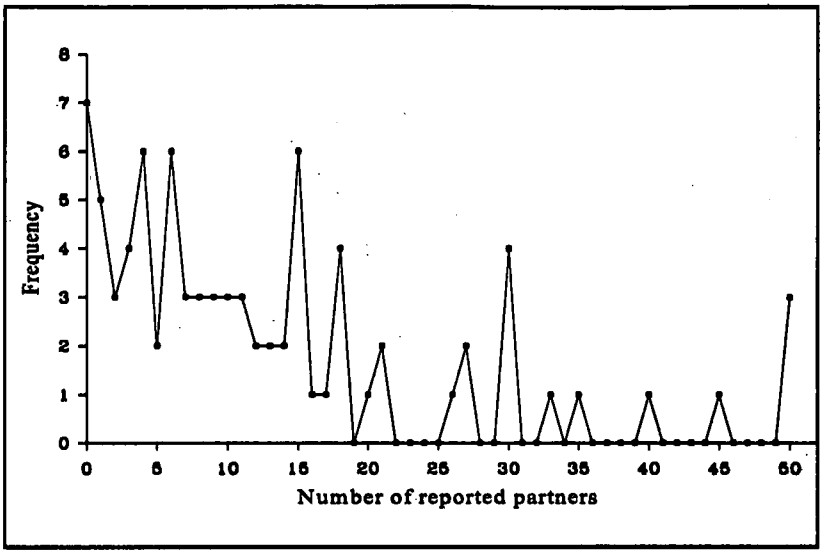


FIGURE 1. *Distribution of lifetime sexual partners (men only).*

respondents who reported 11 or more sexual partners gave an answer that is a multiple of 5.

Other Sexual Topics

Women who answered self-administered questions reported more STDs than those answering questions administered by an interviewer (22.0 percent versus 17.0 percent). This effect of the method of administration was only marginally significant ($\chi^2_1 = 2.93$, $p < 0.10$). No other main effects or interactions were significant. Results were the same for both the logistic regression models and chi-square tests.

The ratio between two items concerning condom use and sexual intercourse in the last 30 days was analyzed; the ratio represented the percentage of time the respondent used a condom in the past month. An analysis of variance was performed to examine this variable. Women who reported that they had not had sexual intercourse in the last 30 days were dropped from this analysis; data from 641 women were included in the analysis. Significantly more condom use was reported with self-administered questionnaires (average reported use 47 percent of the time) than with interviewer-administered questions (35 percent). The main effect for the method of administration variable was significant ($F(1,39) = 8.18$, $p < 0.001$). Apparently, many respondents still regard

the use of condoms as embarrassing behavior. No other significant effects were found on the condom use variable.

DISCUSSION

Effects of Self-Administration

The variable with the most consistent impact on the level of reporting was the method of administering the questions. Women who completed SAQs reported more sexual partners, more STDs, and greater use of condoms than those who responded to questions read by an interviewer. These findings are summarized in table 4, which displays the ratio between the levels of reporting under the self-administered and interviewer-administered conditions. As the table shows, the levels of reporting are substantially higher—from 19 to 37 percent higher—when the questions are self-administered. The effects of self-administration are similar for men. The lack of effects for the site of the interview suggests that respondents may be more concerned about the reactions of the interviewer than about the threat of other family members' overhearing.

TABLE 4. *Reported sexual behavior.*

	Method of administration		
	Self-administered	Administered by interviewer	Ratio
Sexual partners			
Past year, women	1.72	1.44	1.19
Past year, men	4.52	3.88	1.16
Past 5 years, women	3.87	2.62	1.37
Past 5 years, men	14.72	10.43	1.41
Lifetime, women	6.51	5.43	1.20
Lifetime, men	22.76	25.00	0.91
Condom use (women)			
Past 30 days	46.7%	35.3%	1.32
Past year	23.8%	17.9%	1.33
STDs (women)	22.0%	17.0%	1.29

The findings on the impact of self-administration are quite consistent with the results of earlier comparisons of SAQs with face-to-face interviews carried out by field interviewers. The largest studies comparing the two methods of data collection are those reported by Schober and colleagues (1992) and by Turner and colleagues (1992). Both showed that self-administration resulted in higher levels of reported use of illicit drugs. In both studies, the effect of self-administration was restricted to recent as opposed to lifetime drug use; unfortunately, the questionnaire asked only about lifetime use and no effects were found for the experimental variables on drug reporting. The impact of self-administration was not entirely uniform across topics. For example, no significant effect was found on abortion reporting (Jobe et al., in press), even though several earlier studies (such as London and Williams 1990; Mott 1985) found that self-administered questions increased abortion reporting. It is not clear why these earlier findings could not be replicated.

Effects of Site and Medical Context

In contrast to the clear results for self-administration, few effects were observed for the site of the interview. In addition, no effects were observed for either of the attempts to induce a medical context for the questions.

Several studies have attempted to observe the impact of the presence of other family members on reports of sensitive behaviors. For example, in two studies on illicit drug use reporting, interviewers noted whether other family members were present during the interview (Schober et al. 1992; Turner et al. 1992); neither study found an effect of this variable on reported drug use. Mosher and Duffer (1994), on the other hand, report an effect for the site of the interview on abortion reporting. It may be that the effects of this variable are hard to observe consistently. As has already been suggested, respondents may be worried less about the reactions of other household members than about those of the interviewer. In addition, respondents may live alone, or with others (e.g., infants) whose presence is not a cause for concern. Such circumstances will reduce the impact of the site of the interview and make it difficult to demonstrate the effect of this variable.

Neither the version of the questionnaire nor the type of interviewer collecting the data had any discernible effects on reporting. These variables may have made little impression on the respondents. The nurses did not wear distinctive uniforms and, although they introduced themselves as

nurses, this fact probably did not remain very salient to the respondents as the interview progressed. It is also quite possible that respondents see nurses and other medical personnel as authority figures and are no more willing to make embarrassing revelations to them than to ordinary survey interviewers. Several studies demonstrate that respondents admit more sensitive behaviors in an SAQ than they do in interviews conducted by medical personnel (e.g., Boekeloo et al. 1994; Locke et al. 1992); these results suggest that respondents withhold sensitive information from medical personnel just as they do with field interviewers.

Effects of Computerization

Computerization by itself had no consistent effects on levels of reporting among the respondents. Instead, the effects of computer assistance seemed to vary somewhat by the topic of the question and the site of the interview (see table 3). In reports on sexual partners, computer assistance seemed to increase the number of partners reported when data collection took place in the home, but it reduced the number reported when data collection took place outside the home. No compelling explanation for this mode by site interaction suggests itself.

Past investigations of computerized interviewing have tended to emphasize its effects on item nonresponse, timeliness, and cost rather than on the answers that are obtained. Only a few studies have reported effects of computer-assisted data collection on levels of reporting. The experiment comparing CAPI with conventional paper-and-pencil data collection on the National Longitudinal Study of Labor Market Behavior/Youth Cohort found that more respondents reported using birth control under CAPI than under paper-and-pencil interviewing (Baker and Bradburn 1991). Several other studies have shown effects on reporting for computer-assisted self-administration, but in these studies, it is impossible to disentangle the effects of computerization from those of self-administration (e.g., Waterton and Duffy 1984). It appears that computerization by itself has little effect on the answers respondents give, a conclusion consistent with much of the previous literature on computer-assisted telephone interviewing (Groves and Mathiowetz 1987). Nevertheless, computerization may have subtle effects on the respondent, effects that can vary depending on the circumstances of the interview.

REFERENCES

Abernathy, J.; Greenberg, B.; and Horvitz, D. Estimates of induced abortion in urban North Carolina. *Demography* 7:19-29, 1970.

Aquilino, W., and LoSciuto, L. Effect of interview mode on self-reported drug use. *Public Opin Q* 54:362-395, 1990.

Baker, R.P., and Bradburn, N.M. "CAPI: Impacts on Data Quality and Survey Costs." Paper presented at the Public Health Conference on Records and Statistics, Washington, DC, 1991.

Boekeloo, B.; Schiavo, L.; Rabin, D.; Conlon, R.; Jordan, C.; and Mundt, D. Self-reports of HIV risk factors at a sexually transmitted disease clinic: Audio vs written questionnaires. *Am J Public Health* 84:754-760, 1994.

Bradburn, N.M. Response effects. In: Rossi, P.; Wright, J.; and Anderson, A., eds. *Handbook of Survey Research*. New York: Academic Press, 1983.

Bradburn, N.; Sudman, S.; and Associates. *Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*. San Francisco: Jossey-Bass, 1979.

Catania, J.; Gibson, D.; Chitwood, D.; and Coates, T. Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. *Psychol Bull* 108:339-362, 1990.

Groves, R., and Mathiowetz, N. A comparison of CATI and non-CATI questionnaires. In: Thornberry, O., ed. *An Experimental Comparison of Telephone and Health Interview Surveys*. National Center for Health Statistics, Vital and Health Statistics, Series 2, No. 106. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1987.

Hochstim, J. A critical comparison of three strategies of collecting data from households. *J Am Stat Assoc* 62:976-989, 1967.

I-Cheng, C.; Chow, L.P.; and Rider, R.V. The randomized response techniques as used in the Taiwan outcome of pregnancy study. Publication 33, No.11:265-269. New York: Population Council, 1972.

Jobe, J.B.; Pratt, W.F.; Tourangeau, R.; Baldwin, A.; and Rasinski, K. Effects of interview mode on sensitive questions in a fertility survey. In: Lyberg, L., et al., eds. *Survey Measurement and Process Quality*. New York: Wiley, in press.

Jobe, J.B.; Tourangeau, R.; and Smith, A.F. Contributions of survey research to the understanding of memory. *Appl Cog Psychol* 7:567-584, 1993.

Jones, E., and Forrest, J. Underreporting of abortions in surveys of U.S. women: 1976 to 1988. *Demography* 29:113-126, 1992.

Judkins, D.; Mosher, W.; and Botman, S. *National Survey of Family Growth: Design, Estimation, and Inference.* National Center for Health Statistics, Vital and Health Statistics, Series 2, Vol. 109. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1991.

Locke, S.E.; Kowaloff, H.B.; Hoff, R.G.; Safran, C.; Popovsky, M.A.; Cotton, D.J.; Finkelstein, D.M.; Page, P.L.; and Slack, W.V. Computer-based interview for screening blood-donors for risk of HIV transmission. *JAMA* 29:1301-1305, 1992.

London, K., and Williams, L. "A Comparison of Abortion Underreporting in an In-person Interview and Self-administered Questionnaire." Paper presented at the Annual Meeting of the Population Association of America, Toronto, Canada, May 1990.

Lucas, R.W.; Mullen, P.J.; Luna, C.B.X.; and McInroy, D.C. Psychiatrist and computer interrogators of patients with alcohol-related illnesses: A comparison. *Br J Psychiatry* 131:160-167, 1977.

Marquis, K., and Cannell, C. *Effect of Some Experimental Techniques on Reporting in the Health Interview.* National Center for Health Statistics, Vital and Health Statistics, DHEW Pub. No. 1000, Series 2, No. 41. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1971.

Morris, M. Telling tails explain the discrepancy in sexual partner reports. *Nature* 365:437-440, 1993.

Mosher, W.D, and Duffer, A.P., Jr. "Experiments in Survey Data Collection: The National Survey of Family Growth Pretest." Paper presented at the meeting of the Population Association of America, Miami, FL, May 1994.

Mott, F. "Evaluation of Fertility Data and Preliminary Analytic Results from the 1983 Survey of the National Longitudinal Surveys of Work Experience of Youth." Columbus, OH: National Institute of Child Health and Development by the Center for Human Resources Research, January 1985.

Rieger, S.; Judkins, D.; and Sperry, S. *National Survey of Family Growth: Cycle IV CATI Phase. Final Report.* Hyattsville, MD: National Center for Health Statistics, 1991.

Robinson, R., and West, R. A comparison of computer and questionnaire methods of history-taking in a genito-urinary clinic. *Psychol Health* 6:77-84, 1992.

Schober, S.; FeCaces, M.; Pergamit, M.; and Branden, L. Effects of mode of administration on reporting of drug use in the National Longitudinal Survey. In: Turner, C.; Lessler, J.; and Gfroerer, J., eds. *Survey Measurement of Drug Use: Methodological Studies.* Rockville, MD: National Institute on Drug Abuse, 1992. pp. 267-276.

Shimizu, I., and Bonham, G. Randomized response technique in a national survey. *J Am Stat Assoc* 73:35-39, 1978.

Smith, T. Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. *Soc Biol* 39:203-211, 1992.

Turner, C.; Lessler, J.; and Devore, J. Effects of mode of administration and wording on reporting of drug use. In: Turner, C.; Lessler, J.; and Gfroerer, J., eds. *Survey Measurement of Drug Use: Methodological Studies.* Rockville, MD: National Institute on Drug Abuse, 1992. pp. 177-220.

Warner, S. Randomized response: A survey technique for eliminating evasive answer bias. *J Am Stat Assoc* 60:63-69, 1965.

Waterton, J., and Duffy, J. A comparison of computer interviewing techniques and traditional methods for the collection of self-report alcohol consumption data in a field survey. *Int Stat Rev* 52:173-182, 1984.

AUTHORS

Roger Tourangeau, Ph.D.
Senior Scientist
National Opinion Research Center
1350 Connecticut Avenue, NW
Washington, DC 20036

Jared B. Jobe, Ph.D.
Chief
Cognitive Functioning and Aging
National Institute on Aging
7201 Wisconsin Avenue
MSC 9205
Bethesda, MD 20892-9205

William F. Pratt, Ph.D.
Division of Vital Statistics
National Center for Health Statistics (Retired)

Kenneth Rasinski, Ph.D.
Senior Survey Methodologist
National Opinion Research Center
1155 East 60th Street
Chicago, IL 60637

167355

Mode of Interview and Reporting of Sensitive Issues: Design and Implementation of Audio Computer-Assisted Self-Interviewing

Judith T. Lessler and James M. O'Reilly

ABSTRACT

Substantial underreporting is typical in interviewing respondents on their drug use and other sensitive behaviors. This chapter reviews established strategies, self-administered questionnaires and indirect questioning techniques, for increasing the willingness of respondents to report stigmatizing behaviors. While these methods improve reporting, each has shortcomings and burdens which limit their effectiveness. A new computer-based self-interviewing approach which incorporates recorded audio playback of questions offers improved self-administered interviewing. The chapter discusses this technology, audio computer-assisted self-interviewing (audio-CASI), describing its features and positive results from the early research tests of the method.

INTRODUCTION

Drug use is a highly sensitive issue and requires a continual search for new means to both assess and improve the accuracy of self-reported use. In this chapter, the authors briefly discuss the use of various interviewing methods to ask about sensitive behaviors and then describe in detail the design of audio computer-assisted self-interviewing (audio-CASI) and the results of two experiments that compared audio-CASI to other interviewing procedures. Research has generally shown that more private methods of interviewing yield higher reports of sensitive behaviors. The two major approaches that have been adopted to increase the willingness of respondents to report stigmatizing behaviors are the self-administered questionnaire and indirect questioning techniques.

The basic problem with trying to gather information on stigmatizing behaviors is that people do not want to talk about them. In a survey, the respondents might want to conceal their behavior from a number of

entities including the general public, sponsors of the survey (universities, the government), interviewer, and other members of his or her household. Respondents may be concerned with disclosure of specific activities that have specific legal and social consequences or they may have general concerns about how they appear to others. To protect respondents from disclosure of their personal information to the general public and the institutional sponsors, survey research organizations have adopted a number of techniques.

- Requirements that survey protocols be reviewed by institutional review boards (IRBs) to ensure that regulations covering the protection of human subjects are followed and stipulations of the privacy act are met.

- Routine use of confidential data-processing techniques that separate names and addresses from files containing personal information.

- Procurement of specific confidentially pledges from interviewers and staff who have access to the survey data and identifiers.

Interestingly, few respondents are likely to have direct experience that these activities are actually taking place and must rely on the assurances of confidentiality given to them along with explanations of the procedures that are used to maintain this confidentiality. However, researchers' claims of confidentiality are probably enhanced by repeated exposure in the media to reports of surveys and scientific studies in which no person is specifically named. Given that many of the questions that are asked in the National Household Survey on Drug Abuse (NHSDA) focus on illegal behaviors, it is somewhat surprising that anyone reports any illegal drug use. Assurances of confidentiality and appeals to the need for the information that will ultimately contribute to the social well-being of society in many cases seem to outweigh the concerns about self-revelation.

QUESTIONING TECHNIQUES FOR SENSITIVE ISSUES

Both self-administered questionnaires (SAQs) using the sealed ballot approach and indirect questioning techniques serve to conceal the respondent's answers from the interviewer and other household members. In contrast to the above-mentioned procedures, which the respondent must more or less accept on faith, these procedures are often

designed to explicitly demonstrate their privacy-enhancing features. For example, no names are written on SAQs, interviewers stand where they cannot see respondents mark their answers, questionnaires are placed in sealed envelopes, and attempts are made to secure a private place within the home for the interview.

In randomized response, two questions, one sensitive and one not sensitive, are available to the respondent. The respondent uses a randomizing device to select the question to answer (Warner 1965). The interviewer records the answer without being aware of which question was chosen. In item count methods (Droitcour et al. 1991), respondents are given lists of behaviors in which the sensitive behavior is imbedded among a list of nonsensitive behaviors. Respondents indicate the number of the behaviors that apply to them rather than answering questions on the actual behaviors. Random parts of the sample receive lists with and without the sensitive behavior. Each of these methods allows the researcher to use statistical methods to estimate the total number of people who engaged in the sensitive behavior; however, they do not allow one to determine if a particular person engaged in the sensitive behavior. Because of this feature, indirect questioning methods also prevent disclosure to the general public and sponsor as well as to the interviewer and other persons who may be nearby.

Question structure in interviewer-administered questionnaires (IAQs) has also been shown to have an impact on reporting of sensitive behaviors (Bradburn and Sudman 1979; Groves 1989). Open-ended questions, longer questions, and questions incorporating wording that implies that the behavior is more or less common are techniques that have been used to improve response to sensitive questions.

RESEARCH ON SAQs AND INDIRECT QUESTIONNAIRE TECHNIQUES

Research has generally shown that SAQs and indirect questioning techniques yield higher reports of sensitive behaviors (Bradburn 1983; Catania et al. 1990; Miller et al. 1990; Schwarz et al. 1991). For example, in the case of SAQs, Hay (1990) found differences in reported consumption of alcoholic beverages and cigarette use in a study of some 1,500 students in grades 2 through 12 who were randomly assigned to receive either an SAQ or a personal interview. The differences were 74 versus 63 percent for ever using alcohol and 38 versus 30 percent for use of

cigarettes. Turner and colleagues (1992), in a large-scale field experiment in which 3,200 respondents were randomly assigned to either an interviewer or SAQ, found that the difference between the two modes of data collection increased as the sensitivity of the behavior increased. Table 1 shows the ratio between the proportion of SAQ respondents reporting a given behavior to the proportion of respondents reporting that behavior when the interviewer administered the questions. The table displays the results for three time periods and three types of drug.

TABLE 1. *Ratio of prevalence estimates from SAQs and IAQs.*

Drug type	Lifetime	Past 12 months	Past 30 days
Alcohol	0.99	1.04	1.06
Marijuana	1.05	1.3	1.38
Cocaine	1.06	1.58	2.4

Examining table 1, one notes that the superiority of the SAQ relative to the IAQ increases as admitting drug use becomes more sensitive. For alcohol, the ratios are approximately equal to one for all time periods. For marijuana, the ratio is very close to one for lifetime use, indicating that respondents are nearly as willing to report use of marijuana in answer to an IAQ as when answering an SAQ as long as they are talking about use at some time in their life; however, as the reference period becomes more proximate, they are less willing to report use of marijuana to an interviewer. For cocaine, which use is more stigmatized than marijuana, a similar pattern emerges with even larger differences between the SAQ and IAQ; respondents completing an SAQ are nearly 2.5 times more likely to report using cocaine during the past 30 days.

Similarly, Bradburn (1983) notes that randomized response has been demonstrated to yield higher reports of drug use, abortion, and degree of fault in automobile accidents. Miller (1986) found item count techniques resulted in higher reports of heroin use although later tests revealed little differences for marijuana and cocaine use (Droitcour et al. 1991).

LIMITATIONS OF SAQs AND INDIRECT QUESTIONING

There are difficulties with each of these approaches. For SAQs, the most obvious difficulty is that they require that the respondent can read.[1] In addition, the respondents must complete a number of the questionnaire administration tasks such as finding and reading instructions, implementing skip patterns, and marking answers. In addition, respondents are prone to the same types of errors seen in IAQs: missing, out-of-range, and inconsistent answers. Even if a respondent can read, branching or contingent questioning is a particular problem (Turner et al. 1992), and researchers have been advised to use question structures that eliminate branching (Messmer and Seymour 1982). Although attention to the graphical design of the questionnaire has potential to reduce branching errors (Jenkins and Dillman 1994), incorporating branching options may compromise respondents' willingness to report sensitive behaviors in the SAQ (Gfroerer 1994).

Difficulties with indirect questioning techniques include respondents' failure to understand and accept the methods, availability of measurements at the aggregate rather than individual level, and high variance of the resulting estimates. Groves (1989) notes that there has been little research on whether respondents actually believe that the randomized response protects their privacy or on the degree to which respondents implement the procedure correctly. Hubbard and associates (1989) indicated that some respondents had difficulty understanding the privacy-enhancing features of item count techniques and were suspicious of them. In addition, it was demonstrated that in spite of detailed explanations of how to implement the technique, respondents made errors, often responding with the number of the item (i.e., its position in the list) rather than the number of items that applied to them. Although it is possible to use randomized response and item count procedures to make subgroup estimates, for some behaviors that have very low prevalence (and that often are also the most sensitive), the higher variance of these procedures reduces their usefulness in studying subgroup differences. Thus, because of these difficulties, direct questioning using SAQs is often selected over indirect questioning in a survey.

COMPUTER-ASSISTED METHODS FOR SELF-ADMINISTERED QUESTIONING

Computer-assisted self-interviewing (CASI) and audio-CASI systems have been developed to overcome some of the difficulties associated with the response to SAQs. With CASI, respondents read questions as they appear on the screen and enter their answers with the keyboard (or some other input device). The computer takes care of the "housekeeping" or administrative tasks for the respondent. The advantages of CASI are automated control of complex question routing, the ability to tailor questions based on previous responses, real-time control of out-of-range and inconsistent responses, and the general standardization of the interview.

CASI possesses significant disadvantages, however. Most obviously, CASI demands that the respondent can read with some facility. A second, more subtle disadvantage is that, at least with the character-based displays of many CASI applications of today, the visual and reading burden imposed on the respondent appears to be much greater than with an attractively designed paper form. The size of the characters and other qualities of the computer user interface seem to demand more reading and computer screen experience than that possessed by many who might be competent readers of printed material. Graphical user interfaces may reduce or eliminate this problem, but the present software used to developed CASI applications usually lacks this feature.

By adding simultaneous audio renditions of each question and instruction aloud, audio-CASI can remove the literacy barriers to self-administration of either CASI or SAQ. In audio-CASI, an audio box is attached to the computer; respondents put on headphones and listen to the question and answer choices as they are displayed on the screen. Respondents have the option of turning off the screen so that people coming into the room cannot read the questions, turning off the sound if they can read faster than the questions are spoken, or keeping both the sound and video on as they answer the questions. Respondents can enter a response at any time and move to the next question without waiting for completion of the audio question and answer choices for a question.

The advantages of audio-CASI, then, are that the addition of audio makes CASI fully applicable to a very wide range of respondents. Persons with limited or no reading abilities are able to listen, understand, and respond to the full content of the survey instrument. Observers of

audio-CASI interviews also often report that even with seemingly strong readers, audio-CASI interviews seem to more effectively and fully capture respondents' concentration. This may be because wearing headphones increases the insulation of the respondent for external stimuli, and also may be explained by the fact that the recorded human voice in the audio component evokes a more personalized interaction between the respondent and the instrument.

CASI AND AUDIO-CASI RESEARCH

Comparisons of CASI with personal interviews have noted findings similar to those cited above for the comparison of SAQs to IAQs. Waterton and Duffy (1984) compared reports of alcohol consumption under CASI and personal interviews. Overall, reports of alcohol consumption were 30 percent higher under the CASI procedure, and reports of liquor consumption were 58 percent higher. This may understate the potential gains because in this study respondents were first asked by an interviewer whether they had consumed any alcoholic beverages in the past 7 days. Only those respondents who indicated that they had done so received the CASI interview.[2]

Several recent studies comparing CASI to personal interviews in clinic settings have also noted the superiority of this method. Locke and associates (1992) found significant differences between the reporting of risk behaviors for the human immunodeficiency virus (HIV) when CASI was used to administer questions to donors at an American Red Cross donor center (4.4 percent versus 0.3 percent in the traditional interview procedure). Robinson and West (1992) compared reporting of symptoms in a genitourinary clinic using CASI, SAQs, and physician interviews. They found that more symptoms were reported by computer than by paper, and that both methods found more than were found in physician interviews. Levine and colleagues (1989) found that patients who had been admitted to a hospital after harming themselves were more likely to report suicidal ideation in a computer interview than to a physician. The CASI version of the Diagnostic Interview Schedule (DIS) yielded diagnostic information consistent with the traditional interviewer-administered DIS and patients considered the computer contact to be less embarrassing (Erdman et al. 1992); a computer interview with sex offenders yielded large numbers of previously undetected crimes (Weinrott and Saylor 1991); and a comparison of clinician and computer interviews directed at identifying obsessive-compulsive disorders found that the two methods

were equally good at distinguishing those with the disorder and that patients showed no preference for clinician interviews (Rosenfield et al. 1992).

The current authors have participated in two experimental tests that compared audio-CASI with other forms of interviewing. O'Reilly and colleagues (1994) compared paper SAQs, CASI, and audio-CASI in a small-scale experiment designed to assess the technology's potential. Subjects answered questions on drug use, sexual behaviors, and income. A Greco-Latin square design was used to assign subjects to one of three interviewing modes for each topic, producing an experiment that was fully balanced across mode and content. For eight of nine rating scales comparing these modes, respondents reported a preference for one of the two CASI methods. Although the sample size was small, a total of 40, O'Reilly and colleagues found that the two CASI methods tended to produce significantly more reports of marijuana and cocaine use; few differences in sexual behaviors were found. Table 2 summarizes some of the results.

Respondents were also asked which method they thought was better and consistently rated the two CASI methods as better on eight of nine facets rated: "liked best," "best for asking sensitive questions," "easiest to change answers," "most interesting," "easiest to use," "best for getting honest answers," "best for privacy after interview," "best for privacy during the interview," and "overall preference." Respondents felt it was easier to change answers using paper-and-pencil SAQs. Audio-CASI was rated consistently higher than CASI; however, the difference was significant only for three items: "overall preference," "interest," and "ease of use."

ABORTION REPORTING IN THE NATIONAL SURVEY OF FAMILY GROWTH PRETEST

With funding from the National Center for Health Statistics (NCHS), scientists at NCHS, Battelle, and the Research Triangle Institute (RTI) collaborated in a formal field experiment that compared abortion reporting under three different interviewing conditions. Respondents were randomly assigned to receive either an in-home computer-assisted personal interview (CAPI) interview only, an in-home CAPI interview followed by a audio-CASI interview that asked additional questions about abortions, or a CAPI interview at a neutral site away from the respondent's home. Respondents in the audio-CASI treatment were first asked to report their abortions to

TABLE 2. *Proportion of respondents reporting use of drugs by interviewing method.*

	Interviewing method			
	Audio-CASI	CASI	Paper SAQ	P
Alcohol				
Past 30 days	0.43	0.68	0.46	0.82
Past 12 months	0.64	0.76	0.62	0.65
Ever in lifetime	0.86	0.92	0.77	0.02[1]
Marijuana				
Past 30 days	0.21	0.17	0.00	0.09[1]
Past 12 months	0.29	0.6	0.08	0.04[1]
Ever in lifetime	0.64	0.83	0.46	0.10[1]
Cocaine				
Past 30 days	0.00	0.00	0.00	
Past 12 months	0.07	0.08	0.00	0.31
Ever in lifetime	0.29	0.33	0.00	0.03[1]
Ns[2]	14	12	13	

KEY: 1 = Paper SAQ different from CASI and audio-CASI at $p < 0.10$ by t-test. CASI and audio-CASI not significantly different from each other by same test. 2 = Ns shown are the minimum sample size for calculation of any proportion shown in the column.

SOURCE: Data from O'Reilly et al. 1994.

the interviewer during a section of the CAPI interview that asked about the outcome of each pregnancy that they ever had. The question asked:

"Now I'd like to ask some questions about your [N-TH] pregnancy. Please look at Card B-1. Thinking about your [N-TH] pregnancy, in which of the ways shown on Card B-1 did the pregnancy end? (READ LIST. CODE ALL THAT APPLY.)

"Miscarriage? (Occurs naturally, during the first 6 months of pregnancy),

"Stillbirth?	(Baby born dead after 7 or more months of pregnancy),
"Abortion?	(Induced during the first 6 months of pregnancy; include D&C, vacuum extraction, suction, and saline injections),
"Ectopic pregnancy?	(Occurs outside the uterus or womb),
"Live birth by Cesarean section?	
"Live birth by vaginal delivery?	(Includes delivery through natural or induced labor)"

At the end of the interview, respondents were trained in the audio-CASI procedures and were asked additional questions on abortion.

The field experiment included a comparison of audio-CASI, in-home CAPI, and out-of-home CAPI. It was hypothesized that women's willingness to report sensitive information would be increased if they were interviewed outside of their homes because in earlier rounds of the survey respondents had indicated a concern that family members would overhear their responses. An incentive experiment was also included. The out-of-home respondents were paid $40 and the in-home respondents received either no incentive or a $20 incentive.

The National Survey of Family Growth (NSFG) is the major source of information in the United States on pregnancy, family formation, contraceptive use, and childbearing. Prior rounds of the NSFG identified significant underreporting of abortion (Jones and Forrest 1992), and the absence of good information on abortion presents considerable difficulty to analysts who are attempting to understand the relationship between sexual activity, contraceptive use, contraceptive failure, and childbearing. This difficulty in obtaining accurate reports of abortion (and other sensitive behaviors) was the main motivation for the experimental comparison of alternative modes of data collection.

Table 3 compares the results from the audio-CASI question on whether a woman had ever had an abortion and both the pilot questions and pregnancy outcome questions (in section B). There was one refusal of the audio-CASI. Six additional women reported having had an abortion at some time in their life in the audio-CASI interview, which represents a

14 percent increase in the number of women reporting ever having had an abortion.

TABLE 3. *Relationship of abortion reporting in the CAPI and the audio-CASI interview, NSFG Cycle V Pretest.*

	Abortion reported as a birth outcome		
Audio-CASI: Ever had an abortion	Yes	No	N
Yes	42	6	48
No	0	129	129
Total	42	135	177

Results showed that abortion reporting was also increased in the out-of-home interviews, and a higher proportion of the respondents who received an incentive reported having had an abortion. In addition, some women who reported an abortion in section B reported additional abortions in the audio-CASI interview. In all cases when there was a difference in the number of abortions reported between the CAPI interview and the subsequent audio-CASI interview, more abortions were reported, indicating that the different numbers of abortions reported in the audio-CASI is probably not due to random error.

The current authors also fit a series of logistic regression models to determine if there were significant differences due to interviewing conditions. Independent variables included the type of interview (CAPI only, audio-CASI, or neutral site), incentive for in-home interviews (none or $20), race/ethnicity (Hispanic, black, non-Hispanic, non-black), marital status (married, not married), income (unknown, greater than $20,000, or other), and age. A stepwise selection procedure was used in which an independent variable that was significant at the 0.15 level was added to the model. Table 4 summarizes the results.

Based on these results, it was concluded that both the neutral site and the audio-CASI increase the number of women who report that they ever had an abortion. In addition, the incentive has a marginal effect; however, it is not possible to determine if the incentive directly affects willingness to report or if higher reports in this group are due to the higher response rates and a different population of women being included.

TABLE 4. *Analysis of the impact of characteristics of women and interview conditions on abortion reporting, NSFG Cycle V Pretest.*

	Parameter estimate	Standard error	Probability	Odds ratio
Intercept	-2.52	0.49	0.0001	1.081
$20 incentive	0.38	0.27	0.1348	1.488
Married	-0.34	0.23	0.1428	0.714
Age	0.03	0.01	0.0264	1.033
Audio-CASI	0.54	0.27	0.0419	1.723
Neutral site	0.83	0.31	0.0067	2.294

Respondent Attitudes

When asked about their attitudes toward the alternative methods of reporting abortion, women who received the audio-CASI interview indicated that they preferred the audio-CASI method. Table 5 presents the results.

NEED FOR THE AUDIO COMPONENT OF AUDIO-CASI

The above-mentioned results have not demonstrated the need for the audio component of the audio-CASI system. Except for respondent preferences, the feasibility experiment showed no differences in reporting between the audio-CASI and the CASI treatments; the various clinic experiments achieved superior reporting with CASI, not audio-CASI. However, no one can dispute the fact that respondents who cannot read will not be able to complete an SAQ or CASI interview on their own. The solution that survey researchers used in SAQs was to have the interviewer read the questions and responses while the respondent marked the answers. This technique has also been used in a recently reported study by Boekeloo and associates (1994) in which respondents in a sexually transmitted disease (STD) clinic were randomly assigned to complete a self-administered interview either by reading the questions themselves or by marking answer sheets while listening to questions using a cassette player and headphones. These authors found that the

TABLE 5. *Respondents' attitudes toward methods of reporting abortion. NSFG Cycle V Pretest.*

	Response	Percent respondent
How do you rate telling the interviewers your answers to questions on abortion?	Poor	15.2
	Fair	20.3
	Good	30.5
	Very good	17.5
	Excellent	16.4
How do you rate using the computer and earphones to answer questions on abortion?	Poor	2.8
	Fair	8.5
	Good	17.5
	Very good	26.0
	Excellent	45.2
Which method of answering questions on abortion is the most private?	Earphones and computer	62.7
	No difference	32.2
	Telling the interviewer	4.5
	Don't know	0.6
Which method do you recommend for the main study?	Interviewer	16.9
	Computer	58.2
	Do not ask about abortion	2.8
	Does not matter	22.0

audio interview yielded more complete data and identified by more HIV risk behaviors.

Reading questions to respondents, however, completely precludes the use of contingent questioning because branching to the correct followup questions would violate the privacy of the respondent's answers. In addition, reading questions aloud even if the interviewer does not know the answers has the potential to compromise the respondent's ability to conceal responses from household members. While those who can overhear the interviewer reading the questions will be similarly ignorant of the answers, the respondent is subject to a subsequent interrogation as to what the answer was after the interview is complete. This is the reason to obtain a private place for conducting sensitive interviews. The desire to conceal answers from other members of the household may be the factor that is operating to produce the finding from the NSFG Cycle V pretest that those who were interviewed outside the home reported more abortions.

CONCLUSION

Overall, the authors believe that audio-CASI is superior to methods that have been traditionally used to gather data on sensitive issues. It provides the same privacy enhancements that traditional SAQs do and makes it easier to use contingent questioning because it avoids the difficulties associated with having respondents implement complex skip instructions. In contrast to indirect questioning techniques, it allows researchers to know if a particular respondent (who may be anonymous) reported the sensitive behavior, which facilitates analysis of the relationship between the sensitive behaviors and other characteristics. In addition, audio-CASI allows researchers to ask questions in any language of any respondent who can see and hear. Literacy on the part of the respondent is not required. Finally, it is noted that audio-CASI is suitable for use in a variety of settings, including clinics and households.[3]

NOTES

1. The National Adult Literacy Survey (NALS) was conducted in 1992 using a nationally representative sample of 13,600 persons aged 16 and older. Literacy was measured in terms of five proficiency levels on three scales—prose, document, and quantitative. The survey found that the percentage of adults in the lowest level of proficiency was 21 percent for prose literacy, 23 percent in document literacy,

and 22 percent in quantitative literacy (National Center for Education Statistics 1993).

2. In the literature, this study is often reported as a CAPI study. It was actually a CASI study in which computers were taken into the homes and respondents asked to enter their responses on selected questions while the interviewer stood in a part of the room that did not permit observation of the respondent's answers.

3. Since this chapter was originally written, audio-CASI has been used in the homes of over 10,000 women who have responded to Cycle V of the NSFG. As of September 1995, that survey is continuing, and it is expected that more than 11,000 women will have used the audio-CASI by the completion of data collection.

REFERENCES

Bradburn, N.M. Response effects. In: Rossi, P.; Wright, J.; and Anderson, A., eds. *Handbook of Survey Research*. New York: Academic Press, 1983.

Boekeloo, B.O.; Schaivo, L.; Robin, D.; Conlon, R.; Jordon, C.; and Mundt, D. Self-reports of HIV risk factors by patients at a sexually transmitted disease clinic: Audio vs written questionnaires. *Am J Public Health* 84(5):754-760, 1994.

Bradburn, N.M, and Sudman, S. *Improving Interview Methods and Questionnaire Design*. San Francisco: Jossey-Bass, 1979.

Catania, J.A.; Gibson, D.R.; Chitwood, D.D.; and Coates, T.J. Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. *Psychol Bull* 108(3):339-362, 1990.

Droitcour, J.; Caspor, R.; Hubbard, M.; Parsley, T.; Vissher, W.; and Ezzati, T. The item count technique as a method of indirect questioning: A review of its development and a case study application. In: Biemer, P., et al., eds. *Measurement Error in Surveys*. New York: Wiley, 1991.

Erdman, H.P.; Klein, M.H.; Greist, J.H.; Skare, S.S.; Husted, J.J.; Robins, L.N.; Helzer, J.E.; Goldring, E.; Hamburger, M.; and Miller, J.P. A comparison of two computer-administered versions of NIMH Diagnostic Interview Schedule. *J Psychiatric Res* 26(1):85-95, 1992.

Forsyth, B.; Lessler, J.T.; and Hubbard, M. Cognitive evaluation of the questionnaire. In: Turner, C.; Lessler, J.; and Gfroerer, J., eds. *Survey Measurement of Drug Use: Methodological Studies*. Rockville, MD: National Institute on Drug Abuse, 1992.

Groves, R.M. *Survey Errors and Survey Costs*. New York: Wiley, 1989.

Hay, D.A. Does the method matter on sensitive survey topics? *Survey Methodol* 16(1):131-136, 1990.

Hubbard, M.L.; Casper, R.A.; and Lessler, J.T. Respondent reactions to item count lists and randomized response. *Proceedings of the Survey Research Section of the American Statistical Association.* Washington, DC: American Statistical Association, 1989. pp. 544-548.

Jenkins, C.R., and Dillman, D.A. *Combining Cognitive and Motivational Research Perspectives for the Design of Respondent-Friendly Self-Administered Questionnaires*. Washington, DC: U.S. Bureau of the Census, 1994.

Jones, E.F., and Forrest, J.D. Underreporting of abortion in surveys of U.S. women: 1976 to 1988. *Demography* 29(1):113-126, 1992.

Lessler, J., and Holt, M. Using response protocols to identify problems in the U.S. census long form. *Proceedings of the Survey Research Section of the American Statistical Association*. Washington, DC: American Statistical Association, 1987.

Levine, S.; Ancill, R.J.; and Roberts, A.P. Assessment of suicide risk by computer-delivered self-rating questionnaire: Preliminary findings. *Acta Psychiatrica Scand* 80(3):216-220, 1989.

Locke, S.E.; Kowaloff, H.B.; Hoff, R.G.; Safran, C.; Popovsky, M.A.; Cotton, D.J.; Finkelstein, D.M.; Page, P.L.; and Slack, W.V. Computer-based interview for screening blood-donors for risk of HIV transmission. *JAMA* 268(10):1301-1305, 1992.

Messmer, D.J., and Seymour, D.T. The effects of branching on item nonresponse. *Pub Opin Q* 46(2):270-277, 1982.

Miller, J. "A New Technique for Surveying Deviant Behavior: Item-Count Estimates of Marijuana, Cocaine, and Heroin." Paper presented at the American Association of Public Research Conference, St. Petersburg, FL, May 1986.

Miller, H.G.; Turner, C.F.; and Moses, L.E., eds. Methodological issues in AIDS Surveys. In: *AIDS: The Second Decade*. Washington, DC: National Academy Press, 1990.

O'Reilly, J.M.; Hubbard, M.; Lessler, J.; Biemer, P.; and Turner, C.F. Audio and video computer-assisted self-interviewing: Preliminary tests of new technologies for data collection. *J Official Stat* 10(2):197-214, 1994.

Robinson, R., and West, R. A comparison of computer and questionnaire methods of history taking in a genito-urinary clinic. *Psychol Health* 6:77-84, 1992.

Rosenfeld, R.; Dar, R.; Anderson, D.; and Kobak, K. A computer administered version of the Yale-Brown Obsessive-Compulsive scale. *Psychol Assess* 4:329-332, 1992.

Schwarz, N.; Strack, F.; Hippler, H.; and Bishop, G. The impact of administration mode on response effect in survey measurement. *Appl Cog Psychol* 5:193-212, 1991.

Turner, C.F., and Lessler, J.T. "Effects of Mode of Administration and Wording on Data Quality" In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Rockville, MD: National Institute on Drug Abuse, 1994.

Turner, C.F.; Lessler, J.T.; and Devore, J.W. "Effects of Mode of Administration and Wording on Reporting of Drug Use." In: Turner, C.; Lessler, J.; and Gfroerer, J., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Publication (ADM)92-1929. Rockville, MD: National Institution Drug Abuse, 1992.

Warner, S. Randomized response: A survey technique for eliminating evasive answer bias. *J Am Stat Assoc* 60:63-69, 1965.

Waterton, J.J., and Duffy, J.C. A comparison of computer interviewing techniques and traditional methods in the collection of self-report alcohol consumption data in a field survey. *Int Stat Rev* 52:173-182, 1984.

Weinrott, M.R., and Saylor, M. Self-report of crimes committed by sex offenders. *J Interpersonal Violence* 6(3):286-300, 1991.

AUTHORS

Judith T. Lessler, Ph.D.
Director
Statistics Research Division

James M. O'Reilly, Ph.D.
Coordinator of Technology and Marketing

Research Triangle Institute
3040 Cornwallis Road
P.O. Box 12194
Research Triangle Park, NC 27709-2194

Privacy Effects on Self-Reported Drug Use: Interactions With Survey Mode and Respondent Characteristics

William S. Aquilino

ABSTRACT

This chapter examines the impact of interview privacy on self-reported illicit drug use. In 1991, interviews were completed with an urban-suburban sample of 2,417 adults aged 18 to 45. Results show that the presence of third parties during the interview significantly influences respondents' willingness to reveal illicit drug use. Among married respondents, presence of a spouse resulted in higher reporting of illicit drug use, while the presence of adults other than the spouse had a consistent negative effect on drug use reports. A parent's presence during the interview significantly reduced respondents' willingness to report illicit drug use. The pattern of findings suggests that the direction of effects due to third party presence is linked to two factors: the extent of the third party's knowledge of the information requested, and the degree of personal stake the third party may have in the respondent's answers. The differential impact of privacy by interview mode was also examined. Tests of interactions between privacy and interview mode failed to support the hypothesis that the use of self-administered answer sheets reduces privacy effects compared with interviewer-administered interviews.

INTRODUCTION

Most professional survey organizations attempt to conduct personal interviews out of earshot of others (Aquilino 1993) and instruct their interviewers to move to a private location before beginning an interview. Nonetheless, a substantial portion of interviews in most household surveys (often from 25 to 50 percent) are conducted with others nearby or able to overhear the interview (Bradburn and Sudman 1979). Because interviewers are essentially guests in the respondent's home, they often find it hard to insist on complete privacy without jeopardizing the respondent's goodwill.

Interviewers often cannot control the behavior of other household members while the interview is underway.

Although the presence of a third party is a common occurrence in household surveys, their influence on responses to sensitive questions has received relatively little attention in the survey literature on response effects (Aquilino 1993). There is little agreement among extant empirical studies concerning either the direction or magnitude of privacy effects. Some studies have reported significant effects (e.g., Casterline and Chidambaram 1984; Taietz 1962); others have found few or no effects (e.g., Anderson and Silver 1987; Zanes and Matsoukas 1979); while some have reported mixed results (Bradburn and Sudman 1979). The literature on this topic suffers from a lack of any theoretical framework to guide research or explanation. No research has described how privacy effects might differ among modes of interview, and little attention has been given to possible interactions of interview privacy and respondent characteristics in producing response effects.

This chapter examines the impact of interview privacy on self-reported illicit drug use in three interview modes: self-administered, interviewer-administered face-to-face, and telephone. The study was designed to develop and evaluate a theoretical framework predicting the magnitude and direction of expected third-party effects. The primary research questions in this effort were:

1. Does the presence of others during the interview influence respondents' willingness to reveal the lifetime use of illicit substances?

2. Do privacy effects differ according to the identity of the person(s) present, their knowledge of the respondent's past behavior, and their personal stake in learning of the respondent's past drug use?

3. Does the impact of the presence of others on response tendencies differ by mode of interview (including self-administered, face-to-face, and telephone survey)?

4. Does the impact of lack of privacy during the interview differ by respondent characteristics such as sex, age, education, race/ethnicity, and marital/cohabitation status?

EMPIRICAL LITERATURE AND A THEORETICAL FRAMEWORK

The literature on response effects due to lack of interview privacy is surprisingly sparse. Several studies have concluded that privacy effects are small or nonexistent in most surveys. Anderson and Silver (1987) reported that the answers of married couples interviewed together were not more similar than those of couples interviewed apart on both factual and attitudinal questions. Tendencies of respondents to overreport voting behavior did not vary by the presence of others (Silver et al. 1986). With a sample of adolescents from a single school, Zanes and Matsoukas (1979) found that the presence of parents in the same room during the interview had little impact on adolescents' reports of illicit drug use.

Bradburn and Sudman (1979), in a national sample of 1,200 adults, also found no consistent pattern of privacy effects on responses to a variety of sensitive questions. They did report, however, that item nonresponse to sensitive questions was higher when others were present. The presence of a child during the interview also appeared to diminish respondents' willingness to admit they or their friends had ever used marijuana. The overall conclusion of this study was that evidence for third-party presence effects is weak and that lack of privacy does not greatly threaten the validity of sensitive surveys.

In contrast to Bradburn and Sudman's conclusions, several studies have found that the presence of others does affect response tendencies. Studies by Casterline and Chidambaram (1984) and Taietz (1962) found that third-party presence increased tendencies toward socially desirable responses—saying things that would please the person present. Strong privacy effects have been found for adolescents' self-reported drug use. Gfroerer (1985) analyzed data from the 12- to 17-year-old respondents in the 1979 and 1982 National Household Survey on Drug Abuse (NHSDA) surveys and found strong evidence that parents' presence during the interview resulted in less reporting of illegal drug use by adolescents. The use of anonymous self-administered questionnaires for most NHSDA drug categories did not prevent or diminish the influence of parental presence during the interview. Similarly, adolescents have been found to underreport drug use when identifying information is included on the questionnaire cover (Malvin and Moskowitz 1983). Both studies suggest that privacy concerns may be central to the validity of sensitive interviews with adolescents.

Lack of privacy does not always push responses in the direction of increased social desirability (or decreased willingness to reveal sensitive information). Based on a large national sample of married couples, Aquilino (1993) reported that, when spouses were present during the interview, subjective assessments of the utility of marriage were more positive. Higher estimates of spouse contributions to housework were obtained, and men gave lower estimates of the likelihood of marital dissolution. But spouse presence was also linked to a greater willingness to report sensitive factual information concerning the marriage; respondents were more likely to report cohabiting with the spouse before marriage, and self-reported levels of marital conflict were higher. Thus, lack of privacy increased social desirability bias only for subjective assessments of the marriage. Effects were in the opposite direction (more candor) when questions tapped events and behaviors. It is important to note that spouse-presence effects were found despite the use of self-administered forms for items concerning marriage.

This pattern of spouse-presence findings is consistent with a hypothesis of third-party effects proposed by Mitchell (1965), who suggested that when factual information is requested in a survey, the presence of others who are knowledgeable about the subject matter of the interview may actually increase the accuracy of responses, even to sensitive questions. Mitchell hypothesized that it may be harder for respondents to misrepresent (or forget) factual information when someone who knows the truth is nearby.

If this hypothesis is correct, it also suggests that the identity of the others present might moderate third-party effects. The presence of those with the most knowledge of the respondent's behaviors or experiences should elicit more accurate reporting of factual information than would the presence of those with minimal knowledge of the respondent's experiences. A related issue is the extent to which the person present has a stake in learning of the respondent's answers (e.g., wives who would be affected by hearing husbands' assessments of marital relations). The greater the personal stake of the third party in the respondent's answers, the more third-party presence would tend to elicit socially desirable responding.

These propositions suggest that determining whether third-party presence during the interview will affect responses to sensitive questions depends on the answers to a number of questions:

1. Do the survey items ask for factual reports on events and behaviors, or do they ask for subjective assessments of attitudes, feelings, or relationships?

2. If factual reports are requested, how much knowledge does the person present have about the events or behaviors in question? Does the third party know what the respondent's answer should be?

3. If the third party doesn't have knowledge of the factual information requested, or if the interview requests subjective assessments of feelings or relationships, how will the person present be the affected by respondent's answers? Does the third party have a stake in how the respondent answers the question? To what extent will the respondent be concerned about how the listener might react to his or her survey responses, especially if the information would be new to the person overhearing the survey responses?

A first hypothesis based on this framework would be that when purely factual information is requested and the third party has full knowledge of the events or behaviors under question, third-party presence will lead to more accurate reporting of sensitive information (Mitchell 1965). For example, Aquilino (1993) found respondents more likely to report cohabiting before marriage if their spouse was present than if interviewed in private. If sensitive factual information is requested and the third party does not have knowledge of the events or behaviors, third-party presence should not lead to more accurate reporting.

A second prediction of this framework would be that if the third party does not have knowledge of the factual information requested and has a personal stake in the respondent's answers to these sensitive questions, responses will be pulled in the direction of more social desirability (i.e., pulled in the direction that would tend to please the listener). If the questions and answers are irrelevant to the listener, survey responses should be less affected by lack of privacy.

A similar argument can be made for subjective survey questions: If the third party has a stake in learning the respondent's subjective assessments or perceptions, responses will tend toward pleasing the third party (more

social desirability bias). If the subjective assessments are irrelevant to the third party, responses should not be affected by lack of privacy.

Application to This Research

The analyses reported in this chapter estimate the effects of spouse presence, child presence, parent presence, and presence of other adults (relatives or nonrelatives) on self-reports of lifetime drug use among 18- to 45-year-old respondents. The dependent variables request only factual information from the respondent: whether they had ever used marijuana, cocaine, psychotherapeutic drugs, or an illicit drug of any type. The theoretical approach outlined above suggests that these different types of listeners may have different effects on responses to sensitive drug questions based on their knowledge of the respondent's lifetime drug use and their stake in learning about the respondent's past drug use. If the third party already has full knowledge of the respondent's drug use, third-party presence should increase the likelihood of respondents' revealing their lifetime drug use. The presence of third parties who have no knowledge of respondents' past drug use would not increase the probability of positive drug use reports; however, if the third party has little knowledge of respondent drug use and has a great personal stake in learning of such usage, his or her presence should decrease respondents' willingness to reveal illicit drug use.

Among the four categories of potential listeners considered here, spouses (or partners in cohabiting unions) likely have the most knowledge of respondents' past drug use. Many couples may have used illicit drugs together during the dating and courtship phases of the relationship or after marriage or cohabitation. There may be little reluctance among many married or cohabiting couples to reveal past experimentation with drug use, especially when such usage was a short-lived phenomenon of youth. Thus, spouse presence should have a positive impact on drug use reports.

Parents, on the other hand, may be the least likely listeners to have full knowledge of respondents' drug use, and also have the greatest stake in learning about it. Many parents might feel personally responsible and deeply troubled by their child's drug use. It is safe to assume that most children (minor or grown) conceal illicit drug use from parents, most of whom would disapprove. Parents' lack of knowledge and personal stake in the outcome suggest that their presence will decrease respondents' willingness to reveal past drug use.

The analyses in this research focus primarily on spouse or partner presence and parent presence. Presence of children and of other relatives or nonrelatives are included in the models as control variables. This was done so that effects of spouse presence and parent presence could be estimated, controlling for the influence of anyone else who may have been nearby during the interview.

Privacy Effects and Mode of Interview

The possible connection between interview mode and privacy effects has not been investigated in earlier research. Recent studies (Aquilino 1992, 1994; Gfroerer and Hughes 1991; Johnson et al. 1989; Turner et al. 1992) have shown that self-administered interviews, in-person interviewer-administered interviews, and telephone interviews yield different estimates of self-reported drug and alcohol use when effects due to sampling and screening are controlled. Survey mode effects appear strongest among minorities (Aquilino 1994), especially among African Americans.

This chapter explores one of the avenues by which survey mode might influence responses: Different interview modes may either exacerbate or suppress the potential influence of privacy on respondents' willingness to reveal illicit or undesirable behaviors. In particular, the use of self-administered answer sheets to maximize response anonymity during the interview should decrease problems of self-presentation (Sudman and Bradburn 1974). The self-administered format may reduce or eliminate effects due to the presence of others during the interview, compared to interviewer-administered surveys. Multivariate analyses test this prediction.

Privacy Effects and Respondent Characteristics

Very little is known about variation in privacy effects by respondent characteristics such as age, sex, race/ethnicity, or education; interactions of third-party presence with background characteristics have not been tested. The research described here tested for such interactions in all drug use models, but no predictions were made about the direction of possible interaction effects.

METHODS

The Data Set

The data were collected from June through December, 1991. Interviews were completed with 2,417 adults aged 18 to 45 drawn through a multi-stage area probability sample of the 37 largest standard metropolitan statistical areas (SMSAs) in the coterminous United States (an urban-suburban sample). These SMSAs contain about 36 percent of the U.S. population, and had a minimum size of 1.88 million inhabitants. African Americans and Hispanics were double sampled. The sample was restricted to younger adults to maximize the chances of interviewing current and recent users of illicit drugs (12- to 17-year-olds were not included in the sample because of cost constraints). A screening response rate of 94.3 percent and an interview response rate of 80.6 percent were achieved.

Experimental Design and Controls

Screening and Respondent Selection. All households in the sample were screened in person for eligibility. One respondent was randomly selected if more than one adult aged 18 to 45 resided in the household. All respondents in the study were selected using identical sampling, screening, eligibility, and respondent-selection procedures.

Assignment to Mode. Each housing unit in the sample was randomly assigned to one of three interview modes: (1) self-administered questionnaire (SAQ), a face-to-face interview using self-administered answer sheets for drug and alcohol items; (2) personal/no SAQ, a face-to-face interview in which all questions were asked and responses recorded directly by the interviewer; and (3) telephone—the interview was conducted by telephone from the interviewer's home. Households without telephones were excluded from the analyses (N = 169).

Questionnaire. The questionnaire was adapted from the 1990 NHSDA questionnaire; Spanish translation was based on the NHSDA Spanish translation. Question wording, question order, and response categories were identical in all three modes. No show cards were used in the SAQ or personal modes to ensure comparability to the telephone mode.

The SAQ mode used the standard NHSDA procedures for self-administered answer sheets. Answer sheets were sealed in an envelope in the

respondent's presence upon completion of the interview. No names were recorded on the questionnaires or answer sheets.

Interviewers. The same interviewers conducted the interviews in all three modes. About one-third of each interviewer's assignment was done in person with SAQs, one-third in person without SAQs, and one-third by telephone from the interviewer's home.

Experienced interviewers were recruited for this study. As a group, they had an average of more than 11 years interviewing experience. Nearly all interviewers were women; their average age was 48 years, with 14.5 years of school completed (82 percent of the interviewers had at least some college, and about one-third were college graduates). More than 85 percent of the interviews were conducted by someone with previous interviewing experience in drug use surveys.

Multivariate Analyses

Dependent Variables. Four binary dependent variables reflecting lifetime drug use were selected, all coded 1 = yes, 0 = no. The questions were: ever used any illicit drug; ever used marijuana or hashish; ever used cocaine; and ever used pills (nonmedical use of prescription drugs). The "any illicit drug" category includes use of marijuana, cocaine, nonmedical use of prescription drugs (sedatives, tranquilizers, stimulants, and analgesics), inhalants, hallucinogens, and heroin. Thus, this variable indicates the respondent's willingness to reveal any use of an illicit substance.

The dependent variables were restricted to lifetime use due to sample size and use prevalence. For cocaine and pill use especially, there were too few past-year and past-month users to derive reliable estimates of third-party effects.

Independent Variables. When the interview was conducted in the personal or SAQ modes, interviewers recorded (after leaving the respondent's household) who was present during all or part of the interview. Upon completion of the telephone interview, respondents were asked to report who was present or able to overhear the conversation. Four binary independent variables were constructed for third-party presence, all coded 1 = present some or all the time, 0 = not present: spouse/partner present, child present, parent present, and other relative or nonrelative present. A categorical variable for interview mode was

included: personal/no SAQ (the omitted category), SAQ, and telephone. Control variables in the models included sex, race/ethnicity, age, education, cohabitation status (only in models for married/cohabiting respondents), household income, and employment status. In the models predicting lifetime drug use, age (in years) and years of education completed were entered as continuous variables.

It is important to note that the analyses and interpretation of results focus primarily on the effects due to spouse presence and parent presence. Although terms for child presence and other presence are included in the models, they will be treated more as control variables than as independent variables in describing results.

Logistic regression was used to estimate the effects of third-party presence, respondent characteristics, mode of interview, and interaction terms on the dependent variables. To control for household composition, two separate sets of analyses were conducted. The first estimated effects due to spouse/partner presence and the sample was restricted to couple households where the respondent was either currently married or cohabiting (N = 1,118). The second set estimated effects due to parent presence, and the sample was restricted to cases in which the respondent resided with a parent (N = 521). Those living with parents were primarily younger respondents; about 60 percent of this subsample was between ages 18 and 25. After testing the main effects of interview privacy in the models, interaction terms of presence variables by interview mode, age, sex, race/ethnicity, education, and cohabitation status were tested. Case weights were computed and used in the regression analyses and population estimates of lifetime drug use. Marginally significant findings ($p < 0.10$) are noted in the tables of results, but these estimates should be interpreted with caution.

The significance tests in the regression analyses have not been adjusted to reflect two sources of variation: the clustered nature of the sample and the use of case weights. Thus, the true standard errors may be somewhat larger than those reported. The large size of the 37 primary sampling units and the fact that random assignment to survey mode was done within clusters should reduce the effects of sample design on the estimated standard errors. To compensate for not controlling sample design effects in the analyses, two-tailed tests of statistical significance were used in evaluating directional hypotheses (where one-tailed tests would have been appropriate). This raised the critical values needed to achieve significance.

RESULTS

The Likelihood of Third-Party Presence

The proportion of interviews with a spouse/partner or parent present is shown in table 1. Spouses were present for at least some of the time in about one in four interviews with married respondents (throughout the rest of the chapter the terms "spouse" and "married" are intended to include partners and cohabiting couples as well). This is nearly identical to the proportion found in the National Survey of Families and Households (NSFH) sample of nearly 7,000 married couples (Aquilino 1993). Parents were present in about one in six cases where respondents lived with parents.

Spouse Presence. The likelihood of third-party presence varies by respondent and household characteristics. Logistic regression models predicting spouse presence and parent presence are given in table 2. Consistent with previous research (Aquilino 1993), spouse presence was more common when husbands were interviewed, less common when wives were interviewed. Wives may simply be more likely to be at home while husbands are being interviewed, and may have a greater interest in the proceedings than men whose wives are being interviewed. Spouse presence was less likely among African Americans than among whites or Hispanics (15 percent compared to 25 percent), a pattern also found among NSFH married couples (Aquilino 1993). The consistency of results across two diverse samples suggests that this is not an artifact. However, the causes of the racial/ethnic difference are unclear. Spouse presence was most likely among the least educated respondents and among those with the lowest family incomes (33 percent among those with less than $10,000 income in past 12 months). These results parallel findings from the NSFH married couples (Aquilino 1993), which suggest that social class is inversely related to the presence of others during the interview. One possibility is that low-income respondents live in much smaller homes with fewer rooms than do wealthier respondents, making it much more difficult for interviewers to insist on privacy during the interview.

Number of household members was inversely related to spouse presence; the chances of the spouse's being nearby were lower in households of four or more members, compared to two- or three-member households. It may be that there are more distractions and competing

TABLE 1. *Percentage of cases with spouse/partner or parent present during the interview (unweighted estimates).*

	Respondent currently married/cohabiting		Respondent currently living with parent	
	N of cases	% Spouse present	N of cases	% Parent present
Total	1,118	23	521	17
Male	476	30	267	17
Female	642	18	242	17
Hispanic	191	26	77	22
White/other	758	25	305	18
African American	164	15	122	11
Age 18-25	114	22	298	16
26-34	442	26	135	16
35-45	562	22	76	20
Currently married	1,002	24	--	--
Currently cohabiting	115	21	--	--
Education				
Less than high school	136	30	50	24
High school grad	305	22	162	18
Some college	246	24	93	19
College graduate	358	20	80	5
Currently enrolled	55	27	123	18
Work status				
Full time	793	25	313	16
Part time	124	18	94	15
Unemployed	35	20	51	12
Not in labor force	147	18	50	28
Household income				
< $10,000 (omitted)	54	33	30	10
$10,000 - $29,999	228	22	99	24
$30,000 - $49,999	296	26	146	19
$50,000 or higher	456	22	164	12
Income missing	69	17	70	16
Household size				
Two	260	29	76	20
Three	227	27	172	17
Four	370	19	132	14
Five	169	21	69	15
Six+	77	17	60	20
Children coresident				
Yes	804	21	89	18
No	299	29	420	16

TABLE 1. *Percentage of cases with spouse/partner or parent present during the interview (unweighted estimates) (continued).*

	Respondent currently married/cohabiting		Respondent currently living with parent	
	N of cases	% Spouse present	N of cases	% Parent present
Other coresident				
Yes	79	23	72	6
No	1,024	23	437	19
Mode of interview				
Personal, no SAQ	353	27	167	23
SAQ	384	22	166	16
Telephone	366	22	176	11

duties for the spouse in larger households, allowing less free time to monitor the interview.

Parent Presence. Parent presence (among young adults living with parents) did not vary by sex (table 2). Sons and daughters age 18 and older were equally as likely to have a parent nearby during the interview. Consistent with the spouse findings, the racial/ethnic difference was large, with African-American youth only about half as likely as whites or Hispanics to have a parent present. The most educated respondents (college graduates) were by far the least likely to have a parent in the room or able to listen in (5 percent). College graduates may command a bit more respect for their privacy than do less educated sons and daughters, and may be more likely to have the resources to maintain a private telephone line. Surprisingly, however, the pattern for household income was the reverse of that for spouse presence: Respondents in the lowest income households were the least likely to have a parent present during the interview. The reasons for this are not clear.

Parent presence was less likely in households with members other than the respondent and parents, such as siblings, other relatives, or roommates. Again, the suggestion is that other household members may distract or divert parents' attention from the interview.

Finally, mode of interview was strongly related to parent presence. Respondents interviewed by telephone were by far the least likely to

TABLE 2. *Logistic regression models predicting the likelihood of spouse presence and likelihood of parent presence during the interview (standard error in parentheses; unweighted estimates).*

Independent variable	Spouse/partner presence[1]	Parent presence[2]
Sex (female)	-0.64 (0.16)***	0.14 (0.27)
Race/ethnicity		
White/other (omitted)	--	--
African American	-0.57 (0.25)*	-0.84 (0.37)*
Hispanic	-0.15 (0.23)	0.18 (0.37)
Age 18-25	-0.20 (0.27)	-0.09 (0.33)
26-34 (omitted)	--	--
35-45	-0.08 (0.16)	-0.15 (0.41)
Cohabiting	-0.46 (0.27)+	--
Education		
Less than high school	0.61 (0.27)*	0.50 (0.45)
High school graduate	--	--
Some college	0.08 (0.22)	0.19 (0.36)
College graduate	-0.21 (0.22)	-1.29 (0.59)*
Currently enrolled in college	0.33 (0.35)	-0.03 (0.35)
Employment		
Employed full time (omitted)	--	--
Employed part time	-0.30 (0.27)	-0.13 (0.36)
Unemployed	-0.46 (0.46)	-0.29 (0.51)
Not in labor force	-0.10 (0.26)	0.78 (0.41)+
Household income		
Less than $10,000 (omitted)	--	--
$10,000 - $29,999	-0.68 (0.36)+	1.53 (0.72)*
$30,000 - $49,999	-0.42 (0.37)	1.21 (0.74)
$50,000 or higher	-0.72 (0.39)+	0.82 (0.77)
Income missing	-0.94 (0.47)*	0.96 (0.78)
Household size	-0.26 (0.10)*	0.11 (0.13)
Any children coresident	-0.00 (0.25)	-0.13 (0.41)
Any relative/nonrelative coresident	0.20 (0.32)	-1.41 (0.57)*
Interview mode		
Personal, no SAQ (omitted)	--	--
SAQ	-0.25 (0.18)	-0.44 (0.30)
Telephone	-0.26 (0.18)	-0.97 (0.32)**

KEY: 1 = Sample restricted to married and cohabiting respondents (N = 1,118).
2 = Sample restricted to respondents living with parents (N = 521).
+ = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ (2-tailed tests).

have parents listening in. Many homes with more than one telephone have at least one phone in a bedroom or in other more private rooms. Because respondents have a choice of phones (and therefore rooms), it may be easier to achieve privacy in a telephone interview than in the face-to-face interview. In face-to-face mode, the interviewer (as guest) will likely be seated in the more public or shared rooms in the home, such as living room or kitchen, where it becomes more difficult to avoid other family members.

Impact of Privacy on Drug Use Self-Reports

Overview of Main Findings. Multivariate analyses show that third-party presence during the interview significantly influences respondents' willingness to reveal illicit drug use. However, as predicted, the direction of effects depends on the identity of the person present. In the analyses of married and cohabiting respondents (see tables 3 to 7), presence of the spouse or partner had a significant positive effect on the self-reported lifetime use of any illicit drug, marijuana, and cocaine, and nonmedical use of prescription drugs. That is, on all four dependent variables, respondents were more likely to report illicit drug use if the spouse was present than if the interview was conducted in privacy. Significant interactions of spouse presence with respondent's age, race/ethnicity, and sex were found in three of the four models. However, the effects of spouse presence were unrelated to mode of interview.

As the theoretical framework suggested, results were in the opposite direction in the sample of respondents living with parents. A parent's presence during the interview significantly reduced respondents' willingness to report illicit drug use (see tables 8 to 11). Significant negative effects were found for reports of lifetime use of marijuana and any illicit drug. No significant interactions with respondent characteristics were found. Parent-presence effects were linked to mode of interview, but in an unexpected direction: The negative effects were stronger in the SAQ and telephone modes than in the face-to-face mode.

Models for Spouse Presence. Weighted estimates of lifetime drug use by spouse presence, interview mode, and respondent characteristics are given in table 3. Logistic regression models estimating the impact of spouse presence are presented in tables 4 to 7. For the total sample, the presence of the spouse or partner had a consistently positive effect on self-reported drug use on all four dependent variables. These effects are significant in all four logistic regression models with controls for sex,

TABLE 3. *Weighted estimates of drug use by spouse presence and respondent characteristics, married/cohabiting respondents age 18 to 45.*

		Percent who report ever using:							
		Marijuana		Cocaine		Pills[1]		Any illicit drug	
	N of cases	Spouse not present	Spouse present	Spouse not present	Spouse present	Spouse not present	Spouse present	Spouse not present	Spouse present
Total	1,118	57	66	20	32	23	30	61	70
Male	476	65	69	30	33	26	34	66	72
Female	642	51	62	12	29	20	25	56	67
Hispanic	191	37	29	17	29	27	29	45	43
White/other	758	61	74	21	33	24	33	64	77
African American	164	56	55	14	21	11	9	58	55
Age 18-25	114	51	75	19	40	21	29	55	75
25-34	442	63	76	24	39	24	33	67	78
35-45	562	53	54	17	22	23	28	57	61
Education									
Less than HS	137	34	47	13	38	20	30	43	64
HS grad	310	59	68	22	21	27	28	63	70
Some college	248	63	78	23	42	26	37	65	82
College grad.	420	57	64	18	30	19	27	61	65

KEY: 1 = Pills refers to the nonmedical use of four classes of prescription drugs: stimulants, analgesics, tranquilizers, and sedatives.

TABLE 3. *Weighted estimates of drug use by spouse presence and respondent characteristics, married/cohabiting respondents age 18 to 45 (continued).*

		Percent who report ever using:							
		Marijuana		Cocaine		Pills[1]		Any illicit drug	
	N of cases	Spouse not present	Spouse present	Spouse not present	Spouse present	Spouse not present	Spouse present	Spouse not present	Spouse present
Married	1,002	56	65	18	29	22	30	60	70
Cohabiting	115	66	77	35	55	30	34	72	77
Interview mode									
Personal	361	56	66	20	30	20	30	61	71
SAQ	388	58	66	22	39	27	32	62	67
Telephone	369	56	68	18	25	21	29	59	73

TABLE 4. *Impact of spouse presence during the interview on self-reported lifetime use of any illicit drug: Logistic regression models for married/cohabiting respondents age 18 to 45 (N = 1,118; standard errors in parentheses; data are weighted).*

Independent variables	Dependent variable: ever used any illicit drug I	II
Female	-0.49 (0.14)***	-0.51 (0.14)***
White/other (omitted)	--	--
African American (AA)	-0.34 (0.20)+	-0.20 (0.22)
Hispanic	-0.85 (0.20)***	-0.65 (0.22)**
Age	-0.05 (0.01)***	-0.03 (0.01)**
Education (years)	-0.04 (0.03)	-0.04 (0.03)
Cohabiting	0.53 (0.23)*	0.53 (0.23)*
Household income		
< $10,000 (omitted)	--	--
$10,000 - $29,999	0.53 (0.35)	0.48 (0.35)
$30,000 - $49,999	1.10 (0.36)**	1.03 (0.36)**
$50,000 or higher	1.58 (0.37)***	1.55 (0.37)***
Income missing	0.43 (0.42)	0.37 (0.42)
Work full-time (omitted)	--	--
Work part-time	0.14 (0.21)	0.14 (0.21)
Unemployed	0.81 (0.41)+	0.79 (0.41)+
Not in labor force	0.22 (0.21)	0.28 (0.21)
Interview mode		
Personal (omitted)	--	--
SAQ	-0.08 (0.15)	-0.08 (0.15)
Telephone	-0.10 (0.15)	-0.09 (0.15)
Spouse present	0.39 (0.15)*	2.96 (0.88)***
Child present	0.17 (0.15)	0.16 (0.15)
Other adult present	-0.16 (0.26)	-0.20 (0.27)
Significant interaction terms		
AA x spouse present		-0.62 (0.53)
Hispanic x spouse present		-0.85 (0.43)*
Age x spouse present		-0.07 (0.02)**
-2 log likelihood	1552.79	1541.01
Chi-square for model improvement		11.78**
Degrees of freedom		3

KEY: + = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ (2-tailed tests).

TABLE 5. *Impact of spouse presence during the interview on self-reported marijuana use: Logistic regression models for married/cohabiting respondents age 18 to 45 (N=1,118; standard errors in parentheses; data are weighted).*

Independent variables	Dependent variable: ever used marijuana or hashish I	II
Female	-0.65 (0.13)***	-0.68 (0.14)***
White/other (omitted)	--	--
African American (AA)	-0.27 (0.20)	-0.13 (0.22)
Hispanic	-1.06 (0.21)***	-0.75 (0.23)**
Age	-0.05 (0.01)***	-0.03 (0.01)**
Education (years)	-0.03 (0.03)	-0.03 (0.03)
Cohabiting	0.46 (0.22)*	0.46 (0.22)*
Household income		
< $10,000 (omitted)	--	--
$10,000 - $29,999	1.11 (0.40)**	1.02 (0.40)*
$30,000 - $49,999	1.66 (0.42)***	1.55 (0.41)***
$50,000 or higher	2.09 (0.42)***	2.03 (0.42)***
Income missing	0.96 (0.46)*	0.85 (0.47)+
Work full-time (omitted)	--	--
Work part-time	0.14 (0.21)	0.15 (0.21)
Unemployed	1.10 (0.43)**	1.09 (0.43)
Not in labor force	0.21 (0.21)	0.29 (0.21)
Interview mode		
Personal (omitted)	--	--
SAQ	-0.01 (0.15)	-0.01 (0.15)
Telephone	-0.06 (0.15)	-0.05 (0.15)
Spouse present	0.36 (0.15)*	4.06 (0.91)***
Child present	0.10 (0.15)	0.10 (0.15)
Other adult present	-0.22 (0.26)	-0.28 (0.27)
Significant interaction terms		
AA x spouse present		-0.54 (0.54)
Hispanic x spouse present		-1.43 (0.48)**
Age x spouse present		-0.10 (0.03)***
-2 log likelihood	1571.01	1547.64
Chi-square for model improvement		23.37 ***
Degrees of freedom		3

KEY: + = p < 0.10; * = p < 0.05; ** = p < 0.01; *** = p < 0.001 (2-tailed tests).

TABLE 6. *Impact of spouse presence during the interview on self-reported cocaine use: Logistic regression models for married/cohabiting respondents age 18 to 45 (N = 1,118; standard errors in parentheses; data are weighted).*

Independent variables	Dependent variable: ever used cocaine I	II
Female	-0.83 (0.16)***	-1.08 (0.19)***
White/other (omitted)	--	--
African American	-0.63 (0.27)*	-0.64 (0.27)*
Hispanic	-0.10 (0.24)	-0.11 (0.24)
Age	-0.05 (0.01)***	-0.03 (0.01)*
Education (years)	-0.03 (0.03)	-0.03 (0.03)
Cohabiting	0.96 (0.22)***	1.00 (0.22)***
Household income		
< $10,000 (omitted)	--	--
$10,000 - $29,999	0.63 (0.50)	0.66 (0.50)
$30,000 - $49,999	0.79 (0.51)	0.78 (0.51)
$50,000 or higher	1.31 (0.52)*	1.34 (0.53)*
Income missing	-0.25 (0.67)	-0.26 (0.68)
Work full-time (omitted)	--	--
Work part-time	-0.32 (0.28)	-0.30 (0.28)
Unemployed	0.07 (0.47)	0.09 (0.47)
Not in labor force	-0.37 (0.28)	-0.29 (0.29)
Interview mode		
Personal (omitted)	--	--
SAQ	0.18 (0.17)	0.18 (0.17)
Telephone	-0.19 (0.18)	-0.17 (0.18)
Spouse present	0.51 (0.16)**	2.00 (0.88)*
Child present	0.09 (0.17)	0.12 (0.18)
Other adult present	-0.59 (0.34)+	-0.65 (0.34)+
Significant interaction terms		
Sex x spouse present		0.75 (0.32)*
Age x spouse present		-0.05 (0.02)*
-2 log likelihood	1242.20	1230.21
Chi-square for model improvement		12.00**
Degrees of freedom		2

KEY: + = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ (2-tailed tests).

age, race/ethnicity education, income, work status, interview mode, and the presence of household members other than spouse (see model I in tables 4 to 7). There is an especially large presence effect for lifetime cocaine use; lifetime use estimates are over 50 percent higher in cases when the spouse or partner was present.

Significant interactions of spouse presence with respondent characteristics were found in three of the four models. In the models for lifetime use of any illidit drug (table 4) and lifetime marijuana use (table 5), the impact of spouse presence varies by race/ethnicity and age. The positive effect of spouse presence on willingness to reveal past drug use was stronger for whites than for Hispanics or African Americans (although the contrast in both models is significant only for Hispanic versus white; the same direction of effects is obtained for African Americans). The lifetime estimates of any illicit drug use given in table 3 show almost no effects at all of spouse presence on minority respondents, but a sizable effect for whites (e.g., admission of ever using illicit drugs increases from 64 percent to 77 percent among whites when spouse was present). Estimates of lifetime marijuana use suggest no spouse-presence effects for African Americans, positive effects for whites, and negative effects for Hispanics.

Presence effects varied by respondent's age in three of the four models (tables 4 to 6; all but the model for pill use). The negative interaction term (spouse presence x age) shows that the positive impact of spouse presence wanes with respondent age. It is the younger married or cohabiting respondents who are most influenced by spouse or partner presence, especially in the 18 to 25 age group. As can be seen in table 3, the drug use differentials by spouse presence are very small or nonexistent for the middle-aged married respondents (ages 35 to 45), but very large (20 percentage points or more) for the youngest group. The youngest group most likely contains a disproportionate share of new marriages; thus age may be tapping duration-of-marriage effects to some extent. Large differentials by spouse presence can also be seen in table 3 for the 26 to 34 age group, although the effects are somewhat smaller than for the youngest group.

A significant sex-by-spouse-presence interaction term was obtained for cocaine use only (table 6). The positive coefficient shows that spouse presence had a significantly larger influence on women than on men. Women were much more likely to report ever having used cocaine if their husband was nearby during the interview (12 percent to 29 percent); wife's presence had little impact on men's reports of cocaine

TABLE 7. *Impact of spouse presence during the interview on self-reported nonmedical use of prescription drugs: Logistic regression models for married/cohabiting respondents age 18 to 45 (N = 118; standard errors in parentheses; data are weighted).*

Independent variables	Dependent variable: Nonmedical use of prescription drugs I
Female	-0.35 (0.15)*
White/other (omitted)	--
African American	-1.09 (0.30)***
Hispanic	0.02 (0.22)
Age	-0.00 (0.01)
Education (years)	-0.02 (0.03)
Cohabiting	0.44 (0.22)*
Household income	
< $10,000 (omitted)	--
$10,000 - $29,999	-0.47 (0.36)
$30,000 - $49,999	-0.25 (0.37)
$50,000 or higher	-0.15 (0.38)
Income missing	-1.09 (0.50)*
Work full-time (omitted)	--
Work part-time	-0.01 (0.23)
Unemployed	0.34 (0.43)
Not in labor force	-0.16 (0.24)
Interview mode	
Personal (omitted)	--
SAQ	0.27 (0.16)
Telephone	0.01 (0.17)
Spouse present	0.31 (0.15)*
Child present	0.05 (0.16)
Other adult present	-0.73 (0.33)*
Significant interaction terms: none	
-2 log likelihood	1373.33

NOTE: There is no model 2 in this table because none of the tested interaction terms were significant at the 0.10 level.

KEY: + = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ (2-tailed tests).

use. The pattern of stronger spouse-presence effects on women can also be seen in the estimates of marijuana use and any illicit drug use (see table 3), although the interaction terms for those two dependent variables are nonsignificant. There were no significant interactions with respondent characteristics for pill use.

Spouse-presence-by-education interaction terms were nonsignificant in all four models. There were no significant interactions with cohabitation status; the privacy effects were the same for cohabiting and legally married respondents.

Tests for the interaction of spouse presence with interview mode did not support the expectation that the use of self-administered answer sheets (SAQ mode) would eliminate or moderate the effects of third-party presence. The impact of spouse or partner presence did not differ significantly among the three interview modes on any of the four dependent measures. Caution is warranted in interpreting these nonsignificant interactions, because the standard errors of these terms were large. This finding needs to be tested in a larger sample.

Models for Parent Presence. The sample of respondents living with a parent is made up predominantly of young adults, with about three-fifths of this group in the 18 to 25 age range. Most coresidence between parents and adult children involves offspring under age 25 (Aquilino 1991) and becomes relatively rare after age 30. The influence of parent presence in this sample was directly opposite to spouse-presence effects, a pattern consistent with the theoretical model described earlier.

Estimates of the lifetime use of any illicit drug and lifetime marijuana use were substantially lower in the parent-present group (see table 8); 55 percent reported ever using marijuana when parents were not present, compared to 30 percent when parents were nearby during the interview. Self-reported use of any illicit drug dropped from 61 percent to 43 percent when parents were present. Results were in the same direction for lifetime cocaine use (dropping from 18 percent to 12 percent with parent present), but the effect was not significant. Reports of pill use were not affected by parent presence.

The impact of parent presence on survey responses did not differ by respondent characteristics on any of the four dependent measures (all interaction terms for parent presence and sex, age, race/ethnicity, and education were nonsignificant). Marginally significant interaction terms

TABLE 8. *Weighted estimates of drug use by parent presence and respondent characteristics, respondents age 18 to 45 living with parents.*

		Percent who report ever using:							
		Marijuana		Cocaine		Pills[1]		Any illicit drug	
	N of cases	Parent not present	Parent present	Parent not present	Parent present	Parent not present	Parent present	Parent not present	Parent present
Total	521	55	30	18	12	17	18	61	43
Male	273	57	31	20	18	18	16	62	47
Female	248	53	29	15	6	16	20	59	38
Hispanic	79	46	26	20	17	21	17	57	53
White/other	312	57	31	18	12	18	20	62	42
Black	125	51	27	17	10	10	6	56	29
Age 18-25	305	53	24	14	6	13	13	59	38
25-34	137	65	45	30	34	21	29	66	62
35-45	79	48	28	18	3	32	21	59	35
Education									
Less than high school	51	41	56	22	24	30	7	58	60
HS grad	164	63	25	23	9	19	15	67	38
Some college	95	62	28	17	19	22	26	68	37
College grad.	210	50	24	14	6	11	20	54	45
Interview mode									
Personal	171	53	45	13	16	12	15	58	58
SAQ	172	61	24	23	17	19	16	65	27
Telephone	178	52	14	17	2	20	26	59	41

KEY: 1 = Pills refers to the nonmedical use of four classes of prescription drugs: stimulants, analgesics, tranquilizers, and sedatives.

TABLE 9. *Impact of parents' presence during the interview on self-reported use of any illicit drug: Logistic regression models for married/ cohabiting respondents age 18 to 45 coresiding with parents (N = 521; standard errors in parentheses; data are weighted).*

Independent variables	Dependent variable: ever used any illicit drug I	II
Female	-0.06 (0.20)	-0.05 (0.20)
White/other (omitted)	--	--
African American	-0.15 (0.28)	-0.11 (0.28)
Hispanic	0.03 (0.31)	0.03 (0.31)
Age	0.01 (0.02)	0.01 (0.02)
Education (years)	-0.05(0.05)	-0.05 (0.05)
Household income		
< $10,000 (omitted)	--	--
$10,000 - $29,999	-0.57 (0.55)	-0.62 (0.55)
$30,000 - $49,999	-0.17 (0.54)	-0.18 (0.54)
$50,000 or higher	0.21 (0.54)	0.20 (0.55)
Income missing	0.29 (0.57)	0.28 (0.58)
Work full-time (omitted)	--	--
Work part-time	-0.14 (0.26)	-0.09 (0.26)
Unemployed	0.37 (0.37)	0.42 (0.37)
Not in labor force	-1.16 (0.40)**	-1.13 (0.40)**
Interview mode		
Personal (omitted)	--	--
SAQ	0.06 (0.24)	0.34 (0.26)
Telephone	0.02 (0.24)	0.17 (0.26)
Parent present	-0.67 (0.27)*	0.09 (0.43)
Anyone else present	0.13 (0.26)	0.12 (0.26)
Significant interaction terms:		
SAQ x parent present		-1.64 (0.66)*
Telephone x parent present		-0.80 (0.66)
-2 log likelihood	617.33	610.82
Chi-square for model improvement		6.50*
Degrees of freedom		2

KEY: + = p < 0.10;* = p < 0.05; ** = p < 0.01; *** = p < 0.001 (2-tailed tests).

TABLE 10. *Impact of parents' presence during the interview on self-reported marijuana use: Logistic regression models for married/cohabiting respondents age 18 to 45 coresiding with parents (N = 521; standard errors in parentheses; data are weighted).*

Independent variables	Dependent variable: ever used marijuana or hashish I	II
Female	-0.04 (0.20)	-0.04 (0.20)
White/other (omitted)	--	--
African American	-0.13 (0.27)	-0.12 (0.28)
Hispanic	-0.26 (0.31)	-0.25 (0.31)
Age	0.01 (0.01)	0.01 (0.01)
Education (years)	-0.03 (0.05)	0.03 (0.05)
Household income		
< $10,000 (omitted)	--	--
$10,000 - $29,999	-0.32 (0.55)	-0.35 (0.55)
$30,000 - $49,999	-0.06 (0.54)	-0.06 (0.54)
$50,000 or higher	0.23 (0.54)	0.22 (0.54)
Income missing	0.27 (0.57)	0.27 (0.57)
Work full-time (omitted)	--	--
Work part-time	-0.21 (0.26)	-0.17 (0.26)
Unemployed	0.19 (0.35)	0.23 (0.36)
Not in labor force	-1.33 (0.42)**	-1.27 (0.42)**
Interview mode		
Personal (omitted)	--	--
SAQ	0.13 (0.24)	0.34 (0.26)
Telephone	-0.17 (0.24)	0.04 (0.25)
Parent present	-1.02 (0.28)***	-0.28 (0.42)
Anyone else present	-0.20 (0.26)	-0.19 (0.26)
Significant interaction terms		
SAQ x parent present		-1.26(0.66)+
Telephone x parent present		-1.46(0.77)+
-2 log likelihood	626.08	620.46
Chi-square for model improvement		5.63+
Degrees of freedom		2

KEY: + = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ (2-tailed tests).

TABLE 11. *Impact of parents' presence during the interview on self-reported cocaine use and the nonmedical use of prescription drugs: Logistic regression models for respondents age 18 to 45 coresiding with parents (N = 521; standard error in parentheses; data are weighted).*

Independent variables	Cocaine	Nonmedical use of prescription drugs
Female	-0.41 (0.27)	0.02 (0.26)
White/other (omitted)	--	--
African American	-0.05 (0.37)	-0.68 (0.43)
Hispanic	0.25 (0.38)	-0.00 (0.38)
Age	0.04 (0.02)*	0.07 (0.02)***
Education (years)	0.02 (0.07)	-0.08 (0.06)
Household income		
< $10,000 (omitted)	--	--
$10,000 - $29,999	0.96 (0.79)	-0.61 (0.72)
$30,000 - $49,999	0.32 (0.80)	-0.02 (0.68)
$50,000 or higher	0.62 (0.80)	-0.10 (0.69)
Income missing	0.19 (0.85)	-0.20 (0.72)
Work full-time (omitted)	--	--
Work part-time	-0.22 (0.37)	0.27 (0.34)
Unemployed	0.33 (0.43)	0.85 (0.43)*
Not in labor force	-0.80 (0.60)	-0.14 (0.54)
Interview mode		
Personal (omitted)	--	--
SAQ	0.61 (0.31)+	0.47 (0.33)
Telephone	0.15 (0.33)	0.75 (0.33)*
Parent present	-0.39 (0.39)	0.15 (0.35)
Anyone else present	-0.72 (0.39)+	-0.15 (0.34)
Significant interaction terms: none		
-2 log likelihood	413.77	411.51

NOTE: There is no model 2 for either of these variables because none of the tested interaction terms were significant in these models ($p > 0.10$).

KEY: + = $p < 0.10$;* = $p < 0.05$;** = $p < 0.01$;*** = $p < 0.001$ (2-tailed tests).

by mode of interview were found in two of the four models. The interaction terms did not support expectations about the ability of self-administered answer sheets to reduce response effects due to social desirability.

In the model for any illicit drug use (table 9), the negative effect of parent presence on drug use reports was significantly stronger in the SAQ mode than in the face-to-face interview (where respondents answered sensitive questions aloud). The pattern was similar for marijuana, but the interaction terms were not significant at conventional levels ($p < 0.10$). There were no significant presence-by-mode interactions in the models for cocaine or pill use. No firm conclusions about mode differentials in presence effects can be drawn, given the inconsistency of interaction results over the four models for parent presence.

Effects of Child Presence and Other Adult Presence. In the models for married and cohabiting respondents (tables 4 to 7), presence of children and presence of other adults during the interview were included as control variables. The presence of adults other than spouse or partner had a consistent negative impact on drug use self-reports in the models for married/cohabiting respondents, an effect in the opposite direction than spouse presence. The negative coefficients are marginally significant for cocaine ($p < 0.10$) and pill use ($p < 0.05$), but nonsignificant for marijuana and use of any illicit drug. These patterns suggest that the direction and magnitude of third-party effects on survey responses may depend on the identity of the person present. Replication with larger samples is needed to move beyond tentative findings.

The presence of children had no significant effects on any of the dependent measures in the married/cohabiting analyses and the regression coefficients were very close to zero in all models. These findings are similar to results reported by Bradburn and Sudman (1979), who found few consistent effects for child presence. Child presence was not tested in the models for parent presence because few of the respondents coresiding with parents also had coresident children of their own.

A somewhat different specification was used in the models for respondents coresiding with parents. In these models, the age of others present is not distinguished; instead, a dummy variable for the presence of anyone other than a parent was included. Siblings undoubtedly make

up the largest portion of the "others" present in this sample; nearly half of the respondents living with a parent had coresident siblings also, while only about 15 percent reported adults other than parent or sibling in the household. Only one marginally significant effect of others' presence was obtained: Self-reported cocaine use was lower ($p < 0.10$) if someone other than a parent were present during the interview.

DISCUSSION

The findings for spouse/partner presence and parent presence are generally consistent with the theoretical assumptions of this study. The magnitude and direction of third-party effects in surveys appear to be linked to the extent of the third party's knowledge of the information requested, and to the degree of personal stake the third party may have in the respondent's answers. The results summarized below should be taken as tentative and in need of replication in larger and more diverse samples.

Consistent with earlier research (Aquilino 1993) on married couples, the analyses suggest that willingness to report sensitive factual information may be increased by the presence of a spouse or partner during the interview. It is likely that in many relationships, the spouse or cohabiting partner knows of the respondent's past drug use, either because couples have used drugs together or have discussed details of past behavior. The effect is stronger among the younger married and cohabiting couples, especially those in their early twenties. These younger couples are more likely than older ones to be current or recent users of illicit drugs, and this may heighten sensitivity to the presence of someone who knows about this recent behavior. These findings support Mitchell's (1965) contention that it is more difficult to deny past behavior when someone who knows the truth is nearby.

In the married/cohabiting sample, the presence of children had no impact on response tendencies. The presence of other adults (related or unrelated), however, had a consistent negative effect on drug use reports, although the effects became borderline with control variables in the model. These findings suggest that spouse presence and other-adult presence have effects in the opposite direction: Spouse presence increases the probability of reporting illicit drug use, while other-adult presence lowers that probability. This pattern provides tentative support for the

proposition that the identity of the third party is critical in estimating effects related to lack of privacy.

The results for the parent coresident sample are also consistent with the theoretical approach. When someone without knowledge of the illicit behavior is present and that person has a personal stake in learning about such behavior, respondents will be less likely to reveal illegal or socially undesirable behavior. Parents fit both of these conditions in regard to their adult children's drug use; thus, their presence during the interview resulted in significantly lower reports of lifetime illicit drug use and marijuana use.

The subsample of respondents living with parents is predominantly in the 18 to 25 age range, ages at which parent-adult child coresidence tends to be very high. These results suggest that, in a household survey, interviewing young adults in the presence of their parents could introduce a downward bias in the drug use estimates for this group. These results are consistent with Gfroerer's (1985) research showing that younger adolescent respondents in the NHSDA were less willing to reveal illicit drug use when parents were present during the interview.

An unexpected finding of this research was the lack of association of privacy effects with interview mode. There was no significant linkage in the analysis of spouse or partner presence. In analyses of parent presence, interaction effects were inconsistent across models, and the two significant parameters were opposite to what was expected (stronger parent presence effects in the self-administered mode).

The findings for mode-by-presence interactions need to be interpreted very cautiously, however, due to the small sample sizes available to test these effects and the relatively large standard errors in the regression models. If accurate, the pattern of results would suggest that the use of self-administered answer sheets may not shield respondents from the potentially biasing effects of another person's presence during the interview. Having someone other than the interviewer nearby may alter the psychological setting of the interview, even if survey questions and answers are not spoken aloud. The self-administered interview may not feel completely private if someone with an interest in the responses is in the setting. It is also possible that third parties may intrude upon the self-administered interview directly by looking at the forms while they are being filled out, or by asking the respondent about questions and answers. Future research on this topic would benefit from having much

more detailed information on the precise manner in which other household members involved themselves in the interview, whether self- or interviewer-administered.

One primary weaknesses of this study is the crudeness of the measurement of the presence variables. Interviewers were not asked to record which specific parts of the interview were completed in private and which in the presence of others. Much more precise estimates of presence effects could be made if the presence variables could be linked to individual questions and if more detailed measures of the degree and manner of third-party intrusions were included in the data-collection protocol.

Implications for Survey Management

The results of this research suggest that interview setting influences response tendencies when information about sensitive or illegal behavior is sought. Third-party presence has the potential to alter the probability that respondents will admit illicit drug use. Although the effect may be positive (for spouse or partner presence) or negative (for parent presence and other adult presence), it is likely that variation in interview privacy within the sample increases measurement error. These results reinforce the need for interviewers in drug surveys to seek and maintain privacy during the interview. This is true regardless of survey mode, since privacy effects may be large even when the self-administered format is used. Interviewers should know, however, that the presence of a spouse or cohabiting partner during the interview appears to be less of a threat to the validity of self-reported drug use than is the presence of a parent or other adult. Interviewers need to be especially vigilant in preserving privacy when the respondent—whether an adolescent or young adult—coresides with parents. It might also be beneficial in such situations to have interviewers emphasize to respondents that parents will never see the answers recorded on the self-administered forms and that their answers will never be revealed any time after the interview. Respondents living in a parental home may need greater assurances of confidentiality than do those living in their own households. Researchers may need to make greater efforts to develop techniques and strategies for maximizing interview privacy when children of any age are interviewed in the parental home.

REFERENCES

Anderson, B.A., and Silver, B. The validity of survey responses: Insights from interviews of married couples in a survey of Soviet emigrants. *Soc Forces* 66:537-554, 1987.

Aquilino, W.S. Family structure and home leaving: A further specification of the relationship. *J Marriage Fam* 53:999-1010, 1991.

Aquilino, W.S. Telephone versus face-to-face interviewing for household drug use surveys. *Int J Addict* 27:71-91, 1992.

Aquilino, W.S. Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opin Q* 57:358-376, 1993.

Aquilino, W.S. Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opin Q* 58:210-240, 1994.

Bradburn, N.M., and Sudman, S. *Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*. San Francisco: Jossey-Bass, 1979.

Casterline, J., and Chidambaram, V.C. The presence of others during the interview and the reporting of contraceptive knowledge and use. In: Ross, J.A., and McNamara, R., eds. *Survey Analysis for the Guidance of Family Planning Programs*. Liege, Belgium: Ordina Editions, 1984. pp. 267-298.

Gfroerer, J.C. Underreporting of drug use by youths resulting from lack of privacy in household interviews. In: Rouse, B.; Kozel, N.; and Richards, L., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity*. National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985. pp. 22-30.

Gfroerer, J.C., and Hughes, A.L. The feasibility of collecting drug abuse data by telephone. *Public Health Rep* 106:384-393, 1991.

Johnson, T.P.; Hougland, J.; and Clayton, R. Obtaining reports of sensitive behavior: A comparison of telephone and face-to-face interviews. *Soc Sci Q* 70:174-183, 1989.

Malvin, J., and Moskowitz, J. Anonymous versus identifiable self-reports of adolescent drug attitudes, intentions, and use. *Public Opin Q* 47:557-566, 1983.

Mitchell, R. Survey materials collected in the developing countries: Sampling, measurement, and interviewing obstacles to intra- and international comparisons. *Int Soc Sci J* 17:665-685, 1965.

Silver, B.; Abramson, P.; and Anderson, B. The presence of others and overreporting of voting in American national elections. *Public Opin Q* 50:228-239, 1986.

Sudman, S., and Bradburn, N.M. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine Publishing Co., 1974.

Taietz, P. Conflicting group norms and the 'third' person in the interview. *Am J Soc* 68:97-104, 1962.

Turner, C.F.; Lessler, J.; and Devore, J. Effects of mode of administration and wording on reporting of drug use. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS Pub. No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 177-220.

Zanes, A., and Matsoukas, E. Different settings, different results? A comparison of home and school responses. *Public Opin Q* 43:550-557, 1979.

ACKNOWLEDGMENT

This chapter was prepared with support from National Institute on Drug Abuse research grant DA-06614.

AUTHOR

William S. Aquilino, Ph.D.
Associate Professor
Department of Child and Family Studies
University of Wisconsin
1300 Linden Drive
Madison, WI 53706

167357

The Use of the Psychological Laboratory To Study Sensitive Survey Topics

Gordon B. Willis

ABSTRACT

Maximizing the tendency of the survey respondent to answer truthfully when sensitive questions are presented is critical issue in survey methodology. A recent development devoted generally to the reduction of response error in survey data is the use of cognitive laboratory techniques during the survey development phase. The chapter categorizes and describes the various cognitive techniques that have been applied, by Federal agencies and other researchers, to the study of sensitive questions. Based on this analysis and review, a number of recommendations are made concerning specific aspects of survey design, when sensitive questions are administered.

INTRODUCTION

A vital practical problem addressed by survey researchers is the under-reporting, and response error in general, associated with asking questions that are sensitive because the answers are either embarrassing or admissions of illegal activities. A number of studies have addressed this problem through the use of experiments that manipulate, within a fielded survey, variables that are relevant to survey administration. This chapter focuses on an alternative, recently developed methodology—the use of the cognitive laboratory—to study the response to sensitive survey topics. Cognitive laboratory study is distinguished from field experimentation in that this research is carried out as an explicit psychological experiment in the laboratory, rather than under the guise of an actual survey; it is generally small in scale and qualitative in nature; and the aims of laboratory study are somewhat different from those of field experiments.

The results of field experiments have been inconsistent in determining the effects of particular survey administration variables, such as mode (self- or interviewer-administration) and the nature and extent of anonymity and

confidentiality provided to the respondent (Fowler 1993; Jobe et al., in press). The use of the psychological laboratory to study sensitive topics has been, in part, a response to these inconsistent findings. Under this approach, the cognitive aspects of the survey response are emphasized, and explicit attention is paid to the ways in which dominant response trends can be intensively studied through a variety of interviewing techniques that purport to study the thought processes of the tested individual (see Jobe and Herrmann 1994, for a discussion of the cognitive models that are proposed to underlie the survey response process).

Cognitive laboratory procedures generally use either focus groups or group discussions, or cognitive interviewing of individuals. In this chapter, all variants of these methods are termed "laboratory procedures," whether they are carried out in a laboratory or in some other location; the sole intent of the term is to distinguish these investigatory studies from experiments that are embedded in fielded surveys. Approaches that use laboratory procedures will be described in turn. Then, a number of preliminary hypotheses and conclusions arising from these studies will be reviewed.

COGNITIVE LABORATORY PROCEDURES

Focus Groups

Focus groups, in which a small number of individuals are studied in a group discussion format, have been useful early in the development of surveys involving sensitive topics (although see Schechter et al. 1993, who advocate the use of focus groups later in the questionnaire development process). Usually, the major intent of the focus group discussion is to determine the feasibility of a general approach, to develop hypotheses concerning the dominant variables that appear to influence survey responses, and to gather a range of reactions from prospective respondents that may guide survey development (see Krueger 1994 for a practical guide to conducting focus groups). For example, Tourangeau and colleagues (1992) used focus group discussion as part of a study on sensitive questions on women's health to highlight the major concerns that women appeared to have, relating both to the survey questions and the data-collection procedure used to ask the questions. Further, O'Brien (1993) reported on focus groups with gay and bisexual men that were designed to inform the development of survey questionnaires and procedures. Focus groups have the advantage of distributing the

discussion of the topic among a number of participants, so that embarrassing topics can be discussed without any particular person feeling as though he or she is being targeted. On the other hand, the results of focus group discussions are notoriously difficult to document, and a large amount of qualitative information must be reduced to a useful summary form.

Cognitive Interviewing

The use of cognitive interviewing of individual laboratory subjects emphasizes the adaptation of verbal probing techniques, as described by Ericsson and Simon (1980, 1984). As the survey research community increasingly emphasizes sources of response error that impact on overall survey data quality, these techniques have been used extensively to study the response process with respect to survey topics. A general review of these studies is contained in Jobe and Mingay (1991). In general, these studies rely on verbal reports by the laboratory subject that are in addition to the answers given to the survey questions themselves, and are elicited either through think-aloud methods (asking the subject to spontaneously "talk through" their thoughts), or by interviewer probing methods (the interviewer asks additional questions, beyond the written survey questions, to explore the respondent's thinking processes). The qualitative results of these investigations are normally used as a basis for modifying survey questionnaires so that these are in a form that eases completion of the question-answering task for the respondent, presumably resulting in an improvement in data quality.

Variations of verbal report methods have recently been used to study the reactions of individually tested laboratory subjects to a number of presented materials that are relevant to sensitive surveys (see Blair et al. 1992; Forsyth et al. 1992; Holland and Willis 1991; Willis et al. 1994). Two major types of approaches have been taken in these studies, each of which is discussed briefly below.

The Direct Questioning Method. This method closely parallels the approach often taken for testing nonsensitive survey questions: Subjects are asked to answer survey questions in the laboratory, much as they would be asked the questions in an actual survey, and then are administered further verbal probes so the interviewer can explore the basis for their answers. This procedure may be conducted with either subjects of known status with respect to sensitive behaviors (e.g., subjects from drug treatment clinics, such as discussed by Willis et al. 1994), or with

subjects of unknown status (such as recruitment of members of the general public, or of those deemed likely to have engaged in particular behaviors based on age or some other characteristic) (Tourangeau et al. 1992). This procedure allows a straightforward examination of the sensitive topics queried, but is possible only if subjects are willing to divulge and openly discuss sensitive behaviors.

The Indirect Questioning Method. Using this method, subjects are not asked sensitive questions directly, but are instead queried about their perceptions, opinions, or responses associated with a number of issues related to the administration of these questions, and in particular, question content and administration procedure (Willis et al. 1994). This method is more likely to be used when subjects cannot be expected to report in detail on their own behavior with respect to sensitive topics, especially where they may be reticent to do so because of embarrassment or mistrust.

Cognitive Processes Studied Through Cognitive Laboratory Procedures

The focus of both direct and indirect studies is on specific cognitive processes, including the three discussed below.

Comprehension Processes. Survey developers are often interested in respondent comprehension of key terms, phrases, explanations, procedures, and concepts used (e.g., street drugs, sexual partner). For example, in a project to develop questions about teenage sexual behavior, teenage laboratory subjects were asked to circle terms they did not understand (Willis 1991). In other studies conducted at the National Center for Health Statistics (NCHS), subjects paraphrased definitions of key terms used in sensitive questions, and drug users undergoing treatment were asked to provide the best terminology for particular questions (examples of the techniques used to probe subjects can be found in Willis 1994*a*). Further, Tourangeau and colleagues (1992) and Forsyth and colleagues (1992) used cognitive interviewing techniques to study subjects' comprehension of drug questions; Holland and Willis (1991) studied teenagers' comprehension of key terms used in a survey of risk behavior; and Blair and colleagues (1992) assessed understanding of terms used in a survey of human immunodeficiency virus (HIV) risk behaviors. Study of subject comprehension can assess not only the semantic meaning of a term, but its emotional and associative meanings as well.

For example, subjects can be asked to directly or indirectly rate or rank question sensitivity (Willis et al. 1994).

Comprehension can also be assessed with respect to the survey administration procedures used, rather than the survey questions themselves. In particular, cognitive researchers are interested in laboratory subjects' understanding of procedural mechanics, and of their associated perceptions of risks to the survey respondent. For example, Willis and colleagues (1994) have studied laboratory subjects' reactions to different forms of the randomized response technique (RRT), developed by Warner (1965, 1976) and by Horvitz and colleagues (1967), in which anonymity is maintained by the use of a randomizing device (usually a coin) to determine which of a pair of questions the respondent is to answer, and where only the respondent knows which question is being answered. Moriarty and Wiseman (1976) and Soeken and Macready (1982) also conducted psychological laboratory-based experiments with respect to understanding of RRT, and Miller (1984) conducted similar work in a study of the item count procedure for administering sensitive questions in which the respondent counts the number of behaviors that he or she has engaged in and reports only that number.

Recall Processes. In order to study the recall of sensitive information, the investigator may ask about subjects' judgments of confidence in their answers, or mechanisms that subjects report using to retrieve information. These practices are somewhat more difficult than asking about comprehension, because they almost invariably involve a request for the disclosure of sensitive information (i.e., direct questioning). However, for subjects who are willing to disclose such information, it may be valuable to carry out these activities. For example, Forsyth and colleagues (1992) probed the 12-month recall of laboratory subjects in the development of the National Household Survey on Drug Abuse (NHSDA); Keer and colleagues (1992) inquired about long-term effects of drug use; Tourangeau and colleagues (1992) asked about reproductive history; and Holland and Willis (1991) studied teens' recall of topics such as drinking behavior.

Decision Processes. These processes can be divided into two subcategories:

- *Determinants of truthful response.* Subjects can be asked about features of the survey administration context that would lead them to respond truthfully or untruthfully. Note that asking

about whether "you" would respond truthfully inevitably leads to self-disclosure, because the subject will have to then report *why* he or she would do so, and this necessarily involves a direct-questionnaire approach. For an indirect questioning approach, one can instead develop descriptions of hypothetical situations, including vignettes (Willis et al. 1994), so that the subject may make a judgment concerning how a described individual would answer a particular sensitive question.

- *Comprehensive/general study.* As a general category of cognitive decision processes studied, the researcher often investigates laboratory subjects' perceptions of the entire range of stimulus variables that may be relevant to the subjects' decisions to answer sensitive questions truthfully or not. This research emphasizes social psychological factors, including interviewer characteristics, as well as specific survey administration variables.

FINDINGS FROM COGNITIVE LABORATORY-BASED RESEARCH

The following general findings are relevant to survey practice when sensitive questions are asked. They are mainly based on a number of experiments done in the NCHS Questionnaire Design Research Laboratory by researchers under contract to NCHS or by other authors who have applied similar laboratory-based techniques. All findings are presented as hypotheses to be considered by survey methodologists rather than as established truths. Where possible, results from field experiments are compared to those obtained from the cognitive laboratory.

Interviewer Variables/Introduction of Survey

The Stated Justification for the Survey Is Critical. Tourangeau and colleagues (1992) reported that focus group respondents expressed a concern that the results of surveys actually be used to benefit someone, and these authors found that subjects were especially critical of the role of the Federal Government in collecting private information on reproductive behavior. Results of a cognitive interviewing study by Willis (1989) also suggest that the survey researcher should make clear the purpose of the study, or demonstrate that the information collected is

actually useful to members of the general public or the subpopulation surveyed.

Based on these preliminary findings, it may help to justify data collection in some cases by providing respondents with demonstration materials, such as newspaper articles, that describe results from an earlier, related survey. In addition, one might begin survey administration with a letter and a carefully worded brochure that succinctly outline the major uses of the survey. The use of such an approach is supported by suggestions of some authors that even the title of the survey can be important to respondents. For example, Caspar (1992) suggests that responses to the National Household Survey on Drug *Abuse* would be improved by renaming the National Household Survey of Drug *Use*; the title of the survey may communicate the nature of its intended uses.

Respondents May Focus Attention on the Interviewer Rather Than on Other Aspects of Survey Administration. Based on observations of laboratory subjects by Rasinski (1993) and by Willis and colleagues (1994), the usual survey-administration situation is complex and, from the respondent's point of view, contains a number of inherent uncertainties. Further, when sensitive questions are asked, the presented situation represents a classic case of decisionmaking under conditions of uncertainty, and under conditions of potential risk, as the respondent who reveals sensitive information may be placing him or herself in considerable jeopardy if the information is disclosed (Dawes 1988).

Under such conditions, survey administrators often implicitly assume that respondents will determine the credibility of the procedure by assessing the mechanisms used to administer the survey, such as information contained on confidentiality forms and the procedures used to physically protect response security. However, to the extent that laboratory subjects are typical of survey respondents, it appears they do not necessarily focus heavily on these sources of information, but instead make decisions concerning whether to respond truthfully based on a more limited number of known factors, and in particular on their assessment of the interviewer's characteristics. It may be that, quite often, respondents make an initial assessment of the interviewer that, in essence, asks: "Do I trust this person?"

This interpretation concurs with an analysis by Groves and Cialdini (1991), who distinguish between two basic styles of information processing related to decisionmaking: A deliberate, analytical, and

exhaustive consideration of all pertinent features relevant to the decision, and a shortcut heuristic that makes use of a single, highly diagnostic piece of information that has proven to be useful in making past decisions. Groves and Cialdini argue that under conditions in which an individual is required to make a relatively quick decision, with limited information, he or she will rely on the heuristic strategy. Although these authors mention other possible salient cues, it may be that the heuristic often used by survey respondents relates directly to their conclusions concerning the trustworthiness of the interviewer. Thus, the interviewer may often serve as a "salesperson" who is selling him- or herself, rather than the survey per se. Ramifications of this conclusion, especially with regard to interviewer training, will be considered later.

Survey Administration Procedure Variables

Survey Respondents View Administration Procedures Differently than do Survey Administrators. The key assumption of the cognitive approach to survey design is that an understanding of the respondent's viewpoint is vital when developing an optimal administration procedure. Rasinski (1993) took this approach in studying laboratory subjects' interpretations of a confidentiality form used for a questionnaire on abortion reporting, finding that understanding of the form varied widely, and that misinterpretations even included a case in which the subject thought that it meant that the interviewer was requesting identification from the respondent. Rasinski also found that participants often did not understand the term "randomly selected."

Willis (1989) also examined the comprehension of a standard confidentiality form, asking subjects to paraphrase its meaning, and found that it was extremely difficult to understand. A disturbing trend noted by both Rasinski and Willis is that many subjects believed that the information collected would be available to Government agencies in general, including the Internal Revenue Service and the Federal Bureau of Investigation. Finally, Singer and Miller (1992) presented different confidentiality forms to subjects and found that they preferred the most simple, clearly stated version (although their subjects were not particularly impressed by either form).

Further evidence supporting the notion that respondents may view the survey administration situation differently than do survey administrators consists of the finding by Willis and colleagues (1994) that in the laboratory, subjects did not differentiate survey procedures that differed

widely in objective levels of threat (or the level of protection as defined by the survey administrator). Rather, subjects focused on variables that survey researchers do not normally consider, such as the facial expressions of both interviewers and respondents. In one experiment that assessed subjects' comprehension of a complex survey administration procedure, Willis and colleagues (1994) demonstrated to subjects a variation of the randomized response procedure. It was found that subjects fell into discrete subgroups with respect to understanding this procedure, and that members of the group who did not understand statistical aspects of the procedure were very apt to report that they would select a strategy of lying, due to the fear that they were being subjected to a type of "shell game." Smith and colleagues (1974) conducted an earlier, more extensive psychological study, and reported results consistent with this notion.

The above illustrations suggest that potent factors affecting the response cannot always be predicted ahead of time, but should be the focus of debriefing or prior intensive pretesting. Note that even random-digit dialing, which appears to be an anonymous procedure from the point of view of the survey administrator (Fowler 1993; de Leeuw and van der Zouwen 1988), may appear somewhat different to the respondent who is left wondering, "How did they really get my telephone number?"

The Potential for Social Embarrassment Should Be Taken into Account When Devising Administration Procedures. The notion that drug users do not care about what a member of the data-collection establishment thinks of them appears to be invalid; subjects tested by NCHS researchers at a drug treatment clinic indicated that they would prefer not to talk openly about drug use, making comments such as "You don't want the interviewer to think you're a sleaze or something," "You want to keep it upbeat," or "You don't want to tell these things to an old lady" (Willis 1989). Therefore, in asking about sensitive topics such as drug use and sexual behavior, one might avoid direct, face-to-face questioning by a survey interviewer and rely instead on self-administered instruments.

This conclusion is consistent with results from several field-based experiments. For example, Schober and colleagues (1992) found greater reporting of cocaine and marijuana use with self-administration than with face-to-face interviewing. Turner and colleagues (1992), Gfroerer (1992), and Mensch and Kandel (1988) also have higher reports of drug use in a self-administered version than under face-to-face administration, and Waterton and Duffy (1984) found computer-based self-administration to

produce higher reports than did face-to-face (paper questionnaire) administration for alcohol consumption. Further, Jobe and colleagues (in press) found that for questions on number of sex partners in the last year, past 5 years, and lifetime, and for questions on sexually transmitted diseases and condom use, self-administration produced higher estimates than did interviewer administration.

On the other hand, there are also cases when it appears that the factor of social embarrassment may be more in the minds of the survey administrators than in those of respondents. In the initial phase of development of an NCHS survey of teenagers on sexual behavior, drug use, and other risk behaviors, high school health teachers were enlisted to demonstrate the draft questionnaire to their classes and to follow a protocol that prompted student opinions (Willis 1991). Similarly, Holland and Willis (1991) used focus groups to determine the major concerns that teens would have about answering sensitive questions on fighting, drug use, and sexual behavior. In both studies, it became clear that the teens did not feel that questions on sexual behavior or drug use were especially invasive or embarrassing. Rather, teens were mainly concerned about the possibility of their parents' discovering their answers. These results were interpreted as indicating that it is feasible to ask these sensitive questions of teens in a household survey, but that it is important to develop an administration procedure that ensures that parents would not be able to determine respondents' answers.

Technology Used for Administration Can Have an Impact on Responding to Sensitive Questions. Audio computer-based self-administration, or audio-CASI (Turner et al. 1992), has been reported to be a promising technology. The respondent listens through earphones to the microcomputer-based digital speech presentation of the survey questions, and responds directly using the computer keyboard. Preliminary tests of this technology, as reported by Lessler (1994) and by O'Reilly and colleagues (1994), suggest that this could be a useful method for collecting sensitive information because of the inherent privacy afforded by the procedure, and because literacy problems are for the most part circumvented.

Reservations have sometimes been expressed about computers with respect to sensitive topics: The computer may be perceived as too impersonal, or its use may make evident to respondents that the information will be stored in a form that makes it very easy to duplicate and distribute individual responses. However, in a very preliminary

laboratory study, there was no strong indication that computers are seen as objectionable (Willis 1989). This is clearly an unresolved issue, given that Jobe and colleagues (in press) have found in a field study that for interviews conducted outside of the household, using a computer may actually be intimidating.

Laboratory Subjects Express the Inherent Multifactor Complexity of the Impact of Administration Variables. When asked about their feelings concerning a particular set of survey administration circumstances, laboratory subjects are extremely likely to qualify their opinions by stating that "it depends," and then explaining exactly what they think a truthful response will depend on (Willis et al. 1994). In appears that, in effect, subjects are able to spontaneously recognize and articulate potential interaction effects between a large set of relevant variables. This trend is consistent with results of focus group interviews of survey interviewers reported by Groves and Cialdini (1991), who report that interviewers strongly endorse the practice of *tailoring*, or taking into account a number of key contextual variables, and adjusting their survey approach depending on perceived characteristics of the respondents (for example, dressing nicely in a wealthy neighborhood or dressing down in a poor one). It appears that both interviewers and laboratory subjects use implicit knowledge gained through social interaction to make inferences about cues and behaviors that likely will influence the behavior of others.

It is unclear whether attention to contextual features will result in standard procedures that will actually improve compliance with sensitive questions, but this may be an area worthy of intensive study. In particular, it may be possible, in interviewer training, to stress the types of situational factors that respondents are likely to attend to when sensitive questions are administered. Note that this point is consistent with several observations made earlier concerning the ways in which interviewer characteristics, explanations attempting to justify the survey, and other situational variables not normally anticipated by the survey administrators influence respondents' willingness to admit to sensitive behavior.

Laboratory-Based Procedures Can Be Used To Develop Survey Administration Procedures, but the Results Must Be Interpreted With Caution. The various concerns discussed above raise a critical issue with respect to the use of the laboratory study of sensitive topics as an analog to the field environment. Clearly, the practice of using laboratory subjects to select an optimal survey administration procedure is fraught with potential methodological problems because of the suspect validity of

subjects' hypothetical judgments of "What would I do when faced with a particular administration procedure?" Because of this limitation, cognitive laboratory results cannot be used to predict the absolute frequency of truthful responding. Blair and colleagues (1992) also report limitations due to the hypothetical nature of asking, "How would someone react?"

Therefore, laboratory researchers have attempted, at a minimum, to demonstrate different administration procedures to laboratory subjects, and to ask which one subjects would prefer, in order to obtain *relative* measures of preference for these procedures. It remains to be determined whether the ranking of preferences obtained in the cognitive laboratory bears a resemblance to findings from the field. Rasinski (1993) found that for survey questions on abortion, laboratory subjects preferred a self-administered procedure to face-to-face administration, and preferred the telephone the least, an ordering that also reflects the results of some field experiments cited previously. When asked why they prefer self-administration, subjects tend to give explanations that are consistent with the notion of the maintenance of a social distance between interviewer and respondent.

Respondent/Response Variables

Nonresponse Is Not an Adequate Measure of Data Quality. It is sometimes thought that a low nonresponse rate to particular questions indicates a fairly high degree of respondent compliance, and by extension, truthfulness. Laboratory research suggests that this conclusion may be erroneous, however. Laboratory subjects have frequently reported that, for "yes/no" questions asking about whether they have engaged in a particular behavior, they would prefer to lie than to refuse to respond to a question; refusal is seen as tantamount to admitting to having carried out the sensitive behavior (Willis 1989).

The Respondents' Socioeconomic Status and Degree of Prior Disclosure Appear To Be Important for Questions on Drug Use. Two laboratory experiments reported by Willis and colleagues (1994) found that drug clinic subjects believe that drug users are not of one type, but may represent, at the extreme, two subtypes: relatively high-status, recreational users whose drug use is mostly secret, and low-status, heavier users whose drug use has already been generally disclosed. The clinic subjects suggested that the first category of user will be extremely unlikely to reveal drug use in a survey, while the latter will be somewhat more

likely to do so because they have little to lose as a consequence of such an admission.

Some Decision Factors May Not Be Amenable to a Rationally Based Approach. Survey administration approaches that systematically vary justifications, administration procedures, and other elements make the implicit assumption that respondent behavior is rational, and is determined by a systematic integration of perceptions of potential risks, losses, and benefits related to the provision of truthful responses. This hypothesis has been investigated preliminarily in the laboratory by Sirken and colleagues (1991) and by Willis and colleagues (1994). These researchers concluded that to some extent, laboratory subjects do appear to behave according to a rational, probabilistic response mechanism that Willis and colleagues (1994) have labeled the "cognitive utility model."

However, laboratory-based studies also suggest that there are limits to the application of this approach. In particular, the tendency towards absolute denial of certain forms of behavior among some drug users may be relatively intransigent (Keer et al. 1992; Willis et al. 1994). This concept was operationalized by one subject tested by NCHS researchers at a drug treatment center, who suggested that drug users he knew would not respond by answering "yes" to a self-administered question on drug use even if the survey administration procedure involved burning the questionnaire immediately after marking the answer. Therefore, it may be that for a certain proportion of drug users, no administration procedure, and no amount of confidentiality or anonymity, will have a positive effect on the decision of whether to answer truthfully.

Survey Questionnaire Variables

Respondent Perceptions of Question Sensitivity Cannot Be Assumed Ahead of Time. It has already been noted that several studies found teenagers not to be embarrassed or offended by questions on sexual behavior and drug use. Further, in a laboratory-based study of intravenous (IV) drug users, Willis and colleagues (1994) found that, contrary to expectation, anal intercourse was not as sensitive as were behaviors such as needle sharing. Smith (1992) also has found in a field experiment that questions on number of sex partners represent a fairly complex pattern with respect to sensitivity; in general, having too few sex partners is sensitive to men, and too many is sensitive to women. Locander and colleagues (1976) made the point that is clearly represented by these findings: Researchers need to measure, rather than to assume,

level of question or topic sensitivity. If it is the case that the scaling of question sensitivity is somewhat the same in the field as in the laboratory situation, the laboratory interview can be a valuable tool for assessing this factor.

It Is Useful to Distinguish Between Sensitive Questions and Sensitive Answers. Groves and Cialdini (1991, p. 94) define sensitive topics as "those that the respondents believe would reveal socially undesirable attributes they possess." Based on the results of laboratory studies discussed in this chapter, this statement appears to be only partially correct. Reactions of laboratory subjects make clear that some survey questions are sensitive to respondents regardless of the true answer, so that sensitivity is not necessarily dependent on the disclosure of socially undesirable attributes. In particular, even people who have nothing to hide may find detailed questions on sexual behavior to be embarrassing and offensive.

On the other hand, some questions are neither embarrassing nor offensive to ask most individuals, and it is only for those who have engaged in the targeted behaviors that the responses are problematic; in other words, it is the (truthful) *answer*, and not the *question*, that is sensitive. Questions on illegal drug use generally do not seem particularly sensitive to people who have *not* used them, but may be very sensitive to those who have, and who have also taken steps to avoid disclosure of these behaviors.

The Generic Cognitive Processes of Comprehension and Recall Are a Significant Source of Response Error That Can Be Addressed in the Cognitive Laboratory. The primary intent of survey researchers has been to influence the respondent's *decision* processes when sensitive questions are administered so that truthful answers will be obtained. However, laboratory studies have made evident that the basic, requisite cognitive processes of comprehension and recall that underlie response to nonsensitive survey questions are also relevant when sensitive questions are asked.

For example, Blair and colleagues (1992), in a laboratory-based developmental study of the National Household Seroprevalence Survey, found that interpretation of the term "street drugs" is not uniform across IV drug users. Forsyth and colleagues (1992) tested comprehension of terms used in the NHSDA and discovered that many terms used in the questionnaire were vague, misunderstood, or open to multiple

interpretations, and could be effectively respecified (e.g., smokeless tobacco, drug use occasion, or the concept of a month). Further, Tourangeau and colleagues (1992) found that respondent comprehension of concepts such as illegal drugs and drug use during pregnancy was variable, and problematic for the design of survey questions. Finally, Keer and colleagues (1992) have suggested that recall of information on lifetime drug use can be very difficult for users, and significantly, they found that the attribution of cause and effect related to the effects of drug use on one's life is often impossible, especially when multiple drugs have been used.

These above examples suggest that, especially in cases when those who have carried out sensitive behaviors are willing to discuss them, one can use cognitive laboratory techniques to obtain useful information that is otherwise unavailable to the survey designer, and that leads directly to explicit modifications to questionnaire content.

The Use of Long Questionnaires Can Result in Respondent Fatigue and Boredom, Which in Turn Lead to Poor Quality Data. Development of an NCHS National Health Interview Survey drug use questionnaire by Keer and colleagues (1992) suggests that increasing the length of the survey interview may be a larger impediment to response quality when respondents are drug users than when they are nonusers because of the limited amount of time and effort that some drug users will devote to the survey response task. Such a finding does not signify question sensitivity, but rather the opposite (i.e., lack of interest), and suggests the value of using a relatively short questionnaire. This conclusion is supported by the work of Krosnick (1991), who proposes that survey respondents often "satisfice," or expend only the cognitive effort necessary to produce a response that is simply adequate (rather than accurate), especially when they are bored or fatigued.

Open-Ended Responses to Some Sensitive Questions May Reduce Biasing Effects. Blair and colleagues (1977) have suggested that it may be better to ask some sensitive questions in an open-ended format. In support of this conclusion, field experiments by Schwarz and colleagues have shown that the precise intervals given in a closed-ended quantitative behavioral frequency question can influence the response considerably (Schwarz and Bienias 1990; Schwarz et al. 1985). Cognitive laboratory research has also indicated that respondents use multiple sources of information, including the response options given to them, in shaping their responses to nontrivial survey questions. That is,

respondents will often develop response strategies that make use of all available information, including that contained in response categories (Willis and colleagues 1994*b*). Therefore, for a question such as, "In the past 30 days, how many times did you smoke marijuana?", one may elect not to provide response intervals, to avoid possible contamination effects produced by the response categories themselves.

Length of the Reference Period Can Influence Truthfulness of Response. Based on laboratory results, more truthful responding can be expected for lifetime reference periods than for the last 12 months (Willis and colleagues 1994). That finding is consistent with results reported by Gfroerer (1992) and by Harrell and colleagues (1986) from a record-check study, and lends credibility to the use of the laboratory to study these issues. A related point is that, according to reports given by drug users currently undergoing treatment, *past* users of a drug may be much more likely than are *current* users to admit to drug use over any specified time period (Keer et al. 1992). On the other hand, Jick (1982) found through a record-check study that memory for details concerning past drug use is poor, whereas recall for current and recent use is relatively good. Therefore, *willingness* to admit to drug use and *quality* of information reported may be inversely related.

SUMMARY AND CONCLUSIONS

Based on findings from cognitive laboratory-based studies of response processes when sensitive survey questions are administered, this chapter has presented several hypotheses. In assessing the usefulness of these hypotheses, two issues are clearly relevant: the extent to which they are found to be valid in a field survey context; and, if they are true, the prospects that exist for translating them into useful survey practice.

Validity of Conclusions Based on Laboratory Studies

It is very difficult to obtain record or other data that will directly validate the use of the laboratory procedures described in this chapter (or, for that matter, use of the results of field experiments). It is argued here that the value of these procedures generally, and of the specific conclusions presented above, will be obtained through a feedback process between field and laboratory study. The results of field studies will lead researchers to wonder why certain results were obtained, and to endeavor to study these more extensively in the cognitive laboratory. Laboratory testing will

produce hypotheses based on insights gained from intensive, small-scale qualitative study, and lead the researcher to test these ideas out in the field environment. If it is found that the results from field studies support the hypotheses developed in the laboratory environment, this will lend credence to the cognitive laboratory approach.

Applicability of Laboratory Findings

Clearly, the conclusions in this chapter differ with respect to their degree of specificity and, therefore, ease of application. Some suggestions are simply procedural guidelines; for example, the suggestion that intensive cognitive laboratory study be carried out to understand respondents' perceptions of a new survey administration procedure, and use of these findings to make (as now unspecified) changes in the procedure. This suggestion is admittedly limited. Although it describes a series of steps to be taken, it does not necessarily promise an improvement in survey design or data quality.

Other suggestions made here also serve mainly as guidelines to development, but specify outcomes in which researchers might have more confidence. In particular, for purposes of reducing error associated with the basic cognitive processes of comprehension and recall, the cognitive laboratory appears to be very useful; although no studies definitively demonstrate the usefulness of cognitive laboratory procedures in reducing response error, the effectiveness of these techniques with respect to improving the quality of questionnaire data in general has been supported by several survey methodologists (Belson 1981; de Maio et al. 1993; Dippo 1989; Wellens 1994; Willis et al. 1991).

Several suggestions made in this chapter for the survey collection of sensitive information are very specific, and directly applicable to survey designs that incorporate sensitive questions. Most important, it is suggested that, based on the results of laboratory testing, the use of face-to-face or telephone administration of sensitive questions, especially when these questions pertain to illegal behaviors such as drug use, may be inadvisable. Rather, the results point to the value of a survey procedure that provides a degree of legitimacy, while at the same time affording a degree of social distance. The development of the audio-CASI procedure for administering sensitive surveys therefore appears to be a promising one, although further study in both laboratory and field environments is required.

Second, following Blair and colleagues (1977), it is suggested that behavioral frequency questions should use open- rather than closed-ended question formats because this practice will avoid biasing effects due to the nature of specific response categories provided. Third, it is recommended that questionnaires on drug use be relatively short to avoid respondent fatigue and increasing response error over the course of the interview. It may also be advisable to limit the complexity of the concepts presented to respondents, especially where questions require them to provide a self-assessment of cause-and-effect relationships between drug use and deleterious life events.

Looking ahead to the possible further use of cognitive laboratory techniques to study sensitive issues, three points should be realized. First, these techniques are very new, and it is likely that further development and refinement will lead to a significantly enhanced understanding of the most effective means for administering sensitive survey questions. Second, as it becomes more necessary, and more acceptable, to administer sensitive questions in large-scale surveys, it is likely that survey researchers will adopt and apply cognitive techniques more extensively as a normal component of the survey development process. Finally, it is possible that, as the psychological focus of cognitive laboratory research related to sensitive survey questions becomes better developed and more widely known outside the survey research field, basic researchers in psychology, sociology, and other related fields will be motivated to both critique, and to contribute to, this promising approach.

REFERENCES

Belson, W.A. *The Design and Understanding of Survey Questions*. London: Gower, 1981.

Blair, J.; Binson, D.; and Murphy, P. Cognitive assessment of survey instruments and procedures for rare populations: IV drug users and the National Household Seroprevalence Survey. *Proceedings of the Section on Survey Methods Research, American Statistical Association,* 1992. pp. 770-775.

Blair, E.; Sudman, S.; Bradburn, N.M.; and Stocking, C. How to ask questions about drinking and sex: Response effects in measuring consumer behavior. *J Market Res* 14:316-321, 1977.

Caspar, R.A. A follow-up study of nonrespondents to the 1990 National Household Survey on Drug Abuse. *Proceedings of the Section on Survey Methods Research, American Statistical Association,* 1992. pp. 476-481.

Dawes, R.M. *Rational Choice in an Uncertain World.* San Diego: Harcourt Brace Jovanovich, 1988.

de Leeuw, E., and van der Zouwen, J. Data quality in telephone and face to face surveys: A comparative meta-analysis. In: Groves, R.; Biemer, P.; Lyberg, L.; Massey, J.; Nicholls, W.; and Waksberg, J., eds. *Telephone Survey Methodology.* New York: Wiley, 1988. pp. 283-299.

de Maio, T.; Mathiowetz, N.; Rothgeb, J.; Beach, M.E.; and Durant, S. "Protocol for Pretesting Demographic Surveys at the Census Bureau." Report prepared for the Bureau of the Census, Center for Survey Methods Research, 1993.

Dippo, C.S. The use of cognitive laboratory techniques for investigating memory retrieval errors in retrospective surveys. *Proceedings of the International Association of Survey Statisticians, International Statistical Institute.* Paris: International Statistical Institute, 1989. pp. 323-342.

Ericsson, K.A., and Simon, H.A. Verbal reports as data. *Psychol Rev* 87:215-250, 1980.

Ericsson, K.A., and Simon, H.A. *Verbal Reports as Data.* Cambridge, MA: MIT Press, 1984.

Forsyth, B.H.; Lessler, J.T.; and Hubbard, M.L. Cognitive evaluation of the questionnaire. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* DHHS Pub. No. (ADM) 92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off. 1992. pp. 13-52.

Fowler, F.J. *Survey Research Methods.* Newbury Park, CA: Sage, 1993.

Gfroerer, J. An overview of the National Household Survey on Drug Abuse and related methodological research. *Proceedings of the Section on Survey Methods Research, American Statistical Association,* 1992. pp. 464-469.

Groves, R.M., and Cialdini, R.B. Toward a useful theory of survey participation. *Proceedings of the Section on Survey Methods Research, American Statistical Association,* 1991. pp. 88-97.

Harrell, A.V.; Kapsak, K.A.; Cisin, I.H.; and Wirtz, P.W. "The Validity of Self-Reported Drug Use Data: The Accuracy of Responses on Confidential Self-Administered Answer Sheets." Report prepared for the National Institute on Drug Abuse under contract no. 271-85-8305, 1986.

Holland, L., and Willis, G. "Innovative Pretesting Methods for Surveys of Adolescent Populations." Paper presented at the National Field Directors Conference, San Diego, May 1991.

Horvitz, D.G.; Shah, B.V.; and Simmons, W.R. The unrelated question randomized response model. *Proceedings of the Social Statistics Section, American Statistical Association.* Washington, DC: American Statistical Association, 1967. pp. 65-72.

Jick, H. Recall error in interview studies of past drug use. *Am J Public Health* 72:405, 1982.

Jobe, J.B., and Herrmann, D. "Comparison of Models of survey Cognition and Models of Memory." Paper presented at the Third Practical Aspects of Memory Conference, College Park, MD, August 1994.

Jobe, J.B., and Mingay, D.J. Cognition and survey measurement: History and overview. *Appl Cogn Psychol* 5:175-192, 1991.

Jobe, J.B.; Pratt, W.F.; Tourangeau, R.; Mingay, D.J.; Miller, J.D.; Baldwin, A.K.; and Rasinski, K. Effects of interview mode on sensitive questions in a fertility survey. In: Lyberg, L.; Biemer, P.; Collins, E.; DeLeeuw, L.; Decker, C.; Dippo, C.; Schwarz, N.; and Trewin, D., eds. *Survey Measurement and Process Quality*, in press.

Keer, D.W.; Rowe, B.; Rice, S.C., Jr.; and Trunzo, D.B. Asking about drugs and drug dependency: Reports of drug use and the mode of questionnaire administration. *Proceedings of the Section on Survey Methods Research, American Statistical Association,* 1992. pp. 608-613.

Krosnick, J.A. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl Cogn Psychol* 5:213-236, 1991.

Krueger, R.A. *Focus Groups: A Practical Guide for Applied Research.* Thousand Oaks, CA: Sage, 1994.

Lessler, J.T. "Literacy Limitations and Solutions for Self-administered Questionnaires to Enhance Privacy." Paper presented at the meeting of the Council of Professional Associations on Federal Statistics, Bethesda, MD, May 1994.

Locander, W.; Sudman, S.; and Bradburn, N. An investigation of interview method, threat and response distortion. *J Am Stat Assoc* 71:269-275, 1976.

Mensch, B., and Kandel, D. Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects. *Public Opin Q* 52:100-124, 1988.

Miller, J.D. "A New Survey Technique for Studying Deviant Behavior." Ph.D. diss., George Washington University, 1984.

Moriarty, M., and Wiseman, F. On the choice of a randomization technique with the randomized response model. *Proceedings of the Social Statistics Section, American Statistical Association.* Washington, DC: American Statistical Association, 1976. pp. 624-626.

O'Brien, K. Using focus groups to develop health surveys: An example from research on social relationships and AIDS-preventive behavior. *Health Educ Q* 20:361-372, 1993.

O'Reilly, J.M.; Hubbard, M.L.; Lessler, J.T.; Biemer, P.P.; and Turner, C.F. Audio and video computer assisted self-interviewing: Preliminary tests of new technologies for data collection. *J Official Stat* 10:197-214, 1994.

Rasinski, K.A. "Phase I report: Laboratory Research on Responses to Sensitive Survey Questions." Report prepared under National Center for Health Statistics contract no. 200-92-7026. 1993.

Schechter, S.; Trunzo, D.; and Parsons, P. Utilizing focus groups in the final stages of questionnaire design. *Proceedings of the Section of Survey Research Methods, American Statistical Association*, 1993.

Schober, S.E.; Fe Caces, M.; Pergamit, M.R.; and Branden, L. Effects of mode of administration on reporting of drug use in the National Longitudinal Survey. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies.* DHHS Publication No. (ADM)92-1929. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1992. pp. 267-276.

Schwarz, N., and Bienias, J. What mediates the impact of response alternatives on frequency reports of mundane behaviors? *Appl Cogn Psychol* 4:61-72, 1990.

Schwarz, N.; Hippler, H.J.; Deutsch, B.; and Strack, F. Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opin Q* 49:388-395, 1985.

Singer, S.M., and Miller, M. "Confidentiality Assurances in Surveys: A Quantitative Review of the Experimental Literature." Unpublished report. Washington, DC: Bureau of the Census, Center for Survey Methods Research, 1992.

Sirken, M.G.; Willis, G.B.; and Nathan, G. Cognitive aspects of answering sensitive survey questions. *Proceedings of the International Association of Survey Statisticians*. Paris: International Statistical Institute, 1991. pp. 130-131.

Smith, T. Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. *Soc Biol* 39:203-211, 1992.

Smith, L.L.; Federer, W.T.; and Raghavarao, D. A comparison of three techniques for eliciting truthful answers to sensitive questions. *Proceedings of the Social Statistics Section, American Statistical Association*, 1974.

Soeken, K.L., and Macready, G.B. Respondents' perceived protection when using randomized response. *Psychol Bull* 92:487-489, 1982.

Tourangeau, R.; Rasinski, K.; and Lee, L. "Cognitive Research on Health Promotion and Disease Prevention: Women's Health." Report under National Center for Health Statistics contract no. 200-91-7099, 1992.

Turner, C.F.; Lessler, J.T.; and DeVore, J.W. Effects of mode of administration and wording on reporting of drug use. In: Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C., eds. *Survey Measurement of Drug Use: Methodological Studies*. DHHS pub. no. (ADM) 92-1929. Washington, DC: Supt. of Docs.,U.S. Govt. Print. Off., 1992. pp. 177-219.

Warner, S.L. Randomized response: A survey technique for eliminating evasive answer bias. *J Am Stat Assoc* 60:63-69, 1965.

Warner, S.L. Optimal randomized response models (with discussion). *Int Stat Rev* 44:205-212, 1976.

Waterton, J.J., and Duffy, J.C. A comparison of computer interviewing techniques and traditional methods in the collection of self-report alcohol consumption data in a field survey. *Int Stat Rev* 52:173-182, 1984.

Wellens, T. "The Cognitive Evaluation of the Nativity Questions for the Current Population Survey." Paper presented at the meeting of the American Association for Public Opinion Research, Danvers, MA, May 1994.

Willis, G.B. "Laboratory Experiments on Sensitive Questions." Unpublished manuscript. Hyattsville, MD: National Center for Health Statistics, Office of Research and Methodology, 1989.

Willis, G.B. "Development of the Youth Risk Behavior Survey in the Questionnaire Design Research Laboratory." Unpublished manuscript. Hyattsville, MD: National Center for Health Statistics, Office of Research and Methodology, 1991.

Willis, G.B. "Cognitive Interviewing and Questionnaire Design: A Training Manual." Working Paper Series, No. 7. Hyattsville, MD: National Center for Health Statistics, Cognitive Methods Staff, 1994*a*.

Willis, G.B. "The Use of Strategic Processes by Survey Respondents." Paper presented at the Third Practical Aspects of Memory Conference, College Park, MD, August 1994*b*.

Willis, G.B.; Royston, P.; and Bercini, D. The use of verbal report methods in the development and testing of survey questionnaires. *Appl Cogn Psychol* 5:251-267, 1991.

Willis, G.B.; Sirken, M.G.; and Nathan, G. "The Cognitive Aspects of Responses to Sensitive Survey Questions." Working Paper Series, No. 9. Hyattsville, MD: National Center for Health Statistics, Cognitive Methods Staff, 1994.

AUTHOR

Gordon B. Willis, Ph.D.
Survey Statistician
Office of Research and Methodology
National Center for Health Statistics
Room 915
6525 Belcrest Road
Hyattsville, MD 20782

Repeated Measures Estimation of Measurement Bias for Self-Reported Drug Use With Applications to the National Household Survey on Drug Abuse

Paul P. Biemer and Michael Witt

ABSTRACT

Direct estimates of response bias for self-reports of drug use in surveys require that essentially error free determinations of drug use be obtained for a subsample of survey respondents. The difficulty of obtaining determinations which are accurate enough for estimating validity is well-documented in the literature. Methods such as specimen (e.g., hair, urine) analysis, proxy reports, and the use of highly private and anonymous modes of interview all have to contend with error rates which may only be marginally lower than those of the parent survey. Thus, any methodology for direct validity estimation must rely to some extent on approximations and questionable assumptions.

In this chapter, the authors consider a number of methods that rely solely on repeated measures data to assess response bias. Since the assumptions associated with these approaches do not require highly accurate second determinations they may be more easily satisfied in practice. One such method for bias estimation for dichotomous variables that is considered in some detail provides estimates of misclassification probabilities in the initial measurement without requiring that the second measure be accurate or even better than the first. This methodology does require, however, that two subpopulations exist which have different rates of prevalence but whose probabilities of false positive and false negative error are the same.

The applicability of these methods for self-reported drug use are described and illustrated using data from the National Household Survey on Drug Abuse. In the discussion of the results, the importance of these

methods for assessing the validity of self-reported drug use are examined.

INTRODUCTION

The self-report is an integral component of the research methodology for measuring the prevalence of substance abuse and other stigmatized behaviors. While a growing body of literature supports the validity of the self-report, there are also studies that question its validity (see Mieczkowski (1991) for a review of validation research in this area). These studies suggest that response validity for drug use is highly dependent upon the construction of questions, procedures for administration, the investigator's perceived intentions, and the respondent's cognitive fitness. Given the importance of monitoring drug use prevalence, trends, and risk factors for the U.S. population, considerable research has been conducted to improve the validity of the self-report for sensitive topics; examples include the use of more private and/or anonymous reporting methods and attempting to motivate honest reporting by incentives or personal appeals (for example, see chapters by Lessler and O'Reilly, and Tourangeau, this volume).

To compare the accuracy of alternative data-collection methodologies for obtaining self-reports, some information on the reporting error associated with the measurement processes is required. If the objective of a methodological study is to estimate the magnitude of measurement bias, then error-free determinations of drug use are typically required for a sample of study subjects. For other methodological studies, it may only be necessary to obtain determinations that have better measurement error properties than the methodologies being evaluated. If no criterion data are available for estimating measurement bias, it is sometimes sufficient to know the direction of the reporting bias in order to select the best data-collection method. As an example, in many cases it is reasonable to assume that stigmatized behaviors will generally be underreported by the study population. In such cases, the data-collection methodology that produces the highest prevalence rate is deemed the most valid method (see Biemer 1988 for a critique of this approach).

As the preceding discussion affirms, measurement error evaluation methodology is critical for the improvement of the survey design and survey methods. In addition, evaluation methods are used to assess the components of total error in reported estimates from drug use prevalence

studies, and these data help define the limitations of the survey results for policy decisions and other uses of the data. Nonetheless, all methods for estimating validity, reliability, and response bias are themselves subject to questions of validity.

This chapter focuses on a number of methods for assessing the validity of self-reports of drug use. In particular, the discussion is confined to methods for estimating measurement bias that rely on repeated measurements of the same characteristics for the same individuals. Examples are reinterview methods, test-retest, record check studies, and biological test validation methods. In the next section, a number of meaurement error indicators and measures are reviewed that have been used in the literature to describe the measurement accuracy and precision of survey data. In the discussion of reliability and bias estimation, several approaches for estimating these measurement error indicators using repeated measurements methods are presented. Using a general, two-measurement model for measurement error, each estimation approach is seen as a design for restricting the parameter space of the overspecified general model by setting some parameters to zero and/or others to the same, unknown constant. These restrictions then impose requirements on the evaluation designs that must be met in order for the model assumptions to hold.

A substantial part of the chapter presents the results of an evaluation of a recently developed statistical method for estimating false positive and false negative reports from repeated measurement studies. The method was developed by Hui and Walter (1980) for the evaluation of medical diagnostic testing procedures. Sinclair and Gastwirth (1993) applied the method for the evaluation of survey measurements and extended the methodology in ways that enhance the method's applicability for survey evaluation. The method provides estimates of misclassification probabilities in the initial measurement without requiring that the second measurement be without error or even more accurate than the first measure. The methodology does require, however, that two domains can be defined that have different rates of prevalence but have identical probabilities of misclassification.

For this application, the method is applied to data from the National Household Survey on Drug Abuse (NHSDA) to estimate the misclassification errors associated with self-reports of alcohol, marijuana, and cocaine use. False negative and false positive probabilities (and their standard errors) by various demographic

subgroups and geographic areas are presented in the section discussing the application of the Hui-Walter method to NHSDA data.

Finally, the last section summarizes the results of the application of repeated measurement methods for the evaluation of self-reports of drug use and presents conclusions regarding their application to NHSDA data.

REVIEW OF MEASUREMENT ERROR TERMINOLOGY

In this section, several measurement error concepts relevant to the study of self-reported drug use are reviewed. The study is restricted to the error in a single dichotomous response variable, denoted by y, because this type of response is quite often encountered in drug use measurement. As an example, y may denote a "yes or no" response regarding the use of specific drugs during some period or it may denote a response to a category of use in a multiple category response set. Let y denote the measurement for some characteristic associated with the i-th survey respondent where $y_i = 1$ if the respondent possesses the characteristic and $y_i = 0$ if otherwise. Let μ_i denote the corresponding true value for the respondent. Following Cochran (1968), the following misclassification probabilities are defined:

$$\theta = P(y_i = 0 \mid \mu_i = 1)$$

$$\phi = P(y_i = 1 \mid \mu_i = 0) \qquad (1)$$

where θ and ϕ are referred to as the probability of a false negative and the probability of a false positive, respectively. Thus, the expected value of y_i given μ_i is

$$E(y_i \mid \mu_i) = \mu_i(1-\theta) + (1-\mu_i)\phi. \qquad (2)$$

Measurement Bias

Let $\pi = E(\mu_i)$, the true prevalence of the characteristic in the target population; let $P = E(y_i)$, the expected observed prevalence. Let the measurement bias of the measure y be defined as $B(y) = P - \pi$. Thus, from (2)

$$B(y) = -\pi\theta + (1-\pi)\phi \qquad (3)$$

From (3), it can be seen that the measurement bias is 0 or small relative to π if either:

$$(a)\ \theta \approx \phi \approx 0$$

$$(b)\ \pi\theta \approx (1-\pi)\phi$$

Condition (a) implies that there is almost no chance for a misclassification error. Condition (b) implies that the expected number of false positive errors in the population approximately equals the expected number of false negative errors. As Cochran (1968) points out, this latter condition is quite unlikely in most applications, so that zero measurement bias is usually an indication that condition (a) holds.

Note that for drugs with low prevalence rates such as cocaine and heroin, π will be many times smaller than $(1 - \pi)$, and a relatively small false positive rate can have large consequences on the bias. As an example, suppose $\pi = 0.010$, $\theta = 0.30$, and $\phi = 0.010$. Then, $\pi\theta = (0.010)(0.30) = 0.0030$ while $(1 - \pi)\phi = 0.99 \times 0.010 = 0.0099$. Thus, the contribution to bias due to false positives is 3.30 times larger than the contribution due to false negatives, although the probability of a false negative is 30 times greater than the probability of a false positive. Using equation (3), $B(y) = 0.0030 + 0.0099 = 0.0069$, the relative bias, defined by $RB(y) = B(y)/\pi$, is $0.0069/0.010 = 0.69$; that is, estimates of π based on y will be 69 percent larger, on average, than π. Thus, for rare drugs, the consequences of even a small false positive rate can be substantial.

Measurement bias is important in survey work because it is directly related to the bias in estimators of means, proportions, and totals. Let $p = \Sigma y/n$ denote the sample proportion for a simple random sample. Then the bias in p for estimating π, the true population proportion, is defined as $E(p-\pi)$ which is also given by (3).

Reliability

Roughly speaking, reliability refers to the degree of consistency of responses from independent, replicated measurements of the same characteristic. The statistical definition of the reliability ratio, R, is the proportion of the variance that is *not* measurement variance, where *measurement variance* is defined as $E[\mathrm{Var}(y_i \mid i)]$; that is, the average

variance within respondents and between hypothetical, independently replicated measurements. Thus, R can be expressed mathematically as

$$R = 1 - \frac{E[\mathrm{Var}(y_i \mid i)]}{\mathrm{Var}(y_i)} = \frac{\mathrm{Var}[E(y_i \mid i)]}{\mathrm{Var}(y_i)} \tag{4}$$

Biemer and Stokes (1991) show that, for dichotomous responses, R can be quite difficult to interpret because it is a complex function of the misclassification probabilities and π. They show that under the model in (2)

$$R = 1 - \frac{\pi\theta(1-\theta) + (1-\pi)\phi(1-\phi)}{PQ} \tag{5}$$

where $P = \pi(1-\theta)+(1-\pi)\phi$ and $Q = 1 - P$. Further, they show that (a) for two domains or subpopulations having identical probabilities of misclassification, the reliability ratio for one domain can be substantially larger than the ratio for the other solely as a consequence of the difference in their respective prevalence rates. As an example, suppose that for domain 1, $\pi = 0.50$ while for domain 2, $\pi = 0.10$. Further, let $\theta = 0.00$ and $\phi = .10$ for both domains. Then, using equation (5), $R = 0.82$ for domain 1 and $R = 0.47$ for domain 2. On this basis, it would be wrong to conclude that the responses from domain 1 are of higher quality than are those from domain 2. Thus, in this respect, R can be misleading as an indicator of data quality. (b) From equation (5), it can be shown that the reliability ratio can be very high although there is a large amount of misclassification error. As an example, suppose the false positive probability is zero ($\phi = 0$) while the false negative rate is high, say 10 percent ($\theta = 0.10$). Further suppose that $\pi = 0.050$. This situation is often encountered in drug use measurement for rarely used drugs. Then it can be shown that $R = 0.90$, suggesting very high reliability in the measure. Further, the relative bias in the measure is -10 percent, which is nontrivial.

While good reliability is not necessarily an indicator of good data quality, poor reliability usually indicates that the measure is subject to a large measurement bias. This is especially true when the prevalence rate is small, as with cocaine or heroin use. As an example, consider the case where $\theta = 0.10$, $\phi = 0.025$, and $\pi = 0.050$. Here, $1 - R = 0.43$, and the relative bias of the measure defined above is $RB(y) = 0.38$ or 38 percent. This correspondence between $I = 1 - R$, called the index of inconsistency, and the relative bias for small prevalence rates is not coincidental. By comparing (5) to (3) divided by π, it can be verified that I and $RB(y)$ will be close whenever π is small. Further, when π is small, the cause of poor reliability is a high and disproportionate number of false positives compared with the number of false negatives in the population. To illustrate, in the example, the expected number of false negatives in the population is $N \times 0.050 \times 0.10 = 0.0050N$, where N is the population size. This compares with $N \times 0.95 \times 0.025 = 0.024N$ false positives—approximately 5 times as many false positives as false negatives. Thus, one may conclude that when the prevalence of the characteristic is small, poor reliability is usually an indication of a large positive bias in the estimator of the prevalence rate. By a similar argument, one can conclude that when the prevalence rate is large (say, $\pi > 0.90$), poor reliability is usually an indication of a large negative bias in the estimator due to a high false negative rate. For π between 0.10 and 0.90, poor reliability is an indication of a large expected number of false positives and/or false negatives. However, little can be said regarding the direction of the bias or whether the net effect of misclassification error results in either a small or large bias in the estimator of the prevalence rate.

To summarize, this discussion shows that in some situations the reliability ratio can be a good indicator of measurement and estimator bias. Further, a large value for the estimator of R is no assurance of good data accuracy. A low value of R is an indication of large misclassification errors in the data. Finally, in some situations, R can help researchers determine whether the misclassification error problem is a result of high false negative and/or high false positive probabilities.

Validity

Bohrnstedt (1983, p. 97) states that validity is an indicator of "the degree to which an instrument measures the construct under investigation." Bohrnstedt discusses a number of alternative concepts of validity proposed in the psychometric literature for describing data quality. Some of these are predictive validity, concurrent validity, empirical validity, and

theoretical validity. These concepts and others are discussed in some detail in Groves (1989). Of particular relevance to the present discussion is theoretical validity (*TV*) which, in terms of the model, is defined as the correlation between the observed measure and the conditional expectation of the observed measure, called the true score. Thus,

$$TV = \text{Corr}[\text{E}(y_i|\mu_i), \mu_i] \qquad (6)$$

As with most concepts, theoretical validity is defined as a correlation between two constructs (i.e., measures or true scores). Because validity does not depend upon the existence of a true value, it is the preferred indicator for describing the quality of measures of psychological states, attitudes, or knowledge. Biemer and Stokes (1991) show that, under the error model proposed above for dichotomous data, $(TV)^2 = R$, the reliability ratio. They further show that under more general models, $(TV)^2 \leq R$. Thus, reliability is an upper bound on theoretical validity; consequently, a measure may be reliable but lack validity. This is similar to the result shown for measurement bias: A measure may be reliable but still be substantially biased. This result further implies that an unreliable measure cannot be valid. Note, however, that an unreliable measure may still be unbiased. For the classification error models considered here, reliability and validity, while conceptually different, are mathematically equivalent indicators. Thus, as a measure of data quality for categorical variables, the limitations of the reliability ratio are also limitations of validity measures.

It is not uncommon to find the terms "validity" and "measurement bias" used synonymously. It is important to note that these concepts are quite different. As an example, if some positive number, *C*, is added to every measurement, the validity of the measure is unchanged while bias is increased by *C*. The advantage of validity as an indicator of data quality is that, unlike measurement bias, validity does not require that true values exist for the constructs under study.

Mean Square Error

Whereas measurement bias, reliability, and validity are defined at the response level, the mean square error (MSE) is defined at the estimator level. The mean square error of an estimator is a measure of accuracy

that is often used for estimators of population parameters. Let π be any estimator of π, then

$$\text{MSE}(\hat{\pi}) = \text{Bias}(\hat{\pi})^2 + \text{Var}(\hat{\pi}) \tag{7}$$

the sum of the square of the bias and the variance. Suppose that π is the simple expansion mean under simple random sampling, denoted by p. As mentioned above, the bias in p is $B(y)$ defined in (3). Biemer and Stokes (1991) show that for small samples from large populations,

$$\text{Var}\ (\bar{y}) = PQ/n \tag{8}$$

where P was defined before as $E(y_i)$. An unbiased estimator of the variance is the usual estimator,

$$v(\bar{y}) = pq/n \tag{9}$$

where $q = 1 - p$. Thus, under the assumed model, the usual variance estimator is unbiased in the presence of measurement error. It will be shown subsequently that this is not true under more general models.

These variance formulas show that misclassification error can sometimes result in smaller variances for estimators of proportions and totals. Consider the situation where $\pi = 0.5$. In this situation, the variance of the sample proportion is at its maximum. Thus, misclassification can only reduce the variance. One exception to this is when the misclassification errors are correlated, as happens with interviewer error. If interviewers exert influence over the misclassification error for respondents in their assignments, then misclassification errors are correlated and equation (7), which was derived under the assumption of unit-to-unit independent misclassification error, no longer holds. Under a more appropriate model for this situation, misclassification error always results in an increase in estimator variance. Further, the usual estimators of variance may be substantially biased. Biemer and Stokes (1991) discuss models that are appropriate for the study of interviewer errors and dichotomous response variables.

ESTIMATION OF RELIABILITY AND BIAS

This section considers methods for estimating the components of error for the dichotomous measurement error model. These methods are test-retest, true value measurements, and repeated measurements. The assumptions underlying these methods will be discussed in terms of a general model for two measurements. Models for multiple measurements are essentially extensions of this basic model.

Let y_{ti} denote the t-th measurement on unit i for $t = 1,2$ and $i = 1,...,n$. In analogy to the single measurement model, the assumptions are as below.

General model assumptions:

(i) $P(y_{ti}=0|\mu_i=1) = \theta_t,\ t=1,2$

(ii) $P(y_{ti}=1|\mu_i=0) = \phi_t,\ t=1,2$

(iii) $P(y_{2i}=0|y_{1i}=1,\mu_i=1) = \theta_{0|1}$,

(iv) $P(y_{2i}=0|y_{1i}=0,\mu_i=1) = \theta_{0|0}$,

(v) $P(y_{2i}=1|y_{1i}=0,\mu_i=0) = \phi_{1|0}$,

(vi) $P(y_{2i}=1|y_{1i}=1,\mu_i=0) = \phi_{1|1}$

Note that $\theta_{0|1} = \theta_{1|1} = \theta_2$ and $\phi_{1|0} = \phi_{0|0} = \phi_2$ if and only if the false negative errors and the false positive errors, respectively, corresponding to measurements 1 and 2 are independent. Under the general model, the probabilities of misclassification may differ between trials (assumptions i and ii) and further, the second trial outcomes are not independent of the first trial outcomes (assumptions iii - vi). Including π, seven parameters are associated with this model. However, only 3 degrees of freedom are available for estimation for a dichotomous variable with two measurements. Thus, as will be shown subsequently, additional assumptions are needed to estimate any of the parameters.

Test-Retest Methods

As discussed in the previous section, although the interpretation of the reliability ratio is difficult in the dichotomous case, estimates of reliability contain some information on bias that can be useful in studies of the accuracy of drug use measurement. The most common method of estimating reliability for self-reports is the test-retest method. This method includes reinterview studies as well as surveys in which replicate measures are embedded in a single interview and also reinterview studies. In reinterview studies, a subsample of the original respondents is recontacted for the purpose of obtaining a second set of measurements for the original interview characteristics.

Let $t = 1$ denote the first measurement and let $t = 2$ denote the second measurement or reinterview response. For the test-retest measurement model, the assumptions of the general model are replaced as follows.

Test-retest assumptions:

(i) Independence

$$\theta_{0|1} = \theta_{0|0} = \theta_2$$

$$\phi_{1|0} = \phi_{1|1} = \phi_2$$

(ii) Homogeneity

$$\theta_1 = \theta_2$$

$$\phi_1 = \phi_2$$

Assumption (i), (independence), which replaces assumptions (iii) to (vi) in the general model, essentially states that the errors in the two measurements are independent. That is, whether a false positive or false negative error is made for the second measurement does not depend upon whether an error was committed for the first measurement. For embedded replication there is a risk that respondents may simply repeat the erroneous response made on the first on the second measurement. When the second measurement is collected after some time has elapsed since the first measurement, this is less of a risk. Yet, as several researchers have shown, correlated errors can persist even when the

reinterview is conducted weeks after the initial interview (O'Muircheartaigh 1991; Bailar 1968).

Assumption (ii), (homogeneity), which replaces assumptions (i) and (ii) in the general model, states that the false positive and false negative probabilities are the same for both measurements. Thus, the aim of the design of the second measurement is to replicate the first measurement by, for example, using identical procedures, questions, or interviewer competencies. For reinterview surveys where the second measurement is obtained in a separate interview with the respondent, the reinterview design should replicate, to the extent possible, the same essential survey conditions that existed in the first interview. For replicate measures embedded with the same instrument, this assumption is more easily satisfied. Despite the potential difficulties with the test-retest assumptions, the method remains the most commonly used technique in survey methodology for estimating reliability.

To define an estimator of the reliability ratio, R, for dichotomous data, let a, b, c, and d denote the cell counts for the 2 x 2 measurement cross-classification table, as follows:

		y_{1i}		
		1	0	
	1	a	c	
y_{2i}	0	b	d	
				n

FIGURE 1. *Measurement 1 by measurement 2 cross-classification.*

$$p_1 = \frac{a+b}{n}, \; p_2 = \frac{a+c}{n}, \text{ and } n = a+b+c+d$$

Then, an estimator of R is $\hat{R} = 1 - I$ where

$$\hat{I} = \frac{(b+c)/n}{p_1 q_1 + p_2 q_2}$$

where $q_t = 1 - p_t$, $t = 1,2$ and $\hat{I}$ is an estimator of the index of inconsistency. These estimators assume that respondents are sampled using a simple random sampling design; however, for more complex sampling designs, weighted cell counts are typically used to estimate R.

It has been shown (Bureau of the Census 1984) that when the test-retest assumptions are not satisfied, the estimates of R can be substantially biased. Violations of assumption *(i)* are usually due to errors that are positively correlated between trials. In this situation, $\hat{R}$ is an overestimate of R. In the Bureau of the Census work (1984), it is shown that the bias in R is approximately $\rho_T I$ where ρ_T is the between-trial correlation. As an example, if $R = 0.70$ and $\rho_T = 0.20$, then the bias in $\hat{R}$ is approximately $0.20(0.30) = 0.06$ and, thus, $\hat{R}$ overestimates R by approximately 6 percent. When assumption *(ii)* is violated, $\hat{R}$ estimates a complex function of the reliabilities associated with each trial. Thus, interpretations of R based upon the above model can be misleading in these situations.

True Value Measurement Methods

To estimate measurement bias and the misclassification probabilities for self-reports, the traditional methodology has relied upon true value measurements. For drug use measurement, true values have been obtained from:

- administrative records, such as arrest records and drug treatment reports;
- hair, urine, and other specimen analyses to detect the presence of drugs in the specimens; and
- reinterviews using better methods than were used in the first interview, such as more private modes of interview, neutral (out-of-home) settings, and better question design.

With any of these methods, the usual modeling approach is to assume the following:

True value assumptions:

$$\theta_{0|1} = \theta_{0|0} = \theta_2 \equiv 0$$

$$\phi_{1|0} = \phi_{1|1} = \phi_2 \equiv 0$$

That is, it is assumed that the second measurement is the true value, or mathematically, $y_{2i} = \mu_i$. Thus, an estimator of the bias in the measurement y_1, assuming simple random samples is

$$\hat{B}(p_1) = p_1 - p_2 \tag{10}$$

If y_2 in figure 1 now denotes the true value, and using the notation for the cell counts in that table, the estimates of the false negative and false positive probabilities are respectively,

$$\hat{\theta}_1 = \frac{c}{a+c}$$

and

$$\hat{\phi}_1 = \frac{b}{b+d}$$

As before, weighted counts may be used for unequally weighted samples.

Occasionally, the assumptions for the true value model hold only approximately and a more appropriate set of assumptions is:

Improved measurement assumptions:

(i) Independence

$$\theta_{0|1} = \theta_{0|0} = \theta_2$$

$$\phi_{1|0} = \phi_{1|1} = \phi_2$$

(ii) Improved second measurement

$$\theta_2 < \theta_1 \text{ and } \phi_2 < \phi_1$$

In words, it is assumed that the second measurement is not free of error, but that the probability of error in the second measurement is smaller than that for the first measurement. Furthermore, the errors in both

measurements are independent. Under these assumptions, it can be shown that if $\hat{B}(p_1)$ and $\hat{B}(p_2)$ have the same sign,

$$|\mathrm{E}[\hat{B}(p_1)]| < |B(p_1)|$$

where $\hat{B}(p_1)$ is given by (10). Thus, the usual estimator of bias is biased downward. However, if $B(p_2) \approx 0$, then $\hat{B}(p_1)$ may still provide a useful approximation for $B(p_1)$.

It should be noted that, under the improved measurement assumptions, the estimators $\hat{\theta}_1$ and $\hat{\phi}_1$, given above for the true value model, are both biased and the directions of the biases are unknown. However, in this situation the estimation method discussed in the next section can be used to estimate the misclassification probabilities associated with both the first and second measurements.

Repeated Measurements: The Hui-Walter Method

In some studies, two or more measurements of μ_i are available for a sample of respondents; however, the assumptions made for test-retest and true value models are not tenable. For example, the second measurement is not perfect, nor even better than the first measurement. Neither is it plausible to assume that the second measurement is a replication of the first. Hui and Walter (1980) consider this situation in the evaluation of diagnostic tests. In this situation, the presence or absence of a disease may be indicated by two tests, each having probabilities of misclassification that are nonzero, nontrivial, and procedure dependent. Sinclair and Gastwirth (1993) applied the Hui-Walter estimation methodology for estimating the measurement error in self-reports in the evaluation of labor force characteristics in the Current Population Survey (CPS). Here the method is considered for the estimation the false positive and false negative probabilities for self-reported drug use.

Consider the case where two measurements are taken from each individual in two subpopulations or domains indexed by g. For each domain g, let A_g, B_g, C_g, and D_g denote the four cells in figure 1 as follows: A_g = cell (1,1), B_g = cell (1,0), C_g = cell (0,1), and D_g = cell (0,0).

Then the probability that a randomly selected individual from domain g is classified in each cell is as follows:

$$P(A_g) = \pi_g(1-\theta_{g,1})(1-\theta_{g,0|1}) + (1-\pi_g)\phi_{g,1}\phi_{g,1|1}$$
$$P(B_g) = \pi_g(1-\theta_{g,1})\theta_{g,0|1} + (1-\pi_g)\phi_{g,1}(1-\phi_{g,1|1})$$
$$P(C_g) = \pi_g\theta_{g,1}(1-\theta_{g,0|0}) + (1-\pi_g)(1-\phi_{g,1})\phi_{g,1|0}$$
$$P(D_g) = \pi_g\theta_{g,1}\theta_{g,0|0} + (1-\pi_g)(1-\phi_{g,1})(1-\phi_{g,1|0})$$

Assuming independence in the classifications between the two domains, the probability of observing $a_g, b_g, c_g,$ and d_g for $g = 1,2$ is therefore

$$l = \prod_{g=1}^{2} P(A_g)^{a_g} P(B_g)^{b_g} P(C_g)^{c_g} P(D_g)^{d_g}$$

This likelihood function contains 14 parameters and only (2 x 3 =) 6 degrees of freedom for estimation. To reduce the number of parameters, Hui-Walter and Sinclair-Gastwith assume the following.

Hui-Walter independence assumptions:

(i) Independence

$$\theta_{g,0|1} = \theta_{g,0|0} = \theta_{g,2}, \quad g = 1,2$$
$$\phi_{g,1|0} = \phi_{g,1|1} = \phi_{g,2}, \quad g = 1,2$$

(ii) Homogeneity between domains:

$$\theta_{1,t} = \theta_{2,t} = \theta_t, \quad t = 1,2$$
$$\phi_{1,t} = \phi_{2,t} = \phi_t, \quad t = 1,2$$

In words, this assumption says that:

- Misclassification probabilities differ between the two measurements, but are the same for both domains ($g = 1,2$),
- The prevalence rates differ between domains, and

- Misclassification errors are independent between trials.

These assumptions reduce the number of parameters to six, viz., θ_1, θ_2, ϕ_1, ϕ_2, π_1, and π_2. A solution for this formulation can be obtained using maximum likelihood estimation. This model will be referred to as the Hui-Walter independence model.

The assumption of equal error rates across domains is easily justified for many diagnostic tests of the types discussed by Hui and Walter (1980). Their example considers two tests for the detection of tuberculosis that exhibit the same error distributions across socioeconomic subgroups. In the survey setting, the misclassification errors may be highly correlated with the prevalence rates. Therefore, it is important to choose the two domains carefully to ensure proper application of these methods.

For their application to the CPS, Sinclair and Gastwirth (1993) define the two domains based on race and gender: white males and white females. Thus, it is not necessary that the two domains partition the entire population. Although the results of their study only apply to these two domains, important insights may be gleaned for the entire population by studying this part of it. Sinclair and Gastwirth demonstrate the importance of defining the two domains such that their respective prevalence rates for the characteristic of interest are markedly different. Because the characteristic of interest in their study was labor force participation, their choice of race and gender would seem appropriate, as labor force participation rates are considerably higher for white males ($\pi_1 = 0.75$) than for white females ($\pi_2 = 0.55$). Further, the assumption of equal error probabilities for the two domains is also plausible: Each domain is administered the same questions by the same interviewers using the same survey procedures. However, the assumption of independence between the errors for the two trials may not be justified. O'Muircheartaigh (1991) estimates that the between-trial correlation for labor force participation varies in the interval [0.3, 0.5] when the second measurement is obtained using a replicate reinterview survey. Sinclair and Gastwirth consider the effects of between-trial correlations on the resulting estimates and conclude that failure of this assumption to hold can result in large biases in the estimates of the error probabilities.

In this application to self-reported drug use, the estimates using the Hui-Walter independence model as well as a dependent model are compared and evaluated. The latter model is similar to the one proposed by Vacek (1985); however, it uses fewer parameters and therefore requires fewer

degrees of freedom to estimate. For the dependent model, the following is assumed:

Dependent model assumptions:

(i) Homogeneous false negative probabilities

$$\theta_{g,1} = \theta_1, \quad g = 1,2$$

$$\theta_{g,0|1} = \theta_{0|1}, \quad g = 1,2$$

$$\theta_{g,0|0} = \theta_{0|0}, \quad g = 1,2$$

(ii) Independent and homogeneous false positive probabilities

$$\phi_{g,1|0} = \phi_{g,1|1} = \phi_2, \quad g = 1,2$$

$$\phi_1 = \phi_2 = \phi, \quad g = 1,2$$

Thus, it is assumed that a single false positive rate applies to both trials and both domains and, further, that the false positive errors are independent between both trials. Finally, it is assumed that the false negative errors are correlated between trials and that these correlations are equal for the two domains. As with the independent model, the dependent model provides for six parameters, viz., θ_1, $\theta_{0|1}$, $\theta_{0|0}$, ϕ, π_1, and π_2, all of which are estimable.

The rather restrictive assumptions regarding the false positive errors are justified because, for most of the drugs in this study, the false positive rates are expected to be quite small. In this situation, it may be reasonable to assume that $\phi = 0$ rather than estimate ϕ. However, by allowing ϕ to be estimated, it is hoped that the likelihood function is increased and, thus, the estimates for the more important false negative probabilities are improved.

APPLICATION OF THE HUI-WALTER METHOD TO THE NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE

In this section, the Hui-Walter method is implemented to estimate the false negative and false positive probabilities associated with the so-called recency question in the NHSDA. The recency question asks respondents about the most recent time they used a particular drug. For this study, the measurement bias for this question was evaluated for alcohol, marijuana, and cocaine. By design, the NHSDA contains many redundant questions regarding drug use recency, particularly lifetime use. Because of this redundancy, the application of the Hui-Walter method to estimate NHSDA misclassification error is possible. In this section, the use of this methodology for assessing the accuracy of self-reports is demonstrated and the characteristics exhibited by the Hui-Walter estimates are critically examined.

Description of the NHSDA

The NHSDA is a multistage, household survey designed to measure the population's current and previous drug use activities. The 1993 survey was the 13th study conducted in a series initiated in 1971. Since 1990, the survey has been conducted annually, with distinct samples of households and persons selected each year. In October 1992, sponsorship of the survey was transferred from the National Institute on Drug Abuse (NIDA) to the Substance Abuse and Mental Health Services Administration, Office of Applied Studies (SAMHSA/OAS), where it currently resides.

For this research project, data from the 1991, 1992, and 1993 surveys were used in the analysis, a total of 88,000 interviews. Subsequent discussions of the NHSDA will be restricted to design and implementation issues related to these surveys.

Survey Design and Data Collection. The NHSDA is based on a national probability sample of dwelling units in the United States. For the 1991, 1992, and 1993 studies, approximately 118 primary sampling units (PSUs) were selected at the first stage of sampling. These PSUs represent geographic areas in the United States; generally defined as counties, groups of counties, or metropolitan statistical areas (MSAs). At the second stage of selection, smaller geographic areas—segments—within each PSU were selected. The NHSDA segments were defined by joining adjacent census blocks within each PSU. At the third stage of selection, a sample of dwelling units was selected within each segment

and a resident of each occupied, sampled dwelling unit was asked to participate in a screening interview for this survey. Results from this personal visit, screening interview are used to randomly select up to two members of each household. Each selected person was then asked to participate in the personal visit, interview phase of the survey. Data on a person's current and previous drug use activities were collected during this interview phase of the survey.

The target population includes persons 12 years old or older who live in households, certain group quarters (e.g., college dormitories, homeless shelters), and civilians living on military installations. Active military personnel and most transient populations, such as homeless people not residing in shelters, were not included. The sample for the 1991, 1992, and 1993 surveys was approximately 30,000 persons each year. Hispanics, blacks, younger persons, and the residents of six the MSAs were oversampled to ensure that the sample sizes were adequate to produce the subpopulation estimates of interest.

Drug and demographic data were collected from each respondent during the interview phase using a combination of interviewer-administered and self-administered instruments. On average, the interview took about an hour to complete. It began with a set of interviewer-administered questions designed to collect data on the respondent's current and previous use of cigarettes and other forms of tobacco. These initial questions allowed the respondent to become familiar with the format of the NHSDA questions.

The remainder of the questionnaire was divided into sections corresponding to each drug of interest: alcohol; the nonmedical use of sedatives, tranquilizers, stimulants, and analgesics; marijuana; inhalants; cocaine; crack; hallucinogens; and heroin. For each section, the interviewer gave the respondent an answer sheet and asked that responses be recorded on it. Depending on the complexity of an answer sheet, the interviewer either read the questions to the respondent or, if preferred, the respondent read the questions. Upon the completion of an answer sheet, the respondent was requested to place it in an envelope without allowing the interviewer to see the responses. The motivation for conducting the interview in this manner was to ensure that the respondent understood the questions, did not erroneously skip over major parts of the questionnaire and, more important, to guarantee response privacy.

Most of the answer sheets were designed so that even respondents who have never used a particular drug still needed to answer each question about the drug. Since both users and nonusers of a drug were asked to respond to essentially the same number of questions, the interviewer was less likely to guess that the respondent was a user or nonuser based on the time the respondent took to complete an answer sheet. This was another feature of the survey that was designed to protect the privacy of the respondent. In addition, some respondents who indicated that they never used the drug under direct questioning would later answer an indirect question about it in a way that implied use. This redundancy in the questionnaire, therefore, provided additional information regarding drug use that could be used to compensate for underreporting for the direct question.

Data Editing and Estimation. The raw NHSDA data are extensively edited to ensure the internal consistency of drug use responses. For the 1991, 1992, and 1993 surveys, this editing was based on a "most-recent-indication-of-use" rule. As described in the previous section, all respondents were required to respond to essentially the same questions regardless of their drug use. Consequently, use of a particular drug during a particular reference period could be logically established from responses to various questions. These questions included items presented on the specific drug answer sheet, as well as several items on the other answer sheets asking about general drug use activities.

For any particular drug, the logical editing begins with the drug recency question, a question at the beginning of each drug answer sheet that asks respondents about the most recent time they used a particular drug. As an example, on the alcohol answer sheet the recency question is:

> *When was the most recent time that you had an alcohol drink, that is, of beer, wine, or liquor or a mixed alcoholic drink?*
>
> Within the past month (30 days)
> More than 1 month ago but less than 6 months ago
> 6 or more months ago but less than 1 year ago
> 1 or more years ago but less than 3 years ago
> 3 or more years ago
> Never had a drink of beer, wine, or liquor in your life

Thus, the recency question was used to establish the most recent time a drug was used. At this first stage of editing, the recency response

categories are collapsed; for each drug, the respondents are classified into one of the following mutually exclusive categories: a past-month user, past-year user, lifetime user (i.e., any indication of use), or not a lifetime user of the drug under question. Under these editing rules, past-year users do not include past-month users and lifetime users do not include past-year nor past-month users.

After this recoding is completed, it is checked against the responses to all other questions from which drug recency can be implied. These questions include questions related to drug use that are asked on the specific drug answer sheet, as well as questions asked on the drug use activities answer sheets. For example, alcohol use can be implied from other questions on the alcohol answer sheet such as:

- *About how old were you when you first began to drink beer, wine or liquor once a month or more often?* [This question can be used to establish lifetime use of alcohol.]

- *On the average, how often in the past 12 months have you had any alcoholic beverage, that is, beer, wine or liquor?* [This question can be used to establish past year use of alcohol.]

- *What is the most you had to drink on any one day you drank beer, wine or liquor during the past 30 days?* [This question can be used to establish past month use of alcohol.]

Alcohol use also can be implied from questions on the drug use activities answer sheets such as:

- *During the past 12 months, have you gotten any treatment for drinking—such as from a clinic, self-help group, counselor, doctor or other professional?* [From the treatment answer sheet.]

- *During the past 12 months, for which drugs have you consciously tried to cut down on your use?* [From the drugs answer sheet.]

- *In the past 12 months, I felt aggressive or cross while drinking? (Y/N)* [From the drinking experiences answer sheet.]

In almost all cases where there is disagreement between the recency response and the responses to the other questions, the NHSDA editing rules dictate the respondent's final status should be changed to the most recent indication of use. If a response to some other question indicates use in a later recency period, then generally the response to the other question is deleted, and a bad data indicator response is put in its place. Because of this editing phase, a person's most recent use of any drug is determined by looking at all related questions and selecting the response for the most recent use. Unless otherwise noted, drug use estimates produced from the NHSDA are created using these edited most-recent-indication-of-use responses.

By the nature of the editing process, there is the potential for over-correcting for the negative bias in recency estimates and actually over-estimating drug use prevalence for some subgroups. At this writing, work was underway on the 1994 NHSDA to reevaluate the effects of the editing procedures. In addition, comparisons of the Hui-Walter estimates of prevalence—which are adjusted for both false negative and false positive responses—with the usual NHSDA estimates will provide important information regarding the net biases in the NHSDA estimates.

Results of the Hui-Walter Estimation

This analysis of the 1991-1993 NHSDA data is confined to three drugs—alcohol, marijuana, and cocaine. For each analysis, y_1 and y_2 are defined as follows.

$$y_1 = \begin{cases} 1 & \text{if lifetime use was indicated by the recency question response} \\ 0 & \text{if otherwise} \end{cases}$$

and

$$y_2 = \begin{cases} 1 & \text{if lifetime use was indicated by a response to any other question} \\ 0 & \text{if otherwise} \end{cases}$$

As required by the Hui-Walter procedure, two domains were defined for estimation: smokers and nonsmokers. This partitioning of the population seems to satisfy the dual criteria that the difference between the drug prevalence rates for the two domains is large—drug use among smokers tends to be considerably greater than among nonsmokers; and the assumption of equality of misclassification probabilities between the two groups is tenable.

Because y_1 and y_2 are collected in the same interview, the Hui-Walter independence model would not seem appropriate because respondents who intentionally falsify their response to the recency question would likely consistently falsify their reports throughout the questionnaire. However, because subsequent questions regarding lifetime use are less direct than the recency question, it is possible that some lifetime users who falsify on the recency question may unintentionally indicate lifetime use. Then, too, some recency question falsifiers may find the less direct questions on drug use less intimidating and may respond truthfully. There is also the potential that some lifetime users who responded "no lifetime use" in the recency question due to forgetfulness may remember later in the interview and then indicate some use.

Even accepting that some inconsistencies in the responses y_1 and y_2 are likely, the assumption that these inconsistencies satisfy the independence assumption is still questionable. Therefore, these data have been analyzed using both the Hui-Walter independence model and the dependent model assumptions; both set of results are reported.

This development of the Hui-Walter methodology for self-reported drug use is still very much in its preliminary stages. In the analyses presented here, the main objective is to investigate some capabilities and limitations of the methodology and demonstrate its use for surveys such as the NHSDA for which repeated measures are available. For this objective, the usefulness of the methodology for estimating measurement bias is critically evaluated and additional applications in the field of drug use measurement are suggested. It is possible that while the Hui-Walter false positive and negative rates are biased, their relative magnitudes still provide important insights about the causes and remedies of measurement error by identifying the socioeconomic subpopulations, data-collection procedures, and survey designs that are most prone to measurement error.

The more than 88,000 interviews collected in the 1991-93 NHSDA surveys were the object of these analyses. Table 1 gives the results of the analysis for alcohol, marijuana, and cocaine. Note that in this table the false negative rates for the dependent model are generally larger than those for the independent model. This is expected because, as Vacek (1985) has shown, positive between-trial correlations result in a downward bias in the estimated error rates under the independence model. Recall that for the dependent model, only between-trial independence for the false positive errors was assumed. Further, the dependent model provides only one parameter, ϕ, for the false positive rate. The result is a rate that is an average of ϕ_1 and ϕ_2. Because ϕ_2 is usually much smaller than ϕ_1 in the independent model, the result is also expected that ϕ for the dependent model is usually less than ϕ_1 for the independent model.

The pattern exhibited by the prevalence estimate it is also noteworthy, viz., in almost all cases

$$\hat{\pi}_{RECENCY} \leq \hat{\pi}_{INDEP} \leq \hat{\pi}_{DEP} \leq \hat{\pi}_{NHSDA}$$

As anticipated, the estimate of π from the recency question appears to be biased downward, the bias being greatest when the false negative rate is largest. Since estimates of θ for the dependent model are usually larger than for the independent model, it is also anticipated that $\hat{\pi}_{INDEP}$ is usually less than $\hat{\pi}_{DEP}$. Note also that since the NHSDA estimator does not take into account the possibility of false positive errors, it is not surprising that $\hat{\pi}_{DEP} \leq \hat{\pi}_{NHSDA}$. Finally, it is possible that $\hat{\pi}_{DEP} > \hat{\pi}_{NHSDA}$.

Let y_i denote the final edited classification for respondent i. Recall that the NHSDA estimator assigns $y_i = 1$ to any individual i for whom either y_1 or y_2 is 1. Further, if both y_1 and y_2 are 0, the NHSDA estimator assigns $y_i = 0$ to the respondent. However, the Hui-Walter estimator estimates the proportion of respondents in the population who are truly 1s though both y_1 and y_2 are 0. Thus, when these respondents are added to the number of 1 responses, it is possible for the Hui-Walter estimator to produce estimates that are larger than the NHSDA estimates, as can be observed from table 1.

Finally, the validity of the Hui-Walter estimates is considered; attention is given to degree to which the Hui-Walter estimates of measurement bias are themselves biased. Unfortunately, the evaluation of the bias in

TABLE 1. *Comparison of independent and dependent Hui-Walter estimates for the 1991-1993 NHSDA.*

Characteristic	False negative rate				False positive rate				Estimate prevalence rate (as a percent)			
	Independent		Dependent		Independent		Dependent					
	%	S.E.	%	S.E.	%	S.E.	%	S.E.	$\hat{\pi}_{INDE}$	$\hat{\pi}_{DEP}$	$\hat{\pi}_{RECENCY}$	$\hat{\pi}_{NHSDA}$
Lifetime alcohol use												
Total	1.343	(0.106)	1.780	(0.053)	0.002	(0.001)	0.083	(0.279)	83.94	84.31	82.82	84.33
Race/ethnicity												
Hispanic	1.941	(0.294)	3.149	(0.142)	0.193	(0.082)	0.212	(0.089)	76.28	77.23	74.85	77.33
Black	1.971	(0.303)	3.228	(0.157)	0.000	(0.000)	0.764	(0.151)	75.74	76.56	74.25	76.91
White/other	1.207	(0.134)	1.450	(0.063)	0.000	(0.000)	0.000	(0.000)	86.12	86.33	85.08	86.33
Age group												
12-17	0.881	(0.427)	5.671	(0.278)	0.328	(0.085)	0.413	(0.069)	39.97	41.99	39.82	42.47
18-25	0.703	(0.164)	1.482	(0.092)	0.000	(0.000)	0.000	(0.000)	87.50	88.19	86.88	88.19
26-34	1.131	(0.173)	1.114	(0.077)	0.000	(0.000)	0.606	(0.238)	92.72	92.66	91.67	92.75
35+	1.725	(0.101)	1.725	(0.102)	0.000	(0.000)	0.000	(0.000)	88.09	88.09	86.57	88.09
Gender												
Male	1.390	(0.146)	1.620	(0.072)	0.353	(0.109)	0.368	(0.112)	88.26	88.46	87.07	88.55
Female	1.288	(0.159)	1.924	(0.072)	0.000	(0.000)	0.000	(0.000)	79.95	80.46	78.92	80.46

KEY: * = Indicates estimate not available.

TABLE 1. *Comparison of independent and dependent Hui-Walter estimates for the 1991-1993 NHSDA (continued).*

Characteristic	False negative rate				False positive rate				Estimate prevalence rate (as a percent)			
	Independent		Dependent		Independent		Dependent					
	%	S.E.	%	S.E.	%	S.E.	%	S.E.	$\hat{\pi}_{INDE}$	$\hat{\pi}_{DEP}$	$\hat{\pi}_{RECENCY}$	$\hat{\pi}_{NHSDA}$
Lifetime marjuana use												
Total	0.594	(0.229)	3.384	(0.108)	0.011	(0.006)	0.014	(0.008)	33.21	34.17	33.02	34.18
Race/ethnicity												
Hispanic	1.538	(0.841)	5.439	(0.000)	0.003	(0.018)	0.003	(0.021)	27.11	28.23	26.70	28.23
Black	0.997	(0.638)	3.769	(0.261)	0.082	(0.032)	0.099	(0.037)	31.68	32.60	31.42	32.73
White/other	0.480	(0.334)	3.101	(0.142)	0.001	(0.312)	0.003	(0.012)	34.60	35.53	34.43	35.54
Age group												
12-17	0.523	(0.912)	4.711	(0.468)	0.003	(0.013)	0.006	(0.024)	10.68	11.15	10.62	11.16
18-25	0.940	(0.235)	*	*	0.016	(0.024)	0.023	(0.036)	48.72	55.41	48.27	49.10
26-34	0.355	(0.293)	1.175	(0.098)	0.037	(0.032)	0.043	(0.036)	60.23	60.73	60.03	60.76
35+	*	*	0.000	(0.000)	0.007	(0.017)	1.007	(0.060)	25.19	25.16	25.17	26.66
Gender												
Male	1.208	(0.320)	3.338	(0.151)	0.016	(0.013)	0.019	(0.015)	38.59	39.44	38.13	39.46
Female	0.000	(0.000)	0.000	(0.000)	0.007	(0.027)	0.720	(0.033)	28.36	28.36	28.37	29.39

TABLE 1. *Comparison of independent and dependent Hui-Walter estimates for the 1991-1993 NHSDA (continued).*

Characteristic	False negative rate				False positive rate				Estimate prevalence rate (as a percent)			
	Independent		Dependent		Independent		Dependent					
	%	S.E.	%	S.E.	%	S.E.	%	S.E.	$\hat{\pi}_{INDE}$	$\hat{\pi}_{DEP}$	$\hat{\pi}_{RECENCY}$	$\hat{\pi}_{NHSDA}$
Lifetime cocaine use												
Total	3.652	(0.471)	5.314	(0.230)	0.005	(0.004)	0.006	(0.004)	11.62	11.83	11.20	11.84
Race/ethnicity												
Hispanic	1.631	(1.699)	6.915	(0.553)	0.000	(0.000)	0.000	(0.000)	10.17	10.75	10.01	10.75
Black	6.027	(1.177)	8.983	(0.672)	0.003	(0.201)	0.005	(0.076)	9.67	9.98	9.09	9.99
White/other	3.471	(0.634)	4.717	(0.000)	0.005	(0.007)	0.005	(0.007)	12.29	12.45	11.87	12.46
Age group												
12-17	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.190	(0.023)	1.36	1.36	1.36	1.74
18-25	3.763	(0.671)	4.906	(0.000)	0.003	(0.015)	0.004	(0.017)	15.64	15.83	15.05	15.83
26-34	0.713	(0.723)	1.935	(0.196)	0.032	(0.019)	0.036	(0.021)	26.43	26.76	26.27	26.81
35+	6.182	(1.766)	8.299	(0.000)	0.000	(0.000)	0.000	(0.000)	8.03	8.21	7.53	8.21
Gender												
Male	4.904	(0.672)	5.271	(0.000)	0.011	(0.009)	0.011	(0.010)	14.75	14.81	14.04	14.83
Female	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.271	(0.018)	8.63	8.63	8.63	9.12

KEY: * = Indicates estimate not available.

the estimators of the error probabilities and π requires knowledge of the true error probabilities, which is not available. Sinclair (1994) and Sinclair and Gastwirth (1993) examine the sensitivity of the estimates to violations in the model assumptions. For the independent model, they found that the estimates are highly sensitive to violations of the independence assumptions. Moderately large positive correlations between errors in the two measurements can lead to substantial negative biases in the estimates of the error probabilities. Similarly, violations of the between-domain homogeneity assumption can also bias the Hui-Walter estimates; however, differences in the error rates as high as 20 percent between the two domains did not appear to bias the estimates of π appreciably. Since the dependent error model assumes homogeneity between domains but does not assume independence for the false negative errors, the results of Sinclair and Gastwirth (1993) support the claim that the dependent model estimates have greater validity than the independent model estimates.

Another indicator of the validity of the estimates is the degree to which the patterns of errors across demographic variables and the magnitudes of the estimated error rates agree with those in the published literature. Many articles attest to the high potential of underreporting for drug use self-reports, particularly among arrestee reports (see, for example, Mieczkowski 1991; General Accounting Office 1993). These researchers would tend to support the higher estimates of false negative error observed for the dependent error model rather than the smaller estimates produced by the independent model. However, since the true false negative and false positive error probabilities for the NHSDA are unknown, the existing literature is insufficient for assessing the magnitudes of the biases in the error rates obtained from either the dependent or the independent model.

Besides the question of the bias in the estimates, one can, to some extent, investigate the question of the relative validity of the Hui-Walter estimates; that is, the extent to which the estimates of misclassification error provide information regarding the relative bias in self-reports across socioeconomic classes and geographic regions, and for alternate drugs of abuse. For this analysis, the results from Fendrich and Vaughn (1994), who estimated the denial rates for the National Longitudinal Survey of Youth (NLSY) cohort, were used. For nine socioeconomic variables, they computed the proportion of respondents who admitted to using a drug (marijuana or cocaine) in the 1984 survey and then denied ever using the drug in the 1988 survey.

The NLSY is a nationally representative longitudinal sample of 12,686 individuals who were ages 14 to 21 when they were first interviewed in 1979. Twelve waves of interviews were conducted between 1979 and 1990 for the sample analyzed by Fendrich and Vaughn. Retention rates averaged about 90 percent in each of the survey years. Questions about illicit substance use were asked in 1980, 1984, and 1988. In 1988, an experiment was conducted in which half the subjects (in a selected sample) were randomly assigned to an interviewer-assisted mode and the other half to the self-administered mode.

The focus of Fendrich and Vaughn's study is on responses to the surveys administered in 1984 and 1988, since these two surveys included nearly identical questions about lifetime use for two illicit drugs—cocaine and marijuana. Their study considers two subsamples as follows: all respondents who completed the questions about marijuana use in 1984 and 1988 and also reported lifetime use of marijuana in 1984 (N = 6,204); and all respondents who completed the questions about cocaine use in 1984 and 1988 and also reported lifetime use of cocaine in 1984 (N = 1,589).

Although denial rates estimated by Fendrich and Vaughn provide direct evidence of false negative error in the NLSY, they should not be taken as estimates of the false negative probabilities because they refer only to respondents who reported any use of a drug in the first interview. Thus, the rates exclude persons who used the drug but did not report their use and respondents who never used the drug but reported that they did in the first interview.

Further, the magnitudes of the Fendrich and Vaughn denial rates are not useful for predicting the magnitudes of the NHSDA false negative error rates for a number of reasons. First, they are denial rates, not false negative rates. Second, the interview setting and mode in the NLSY are quite different from the NHSDA. While the NLSY is a panel study in which the interviewer returns annually to reinterview the respondents and may become quite familiar with them, the NHSDA is a one-time cross-sectional survey in which the interviewer and respondent have never met before. In the NHSDA, great care is taken to preserve the anonymity of the respondents and to protect their responses from discovery by the interviewer. In the NLSY, this type of confidentiality is not possible because of the nature of the survey. Finally, in the NLSY, the two measurements were separated by a period of 4 years, while in the NHSDA, the two measurements were separated by only a few minutes.

Thus, in the NLSY, there is a greater chance that the respondent's response on measurement 1 will change by the time measurement 2 is taken.

Despite these limitations of comparisons between the NHSDA and the NLSY estimates, such comparisons may still be quite fruitful. To the extent that the denial rates estimated in Fendrich and Vaughn reflect general tendencies of various socioeconomic domains to underreport their drug use, and to the extent that these tendencies and patterns for underreporting are stable over time, the estimates of false negative rates from NHSDA should be correlated, to some extent, with the denial rates from the NLSY for the same subpopulations. Lack of concordance between the two sets of estimates may not be evidence of the invalidity of either set of estimates for the reasons cited above. However, significant correlations between the two estimates are evidence of the validity of both sets of estimates as measures of the relative true-false negative error in self-reported drug use in surveys.

Table 2 shows Fendrich and Vaughn's NLSY denial rates, the NHSDA independent model false negative error estimates (NHSDA-IND), and the NHSDA dependent model false negative error estimates (NHSDA-DEP). Note first that the NLSY denial rates are considerably larger than both sets of NHSDA estimates. However, what is important here is the correlation between the NLSY and the NHSDA estimates. Table 3 displays the correlations for all pairs of the three sets of estimates for marijuana and cocaine. The "across variables" correlation is Corr (NLSY, NHSDA) across all 29 variable categories shown in table 2. The NHSDA-INDEP estimates exhibited highly significant correlation with NLSY denial rates for both marijuana (0.76) and cocaine (0.58). Surprisingly, the across variables correlations for the NHSDA-DEP estimates are not significant. The "within variables" correlation is the average correlation between categories within each of the nine variables in table 2. Here, both the NHSDA-INDEP and the NHSDA-DEP estimates exhibit highly significant correlations with the NLSY estimates for cocaine, while for marijuana the correlations are not distinguishable from 0. These results support the validity of the Hui-Walter estimates when viewed as measures of relative bias (between socioeconomic domains).

TABLE 2. *Comparison of NLSY denial rates and 1991-1993 NHSDA false negative rates: Lifetime use for 23 to 32-year-olds.*

Characteristic	Marijuana (percent)			Cocaine (percent)		
	NLSY	NHSDA-IND	NHSDA-DEP	NLSY	NHSDA-IND	NHSDA-DEP
Total 23-32 year olds	11.7	0.77	1.38	18.9	1.72	48.07
Privacy						
Private interview	12.5	0.93	0.38	18.6	1.50	2.17
Others present	10.3	0.55	0.39	22.1	1.93	2.36
Race/ethnicity						
Hispanic	14.9	2.58	3.40	20.8	0.85	3.21
Black	19.3	2.66	2.87	33.2	3.99	6.79
White/other	8.0	0.38	0.99	15.0	1.55	1.66
Gender						
Male	11.3	0.90	*	19.4	1.79	*
Female	12.2	0.62	1.40	18.3	0.00	0.00
Income						
0-$11,999	15.0	0.65	2.57	19.7	0.35	2.85
$12,000 - $19,999	11.1	0.57	0.92	20.3	2.46	3.10
$20,000 - $29,999	10.6	1.38	1.49	16.5	0.55	2.45
$30,000 - $42,999	10.2	0.49	1.53	22.4	0.74	1.48
$43,000+	9.0	0.16	0.61	22.6	0.11	2.00

KEY: * = Indicates estimate not available.

TABLE 2. *Comparison of NLSY denial rates and 1991-1993 NHSDA false negative rates: Lifetime use for 23 to 32-year-olds (continued).*

Characteristic	Marijuana (Percent)			Cocaine (Percent)		
	NLSY	NHSDA-IND	NHSDA-DEP	NLSY	NHSDA-IND	NHSDA-DEP
Education						
< High school	15.4	1.56	2.09	26.6	1.99	3.18
High school	11.6	0.29	1.69	18.9	0.56	2.92
Some college	11.3	0.56	0.88	18.7	0.16	1.14
College graduate	8.3	0.00	0.00	12.8	0.19	0.00
Labor force						
Employed	11.2	0.68	1.40	18.3	0.22	1.86
Unemployed	12.3	1.26	1.26	22.8	3.74	3.74
Not in labor force	14.6	0.77	1.35	19.7	1.44	3.39
Marital status						
Single	11.7	0.89	1.66	17.5	2.21	2.39
Married	11.8	0.78	1.28	22.1	1.58	2.45
Widowed/div/sep	11.6	0.46	1.06	14.3	0.22	1.22
Residency						
Urban	11.7	0.86	10.56	17.9	1.78	2.46
Rural	11.7	0.57	0.90	25.0	1.27	1.27
Age						
23-25	11.7	1.22	1.81	21.4	3.09	3.09
26-27	12.4	0.49	0.63	19.7	0.92	2.36
28-29	11.4	0.72	1.16	17.6	1.40	1.94
30-32	11.3	0.52	1.62	16.6	0.00	0.00

TABLE 3. *Correlational analysis of NHSDA false negative rates and NLSY denial rates for characteristics in Fendrich and Vaughn (1994).*

Correlation	Marijuana		Cocaine	
	Across var. (N = 29)	Within var. (N = 9)	Across var. (N = 29)	Within var. (N = 9)
NHSDA-IND with NLSY	0.76*	0.28	0.58*	0.57*
NHSDA-DEP with NLSY	0.06	0.01	0.02	0.55**
NHSDA-IND with NHSDA-DEP	0.15	0.41***	0.19	0.87***

KEY: * = Significant at $\alpha = 0.05$; ** = significant at $\alpha = 0.01$; *** = significant at $\alpha = 0.001$.

SUMMARY AND CONCLUSIONS

In this chapter, a general model for studying misclassification in self-reported drug use was presented and the model was then extended to the case where two measurements of the same characteristic are available for a the sample of respondents. For the two-measurements case, the general model requires seven parameters while only 3 degrees of freedom are available for estimation. Thus, some additional assumptions are required to reduce the set of unknown parameters to three or less. It was shown how the assumptions typically made for test-retest, true value, improved value, and Hui-Walter methods relate to the general model. Further, it was shown how the measures of reliability, measurement bias, estimator bias, mean squared error, false negative, and false positive probability can be defined in the context of the general model and how they may be estimated under the appropriate study designs.

Finally, the use of Hui and Walter's method for estimating misclassification error based upon two erroneous reports was demonstrated. The reports may be self-reports, biological tests, administrative record values, or any

other measure. For the general case of two measurements, the Hui-Walter method used maximum likelihood estimation to obtain estimates of the false negative and false positive probabilities associated with each measurement as well as the error adjusted estimates of prevalence based upon both measurements. The method requires that the population be divided into two domains that have markedly different prevalence rates and that satisfy the assumption of homogeneity of error probabilities.

To demonstrate the use of the Hui-Walter method for evaluating the error in self-reported drug use, the method was applied to the 1991-93 NHSDA data. Two sets of model assumptions were evaluated: the independent model and the dependent model. The dependent model yielded estimates of false negative error that were generally larger than those for the independent model. Further, the dependent model produced estimates of drug use prevalence that were very nearly the same as the NHSDA published estimates. However, an important advantage of the Hui-Walter method is that it has a probability basis for the estimation that is lacking in the NHSDA estimation procedure. In addition, the Hui-Walter estimators are adjusted for false positive errors and consistent false negative errors, while the NHSDA estimator ignores these errors.

To provide evidence of the validity of the Hui-Walter estimates, correlations between the NHSDA model-based estimates of false negative error and the NLSY denial rates were computed. The independent model exhibited highly significant average correlations across categories within the nine socioeconomic variables reported in Fendrich and Vaughn (1994). For cocaine, both models produced estimates that were significantly correlated with the NLSY within variables. This evidence suggests that the Hui-Walter method is at least useful for comparing false negative rates across socioeconomic subgroups within the same survey in order to identify which groups are most prone to false negative error. The available data were inadequate to determine whether the false positive and false negative error rates produced by the Hui-Walter method are unbiased for this application.

Future work in this area will include further study of the bias and validity of the Hui-Walter estimation method. As an example, in this application, the joint likelihood of smokers and nonsmokers was considered because this partitioning of the population seemed to fit the Hui-Walter criterion well. Other definitions for the two domains that also a priori seem to meet the Hui-Walter criteria will also be considered and the estimates

produced by each definition will be compared. Finally, attempts will be made to relate the estimates as dependent variables to subpopulation characteristics using logistic models that predict the false negative rate from variables such as age, race, sex, and income. In this way, the concurrent validity and predictive validity of the Hui-Walter estimates can be investigated.

Finally, the Hui-Walter method should be considered for studies of drug use reporting error that use a biological test (hair, urine, or nail) to evaluate the error in the self-report. As reported in the literature (e.g., Cone, this monograph), biological tests are themselves subject to considerable error, even when the period for drug use is restricted to maximize the accuracy of the test results. Self-report validity studies employing biological testing have assumed the true value or preferred value assumptions described earlier. However, the general two-measurement model in this discussion may be more appropriate for these studies. As mentioned, when the second measurement is a biological test, the assumption of between-measurement independence is likely satisfied and thus the Hui-Walter independence model can be used. Under this model, the procedure will provide estimates of false positive and false negative errors for both the self-report and the biological test result. In this way, the accuracies of both self-reports and biological tests for drug use measurement can be studied.

ACKNOWLEDGMENT

The authors gratefully acknowledge the assistance of Michael Sinclair, who generously provided them with a copies of his Ph.D. dissertation and the SAS software he developed for the Hui-Walter procedure.

REFERENCES

Bailar, B.A. Recent research in reinterview procedures. *J Am Stat Assoc* 63(1):41-63, 1968.

Biemer, P.P. Measuring data quality. In: Groves, R.M., ed. *Telephone Survey Methodology*. New York: John Wiley & Sons, 1988.

Biemer, P.P., and Stokes, S.L. Approaches to the modeling of measurement errors in surveys. In: Biemer, P.P.; Lyberg, L.E.; Massey, J.T.; Nicholls, W.L., III; and Waksberg, J., eds. *Measurement Errors in Surveys*. New York: John Wiley & Sons, 1991.

Bohrnstedt, G.W. Measurement. In: Rossi, P.H.; Wright, J.D.; and Anderson, A.B., eds. *Handbook of Survey Research.* New York: Academic Press, 1983.

Bureau of the Census. *Evaluating Censuses of Population and Housing.* Pub. No. STD-ISP-TR-5. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1984.

Cochran, W.G. Errors of measurement in statistics. *Technometrics* 10:637-666, 1968.

Fendrich, M., and Vaughn, C. Diminished lifetime substance use over time: An inquiry into differential underreporting. *Public Opin Q* 58(1):96-124, 1994.

General Accounting Office. *Drug Use Measurement: Strengths, Limitations, and Recommendations for Improvement.* Report to the Chairman, Committee on Government Operations, House of Representatives, GAO/PEMD-93-18, June, 1993.

Groves, R.M. *Survey Errors and Survey Costs.* New York: John Wiley & Sons, 1989.

Hui, S.L., and Walter, S.D. Estimating the error rates of diagnostic tests. *Biometrics* 36:167-171, 1980.

Mieczkowski, T. The accuracy of self-reported drug use: An evaluation and analysis of new data. In: Weisheit. R., ed. *Drugs, Crime and the Criminal Justice System.* Cincinnati, OH: Anderson Publishing Co., 1991.

O'Muircheartaigh, C. Simple response variance: Estimation and determinants. In: Biemer, P.; Lyberg, L.E.; Massey, J.T.; Nicholls, W.L., III; and Waksberg, J., eds. *Measurement Errors in Surveys.* New York: John Wiley & Sons, 1991.

Sinclair, M.D. "Evaluating Reinterview Survey Methods for Measuring Response Errors." Ph.D. diss, George Washington University, 1994.

Sinclair, M.D., and Gastwirth, J.L. Evaluating reinterview survey methods for measuring response errors. *Proceedings of the 1993 Annual Research Conference of the Bureau of the Census.* Washington, DC: U.S. Dept. of Commerce, 1993.

Vacek, P.M. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* 41:959-968, 1985.

AUTHORS

Paul P. Biemer, Ph.D.
Director
Survey Methods Research Program

Michael Witt, M.S.
Research Statistician
Statistics Research Division

Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709-2194

167359

The Use of External Data Sources and Ratio Estimation To Improve Estimates of Hardcore Drug Use from the NHSDA

Douglas Wright, Joseph Gfroerer, and Joan Epstein

ABSTRACT

Levels of hardcore drug use have been especially difficult to estimate because of the relative rarity of the behavior, the difficulty of locating hardcore drug users, and the tendency to underreport stigmatized behavior. This chapter presents a new application of ratio estimation, combining sample data from the National Household Survey on Drug Abuse (NHSDA) together with population counts of the number of persons arrested in the past year from the Uniform Crime Report (UCR) and the number of persons in drug treatment programs in the past year from the National Drug and Alcoholism Treatment Unit Survey (NDATUS). The population counts serve as a benchmark accounting for undercoverage and underreporting of hard drug users.

INTRODUCTION

The need for accurate estimates of the size of the so-called hardcore drug-using population is substantial. Studies of hardcore drug users typically include heavy chronic users of drugs such as cocaine and heroin. While there is no standard definition of hardcore drug users, and the authors here employ several alternative measures, this population of heavy drug users is likely to need significant resources for treatment of their drug problems and associated medical and other problems. Hardcore drug users have also been shown to be responsible for a disproportionate amount of crime (Nurco et al. 1991).

This chapter describes a method for estimating the prevalence of hardcore drug use based on the NHSDA in conjunction with outside sources and the methodology of ratio estimation. In ratio estimation, one can often obtain a better estimate of a population total if there is a known population total for a related variable. Then the estimate of the total is

$Y' = (y/x)*X$, where y is the variable of interest, x is the related variable, and X is the known population total for the related variable (Cochran 1977).

Another way of describing this method is to say that it inflates (i.e., gives more weight to) the drug prevalence data from the NHSDA for populations with characteristics that are known to be related to hardcore drug use but are also underestimated. In this case, it is known that NHSDA undercounts arrestees and drug treatment populations, so there is a ratio adjustment of the NHSDA hardcore drug use estimates upward to externally derived counts of arrestees and treatment clients that are believed to be accurate.

In survey sampling theory, ratio estimation is often associated with the desire to improve the precision of an estimate. The ratio estimate will be better, in the sense that it will have a smaller variance, than the simple expansion estimator $Y'' = \sum w_i y_i$ that is commonly used when certain conditions are met. Here, Y is the variable of interest, y_i is the value reported by the *i*-th sample case, and w_i is the inverse of the sampling probability for the *i*-th case. (Precision of the estimates is discussed later in this chapter.)

However, in this application, one is less interested in variance reduction and more interested in bias reduction. Ratio estimates have been used for a number of years to adjust for nonresponse and to adjust to known population counts, often based on a census. This application uses known population counts to adjust NHSDA sample estimates for underreporting and undercoverage. Early examples of using ratio estimation to adjust for undercoverage and nonresponse are the Health Interview Survey (Thornberry and Massey 1978) and the 1975 Survey of Scientific and Technical Personnel (Tupek and Richardson 1978), respectively. The NHSDA methodology is described elsewhere (Substance Abuse and Mental Health Services Administration (SAMHSA) 1993*a*, 1994).

BACKGROUND

The major problems with obtaining good estimates of the number of hardcore drug users are locating them and getting them to respond honestly about their usage. Studies have shown that hardcore drug users tend to be more transient, being found in significant numbers outside households (National Institute on Drug Abuse (NIDA) 1994). Other attempts to estimate the number of hardcore drug users typically have

been based on model-based assumptions or nonrandom samples. (See section on comparing with other methods.)

Estimating the number of hardcore drug users has historically been a difficult problem. Household interview surveys such as the NHSDA were not designed for this type of estimation and are believed to be inadequate tools for measuring hardcore drug use because of the low prevalence of the behavior and difficulties in accessing this population. Underreporting (survey participants who do not report their drug use) and undercoverage (inability to roster hardcore drug users) also affect this estimation.

In comparing the results of NHSDA estimates to those from various administrative records systems (e.g., drug treatment program data, parole, probation, or arrest data from the Federal Bureau of Investigation (FBI)), the apparent underreporting of these types of characteristics by the sample respondents has been significant. It should be noted that FBI Index Crimes appear to be fairly well estimated by the NHSDA when allowing for differences due to incarceration (Harrison and Gfroerer 1992).

Research has shown that underreporting of drug use increases as the reference period approaches the present and as the perceived social disapproval increases. This suggests that hardcore drug use is underestimated more than casual drug use. The underestimates could also be the result of undercoverage of the populations with these characteristics (Gfroerer 1993; Turner et al. 1992).

Various methods have been used to estimate hardcore drug prevalence including capture-recapture techniques, truncated Poisson, and modeling methods generally (c.f., Brodsky 1985; Hser et al. 1992; Woodward et al. 1985). These methods have been based primarily on data from administrative records such as treatment admission data, essentially ignoring household survey data. Other methods have combined household survey data with other sources of data (e.g., arrest data) to construct composite estimates of hardcore drug use (Rhodes 1993; Wish 1990-91).

There also has been significant research on various data-collection methods that encourage honest responses to sensitive questions. Such methods include randomized response and nominative techniques (Miller 1985; Warner 1965; Zdep et al. 1979).

Randomized response involves a randomizing device, such as a pair of dice, and two questions, one of which is sensitive ("Have you used heroin in the past year?") and one of which is innocuous (e.g., "Were you born in April?"). The respondent uses the randomizing device to determine which question to answer, and the interviewer records the answer (yes or no), but not the question. This method can be used to estimate the proportion having the sensitive characteristic.

The nominative technique involves first asking the respondent to indicate how many of his or her close friends have the sensitive characteristic (e.g., use of heroin in the past year). The second question is then asked about each of the close friends: "How many of this person's other close friends (besides yourself) know that he (or she) has used heroin?" With these two questions the count of the number of past-year heroin users can be corrected for duplication.

In the following discussion, the focus is on the ratio estimate's ability to reduce bias (in particular, the undercounting of hard drug users in the NHSDA) given a true population value of a related variable. To make the discussion more concrete, the estimation procedure will be applied to four separate, but overlapping, measures of hardcore drug use for 1992: the number of past year users of heroin, weekly users of cocaine in the past year, past-year users who are dependent on some illicit drug, and past-year intravenous drug users.

BASIC METHODOLOGY

The information that the authors wish to use are the count of the number of persons in treatment centers for drug abuse during 1992 from the NDATUS (SAMHSA 1993*b*) and the known count of the number of arrests (for any crime other than minor traffic violations) during 1991 from the FBI Uniform Crime Reporting Program (Maguire et al. 1993).

Using a Single Population Count

Let Nt equal the estimated count of the number who received treatment for a drug problem during 1992 derived from the NDATUS. The count was computed by multiplying the number of treatment slots times the average number of persons treated per year per slot, and includes an adjustment for multiple episodes by the same individual.

Na equals the estimated count of the number of persons arrested during 1991. Na was calculated by taking the latest available FBI Uniform Crime Report estimated number of arrests—14,211,900—and dividing this by the average number of arrests per person arrested calculated from the NHSDA, approximately 1.46, resulting in an estimate of 9,722,671. Based on recent trends, the 1992 estimate would be expected to be slightly higher than the 1991 estimate.

The typical use of a ratio estimate occurs when the outside source fully overlaps the population of interest. For example, if one wanted to estimate the number of students in school and had information on the total number of teachers, then this source overlaps the population of interest since every student has a teacher. Then, if an estimate of the average number of students per teacher were developed, one could multiply the total number of teachers times the average number of students per teacher to obtain an estimate of the total number of students.

The authors' situation is slightly different in that neither the treatment nor arrestee population counts fully overlap the population of hard drug users. However, one can construct counts from the Census Bureau estimates for 1992 that are so precise at the national level that they can be considered population counts. With these data, one can develop counts that cover the population.

Given the Census Bureau estimate of the number of noninstitutionalized persons 12 and older, N = 205,713,000 (for July 1, 1992, the count at the midpoint of data collection for the NHSDA target population), one can form two pairs of counts that cover the population.

Number of persons in treatment: Nt = 1,789,000
Number of persons not in treatment: N-Nt = 203,924,000

Number of persons arrested: Na = 9,722,671
Number of persons not arrested: N-Na = 195,990,329.

From these counts, two estimates of the number of persons using heroin (H in the following equation) during the past year (1992) are possible.

$$H(t) = r(t)*Nt + r(t)^{*}*(N-Nt) = (139,003/834,702) * 1,789,000(171,136/200,656,309) * 203,924,000$$
$$= .166*1,789,000 + .00085*203,924,000$$
$$= 297,922 + 173,922$$
$$= 471,844.$$

$$H(a) = r(a)*Na + r(a)^{*}*(N-Na) = (184,277/4,743,706) * 9,722,671 + (125,861/196,747,306) * 195,990,329$$
$$= .039*9,722,671 + .00064*195,990,329$$
$$= 377,693 + 125,377$$
$$= 503,070$$

where $r(t) = h_t/t$: the estimated rate of hardcore drug use (heroin, in this example) in population Nt. For heroin, it is the estimated number (from the NHSDA sample) in treatment and using heroin in the past year divided by the estimated number (from the sample) in treatment centers for drug use in the past year. (HERREC was the heroin variable used from the 1992 NHSDA file to code if the respondent used heroin in the past year; the treatment variable was the union of three variables TRMTTRMT (received treatment for drug use in a treatment or rehab facility in the past year), TRMTHOSP (received treatment for drug use in a hospital in the past year), and TRMTMHC (received treatment for drug use in a mental health center in the past year.) The treatment variable was defined in this way in order to be as consistent as possible with the data collected in NDATUS.)

The estimated rate of hardcore drug use in population N-Nt is denoted $r(t)^{*}$. For heroin, it is the estimated number (from the sample) not in treatment but using heroin in the past year divided by the estimated number (from the sample) not in treatment centers for drug use during the past year.

The estimated rate of hardcore drug use in population Na is denoted $r(a) = h_a/a$. In this example, it is the estimated number (from the sample) arrested and booked and using heroin in the past year divided by the estimated number (from the sample) arrested and booked in the past year (the union of the variable NOBOOKYR with BKLARCNY, BKBURGL, BKAGASLT, BKSMASLT, BKMVTHFT, BKROB, BKRAPE, BKMURDER, BKARSON BKDRVINF, BKDRUNK, BKDRUG, BKPROS, BKVANDAL, and BKOTHOF).

The estimated rate of hardcore drug use in population N-Na is denoted $r(a)^*$. It is the estimated number (from the sample) not arrested and not booked but using heroin during the past year divided by the estimated number (from the sample) neither arrested nor booked in the past year.

The above estimates can be compared to the standard published estimate of 323,000 for the number of past-year users of heroin in 1992. The standard published estimate is a complex estimator incorporating the inverse of the selection probability, adjustments for household and person-level nonresponse, and a poststratification adjustment to census projections of age by gender by race/ethnicity distributions.

Using Two Known Population Counts

Having the two separate estimates based on treatment and arrests counts raises the question, "Is there an alternate method that would make simultaneous use of both the treatment and arrest counts?"

The ideal situation when one has two variables, such as the number receiving treatment in the past year and the number arrested and booked in the past year, is to use known counts for the interior cells. In other words, one can make consistent estimates that make use of ratio estimation for each of the cells using the following matrix:

	Treatment	No treatment
Arrested	N(11)	N(12)
Not arrested	N(21)	N(22)

where N(11) is the known count of the number in treatment and arrested and booked in the past year and so forth for the other cells. (It is interesting to note that if there is no partial nonresponse for the variables of interest, and estimates are used of the number in each cell based on the NHSDA sample multiplied by the sample cell estimates of prevalence, the NHSDA standard published estimate is obtained. The impetus for using accurate external counts is that the sample counts tend to underestimate counts of those in treatment and those arrested and booked.)

In the earlier description of the calculation for heroin, all that was known at the national level were the marginals Nt, the number in treatment in

the past year, and Na, the number arrested and booked in the past year. In the absence of known population counts, exploring a few alternatives is possible. One alternative is to fill in the interior cell counts by raking the marginals (or iteratively proportionally fitting); for example, by raking based on the least squares solution. However, this solution, often suggested when the initial interior cells exhibit variation due to sampling error, minimizes the change of the interior estimates. This would be inappropriate here because the authors believe the cell estimates are biased downward.

A second alternative, and the one the authors employ, is to use independent sample estimates of one cell to estimate the remaining cells. To estimate one of the interior cells, sample data from the 1990 Drug Services Research Survey (DSRS) was used in conjunction with data from the NHSDA and the estimated marginal count from NDATUS.

DSRS is a national sample survey of treatment centers and records of discharged clients. In 1990, a stratified random sample of 1,183 drug treatment facilities was selected from respondents to the 1990 NDATUS file. In the summer of 1990, a stratified subsample of 118 combined drug and alcohol treatment facilities was selected. In each facility, a sample of client discharge records (discharged during the 12 months between September 1, 1989 and August 31, 1990) was selected. The total sample included over 2,000 records (Bigel Institute for Health Policy 1992).

The aim is to estimate the number of persons in treatment during 1992 who also were arrested and booked in 1992, represented as N(11). Because the number of persons in treatment (Nt) is known, the only task is to estimate the percent of those in treatment who also were arrested and booked in 1992.

From DSRS, one can estimate that the proportion of those in treatment who were ever arrested was 77 percent. This number is based on records having information on arrests and on the number of episodes of treatment in the past year. Approximately 30 percent of the records were missing one or more of these variables. The estimate from the 1990 DSRS survey has been assumed to hold true for the target year, 1992. From the NHSDA, one can estimate that the proportion of those in treatment during the past year and ever arrested who were also arrested and booked in the past year was 52 percent. Therefore, it is necessary to multiply 0.77 by 0.52 times the number in treatment, 1,789,000, to obtain

716,315, the number in treatment and arrested and booked in the past year.

Counts for the interior cells N(11), N(12), N(21), and N(22) that are consistent with the marginal counts used earlier, can now be obtained.

	Treatment	No treatment	Total
Arrested	716,315	9,006,356	9,722,671
Not arrested	1,072,685	194,917,644	195,990,329
Total	1,789,000	203,924,000	205,713,000

The following notation can be used to represent the corresponding ratios (i.e., prevalence rates):

	Treatment	No treatment	Marginal avg.
Arrested	r(11)	r(12)	r(a)
Not arrested	r(21)	r(22)	r(a)*
Marginal avg.	r(t)	r(t)*	r

Then the estimate based on interior cells can be written as:

Estimate = H(t,a) = r(11)*N(11) + r(12)*N(12) + r(21)*N(21) r(22)*N(22).

Returning to the estimation of heroin, the prevalence rates in the cells are:

	Treatment	No treatment	Marginal avg.
Arrested	0.308	0.02100	0.03900
Not arrested	0.092	0.00038	0.00064
Marginal avg.	0.166	0.00085	0.00160

(The grand average is the standard published estimate divided by the population total, 205,713,000.)

The estimate based on all four cells rather than the marginals is 587,966, which was calculated as follows:

$$H(t,a) = (88,495/287,200)*716,315 + (50,507/547,502)*1,072,685 + (95,781/4,456,505)*9,006,356 + (75,354/196,199,804)*194,917,644 = 587,966.$$

This estimate is larger than either of the estimates based on marginals. By analogy with stratification and poststratification, the interior cells estimate is generally to be preferred to marginal estimates, especially if there are large differences in the usage ratios among the cells.

Two comments about interpretation of the above numbers are in order. First, since the above ratios have been calculated only for those records for which all the variables exist, including the treatment and arrest variables, one would expect the denominators to be somewhat of an underestimate of the independent count totals. Second, notwithstanding that fact, two of the cells are significant underestimates by a factor of 2 or more. For example, the NHSDA estimated number of those arrested and booked in the past year but not in treatment, 4,456,505, is less than half of the independent count of 9,006,356. The third cell, those in treatment but not arrested, is underestimated by a factor slightly less than 2. The last cell, those not in treatment and not arrested in the past year, is very close to the independent count, so that the impact of the ratio estimate on this cell is minimal.

To obtain the most accurate estimate, one should utilize independent population counts that are believed to have the best coverage and are most closely related to the variable of interest. One can also calculate the estimated variance of alternative estimates, preferring the one with the smallest variance, other things being equal.

The mean squared error of the estimate based on the interior cells (given that the interior cells are known counts) will typically be smaller as well, as long as the cell sample sizes are sufficiently large and certain relationships between the numerator and denominator of cell prevalence rates hold (in particular, that the correlation between the numerator and denominator is less than the population coefficient of variation (cv) of the denominator divided by twice the population cv of the numerator; the variances should always be estimated to verify that this is the case).

Assumptions

One major assumption made in the above estimation methodology concerns the accuracy of the estimated ratios given the expectation of underestimation of these hard drug populations from the household sample. A basic assumption being made in these ratios is that both numerator and denominator are being similarly underestimated. This would be the case, for example, if drug users underreport their drug use (or it is undercovered) at the same rate as the treatment population underreports their treatment and arrestees underreport being arrested. Taking the estimate of r(t), the expected value of h_t is assumed to equal $c_{ht}*H_t$, where c_{ht} is a constant and H_t is the true value. Similarly, the expected value of t is assumed to equal c_t*Nt, so that the expected value of r(t) equals approximately H_t/Nt when $c_{ht}=c_t$. (Another possible assumption is that $c_{ht}<c_t$ because some will assert that the least accurate NHSDA coverage is of the heaviest users.)

For the complementary cell in the 2-cell estimate, the numerator of the ratio r(t)*, the number of heroin users who are not in treatment, is probably underestimated. But the sample estimate of the denominator will generally not be an underestimate, so that the impact of ratio estimation on this cell is minimal.

Similarly, in the 4-cell estimate, the cell estimate of those who have not been in treatment or arrested and booked in the past year is very similar to the independent count for that cell. Here also, one expects that the assumption of equal underestimating of the numerator and denominator would not hold. Therefore, the ratio estimate still would underestimate this cell for any measure of hard drug use.

Another issue is the extent to which the authors' sample ratios cover the entire population (noninstitutionalized individuals age 12 and over). By design, the NHSDA only covers those who spend some portion of the year in a household, dormitory, or shelter. It does not cover the population that is composed of persons who are in a residential treatment facility for a full year. These occur relatively infrequently since most treatment is outpatient. For those arrested and booked, the group missed in the household survey is those prisoners who are in prison for the full 12 months because they cannot be in the household sample. The "permanently" homeless (those who do not appear in a household or a shelter for an entire year) also are missing.

To calculate such ratios as r(t), r(a), and others, only those cases that had no missing data for any of the variables mentioned above were used. In other words, each variable was coded 1, 0, or missing. Over three variables there are 27 combinations, but only the 8 combinations of 1s and 0s were used. The implicit assumption is that the item nonresponse is at random. About 1,000 cases had one or more variables missing, but out of a sample of 28,000, this only represents about 3 percent.

It should be noted that the ratio adjustment can be looked at differently, as, for example, $h_t * (N_t/t)$. Looked at this way, the ratio adjustment is a weight adjustment that differs by cell, applied to the usual weight. This simplifies the application of this method and provides the capability of estimating hardcore drug use for population subgroups or for other drugs. Thus, the standard NHSDA estimate Y″ is replaced by $Y'' (adj) = \sum w_i(adj)*y_i$, where $w_i(adj)$ is the usual weight adjusted by the appropriate cell ratio.

ESTIMATES OF OTHER MEASURES OF HARDCORE DRUG USE

Below are estimates of other measures of hardcore drug use based on the above methodology. Table 1 displays the prevalence estimates by cell, and table 2 shows the resulting estimates.

Implications of Components of the Estimator

It is instructive to look at the relative size of certain components to gain insight into how the estimate might be improved and where its weaknesses are. Note that in the above assumptions, the levels of the components that make up the ratios r(1) and r(2) are not really important to the size of the estimates H and C, respectively, if one believes that the undercounting of the numerator and denominator are similar. The estimate for H, for example, would be larger if one believes that there has been more underreporting of heroin users in treatment than there has been of the total persons in treatment.

The next components to look at are the denominators of the ratios as compared to the independent counts (actually, partly estimated). The sample-based estimate from the 1992 NHSDA for the number of persons in treatment during the past year is 834,702. The universe estimate from NDATUS (adjusted for certain double-counting and for nonresponse, but

TABLE 1. *Cell estimates for other measures of hardcore drug use (variables in parentheses indicate variable names on the NHSDA public use file).*

Cocaine (COCWKF)	Treatment	No treatment	Marginal avg.
Arrested	0.259	0.0160	0.0310
Not arrested	0.108	0.0020	0.0022
Marginal avg.	0.160	0.0023	0.0031
Dependence on any illicit drug*			
Arrested	0.519	0.0930	0.1180
Not arrested	0.348	0.0066	0.0076
Marginal avg.	0.406	0.0085	0.0102
Past year use of needles (NEDYR3)			
Arrested	0.238	0.0560	0.0670
Not arrested	0.068	0.0014	0.0016
Marginal avg.	0.127	0.0026	0.0032

KEY: * = The number of persons dependent on any illicit drug in the past year is based on an algorithm to approximate the "Diagnostic and Statistical Manual of Mental Disorders," 3d ed. revised (DSM-III-R) criteria (Epstein and Gfroerer 1995). This algorithm combines items on symptoms and problems included in the NHSDA questionnaire to approximate five of the nine DSM-III-R criteria for substance dependence and defines as dependent a person who meets two of the five criteria.

not adjusted for any undercoverage from having a incomplete frame) is 1,789,000.

This would imply that if these two numbers are similarly underestimated, then the NHSDA survey undercounts the NDATUS number by 53 percent (or more). Similarly, the sample-based estimate for the number arrested and booked in the past year is 4,743,706, while the estimate of the number of persons arrested in the past year derived from the UCR is 9,722,671. The sample-based estimate "underestimates" the universe estimate of 9,722,671 by 51 percent. "Underestimates" is put in quotes because the two numbers may not be completely analogous. The sample estimate is of persons arrested and booked, while the universe count is for those arrested. In most arrests, however, if one is arrested, one is usually processed or booked. It is apparent for both of these questions that they are significantly underreported to the NHSDA. The

TABLE 2. *Comparison of ratio estimate to standard estimate for various drugs.*

	Ratio estimates			Standard estimate
Drug usage	Treatment/ arrest ratio	Treatment marginal ratio	Arrest marginal ratio	Standard published estimate
Past year heroin	587,966	471,844	503,070	323,000
Weekly use of cocaine	829,017	750,504	742,202	642,221
Past year dependence on any drug	2,869,242	2,467,074	2,635,084	2,104,508
Past year needle use	1,019,165	755,977	960,773	659,292

independent counts of the variables enable one to adjust for this underreporting.

What is the impact of this kind of estimation on other non-hardcore drugs? Generally speaking, it is not as dramatic as with the hardcore drugs. The impact was calculated on use of marijuana in the past year (variable MRJYR). The standard published estimate was 17,400,273, while the ratio estimate based on four cells was 19,461,280. The latter estimate is only 12 percent larger.

The reason for this is that marijuana is used more widely in the population, and most users fall in the no treatment/not arrested cell. Therefore, this cell (13,644,235 users) dominates the estimate. The relative differences in prevalence rates among cells is not as dramatic as with the hardcore drugs. The cell rates are as follows:

	Treatment	No treatment
Arrested	0.91	0.46
Not arrested	0.45	0.07

Precision of Estimates

Estimates of variance were calculated for each of the above estimates using a software package that calculates the variance of complex sample surveys using Taylor series (Research Triangle Institute 1992). The current authors used the ratio estimation procedure with poststratification weights. It was assumed that the independent counts were estimated without error.

The estimated standard error for the ratio estimate of heroin was very similar to the estimated standard error for the standard published estimate, approximately 106,000. Even though the cell count of those who have neither been in treatment nor arrested and booked in the past year is the largest of the four cells, its contribution both to the estimate and to the estimate of variance was relatively small because of the estimated low incidence in that cell. The 95 percent (2 σ) confidence interval for the estimate of past-year users of heroin was 587,966 plus or minus 212,000.

For the other variables, the standard errors for the standard NHSDA estimates were similar to those for the ratio estimates. Since the standard errors of the estimates have remained similar while the estimates themselves have increased, the coefficients of variation (the relative precision of the estimates) have been somewhat improved.

COMPARISON WITH ESTIMATES FROM OTHER METHODS

Previous national estimates of hardcore drug use have used widely varying methods. Estimates of heroin prevalence published by NIDA in the 1970s relied on a small number of locally derived prevalence estimates that were projected to the entire nation using available heroin problem indicators available in other locations (Person et al. 1977). These estimates of the number of heroin addicts ranged from 584,000 in 1974 to 420,000 in 1979. However, these are not comparable to estimates of any past-year heroin use because past-year users include both addicts and infrequent users.

A nominative method of estimating heroin prevalence from the NHSDA produced an estimate of 1.9 million past-year heroin users in 1982 (Miller 1985). A recent estimate of 658,000 weekly heroin users in 1990 was derived from a synthetic estimation procedure that involved

combining multiple data sources under various assumptions (Rhodes 1993). This same methodology was used to derive an estimate of 2.1 million weekly cocaine users in 1991. These recent synthetic estimates represent the most rigorous attempts to utilize multiple sources of data in estimating hardcore drug use prevalence.

While there are many differences between the synthetic estimation model and the NHSDA ratio estimation, the large discrepancies in estimates from the two methods are largely explained by the assumptions made regarding the arrestee population. The synthetic model relied heavily on drug prevalence data from the Drug Use Forecasting system (not a representative sample of arrestees), which resulted in an estimate of 1.8 million weekly cocaine users and 500,000 weekly heroin users among arrestees. By contrast, the ratio estimation method relies more heavily on NHSDA drug prevalence data for arrestees, and resulted in an estimated 329,626 weekly cocaine users and 414,265 past-year heroin users among arrestees.

A complete evaluation and comparison of the ratio estimation procedure with other methods of estimating hardcore drug use is beyond the scope of this chapter. However, one can make some overall statements about ratio estimation.

- Ratio estimation does not fully account for underreporting and undercoverage in the NHSDA. In particular, for the population not arrested and not in treatment, the method does not adjust for under- reporting at all. Thus, the authors consider these estimates of hardcore drug use to be improvements on the standard published NHSDA estimates but still conservative. The standard published NHSDA estimates already adjust for under- or overcoverage by age, race/ethnicity, and gender.

- Because ratio estimation can be looked at as an adjustment to the NHSDA analytic weights (which are based on a probability-based sample design), it provides analytic capabilities that are not possible in any of the previously used methods. While other methods essentially focus on obtaining the bottom line estimate of the number of hardcore drug users, by constructing estimates within the framework of the NHSDA data set one can extend the estimation to population subgroups, such as by region, gender, race/ethnicity, and income, taking advantage of the multitude of data collected in the NHSDA. It must be acknowledged, however,

that there are limitations to these secondary applications of the ratio estimation procedure that are not yet determined. Because the procedure is designed to improve national hardcore drug use estimates, it may not be appropriate (without modification) for certain other estimations, such as for some subgroups and for other drug use measures (e.g., casual use).

- The ratio estimation model, as applied in this case, relies primarily on regularly updated and consistently collected data from the NHSDA, NDATUS, and UCR, and a relatively small number of easily understood assumptions. Thus, it is likely to be able to provide more reliable trend information (given constant levels of underreporting) than the previously used methods, which rely more heavily on assumptions that could change over time.

- Because ratio estimation relies primarily on the NHSDA sample design and weighting, it is possible to develop estimates of the variances of ratio-adjusted estimates. This is generally not possible in the methods previously used.

POSSIBLE FUTURE RESEARCH/APPLICATIONS

There are three primary areas for further investigation. One is in the population counts. Another is the assumptions made about the ratios used. The third involves a search for unbiased methods to estimate the ratio.

1. It would be useful to explore the development of more accurate estimates for the four-cell counts or of alternative counts based on different variables. Estimating the counts used in this chapter necessitated using multiple sources to make the counts comparable to what is collected by NHSDA. Generally, this is best accomplished by coordinating the questions on the NHSDA and other surveys with systems used to develop administrative counts so that the definitions are as consistent as possible. Coordination of item wording among surveys will at a minimum make it possible to compare estimates across surveys. For 1994, the NHSDA question on being in treatment has been changed to agree exactly with the definition used in NDATUS.

Since it is known that age and race are major correlates of the rate of drug usage, another improvement would be to seek a source or a method of estimation that could provide further age/race breakouts to the treatment/arrest cell counts.

2. In the area of assumptions, when possible, one can compare the distributions of persons for a variable used in the cross-classification based on the NHSDA to those of the population frames to see if they are similar. For example, one can compare the distribution of those in treatment from the NHSDA to the distribution of the population values from NDATUS that are available by age, race, and gender.

Another possibility is the introduction of additional weights reflecting the proportion of the year that a person is in treatment or living in a household. This would serve to increase the size of the populations that are not year-round household residents.

3. With respect to the instrument, one could perhaps try to introduce methodology that would result in less undercounting of the variables that form the ratios: heroin, treatment, arrested and booked, and others, possibly using multiplicity methods, nominative techniques, or using some new method such as hair tests (if the methodology proves to be feasible) to confirm drug use or nonuse.

4. Other applications: While population counts were not directly available at the national level for each of the interior cells, similar methodology may be useful for smaller geographic entities such as States, where the interior cell counts may be known. To use this methodology, one would have to conduct a prevalence survey including questions as similar as possible to what is collected in existing population counts of related social indicators. Again, more coordination of item wording between sample surveys and administrative file systems would enhance this kind of estimation.

REFERENCES

Bigel Institute for Health Policy, Brandeis University. *Drug Services Research Survey - Final Report: Phase II.* Final report of activities under National Institute on Drug Abuse contract no. 271-90-8319/1. Waltham, MA: Brandeis University, 1992.

Brodsky, M.D. History of heroin prevalence estimation techniques. In: Rouse, B.; Kozel, N.; and Richards, L., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity*. National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985. pp. 94-103.

Cochran, W.G. *Sampling Techniques*. New York: John Wiley & Sons, Inc., 1977.

Epstein, J.F., and Gfroerer, J.C. "Estimating Substance Abuse Treatment Need From a National Household Survey." Paper presented at the 37th International Congress on Drug and Alcohol Dependence, San Diego, 1995.

Gfroerer, J.C. An overview of the National Household Survey on Drug Abuse and related methodological research. *Proceedings of the Survey Research Section of the American Statistical Association*. Washington, DC: American Statistical Association, 1993.

Harrison, L., and Gfroerer, J. The intersection of drug use and criminal behavior—Results from the National Household Survey on Drug Abuse. *Crime Delinquency* 38(4):422-443, 1992.

Hser, Y.I.; Anglin, M.D.; Wickens, T.D.; Brecht, M.L.; and Homer, J. *Techniques for the Estimation of Illicit Drug-Use Prevalence: An Overview of Relevant Issues*. National Institute of Justice Pub. No.(NCJ)133786. Washington, DC: National Institute of Justice, 1992.

Maguire, K.; Pastore, A.L.; and Flanagan, T.J., eds. *Sourcebook of Criminal Justice Statistics*. U.S. Department of Justice, Bureau of Justice Statistics. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1993.

Miller, J.D. The nominative technique: A new method of estimating heroin prevalence. In: Rouse, B.; Kozel, N.; and Richards, L., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity*. National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985. pp. 104-118.

National Institute on Drug Abuse. *Prevalence of Drug Use in the DC Metropolitan Area Household and Nonhousehold Populations: 1991*. Technical Report 8. Rockville, MD: National Institute on Drug Abuse, 1994.

Nurco, D.N.; Hanlon, T.E.; and Kinlock, T.W. *Recent Research on the Relationship Between Illicit Drug Use and Crimes*. Vol. 9. *Behavioral Sciences and the Law*. New York: John Wiley & Sons, 1991.

Person, P.H.; Retka, R.L.; and Woodward, A.J. *A Method for Estimating Heroin Use Prevalence.* DHEW Publication No.(ADM)77-439. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1977.

Research Triangle Institute. *SUDAAN Users Manual.* Research Triangle Park, NC: Research Triangle Institute, 1992.

Rhodes, W. Synthetic estimation applied to the prevalence of drug use. *J Drug Issues* 23:297-321, 1993.

Substance Abuse and Mental Health Services Administration. *National Household Survey on Drug Abuse: Population Estimates, 1992.* DHHS Pub. No. (SMA)93-2053. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1993*a*.

Substance Abuse and Mental Health Services Administration. *National Drug and Alcohol Treatment Unit Survey (NDATUS) 1991 Main Findings Report.* DHHS Pub. No. (SMA)93-2007. Rockville, MD: U.S. Department of Health and Human Services, 1993*b*.

Substance Abuse and Mental Health Services Administration. *Preliminary Estimates from the 1993 National Household Survey on Drug Abuse.* Advance report number 7. Rockville, MD: U.S. Department of Health and Human Services, 1994.

Thornberry, O.T, and Massey, J.T. Correcting for undercoverage bias in random digit dialed national health surveys. *Annual Proceedings of the American Statistical Association.* Washington, DC: American Statistical Association, 1978.

Tupek, A.R., and Richardson, W.J. Use of ratio estimates to compensate for nonresponse bias in certain economic surveys. *Annual Proceedings of the American Statistical Association.* Washington, DC: American Statistical Association, 1978.

Turner, C.F.; Lessler, J.T.; and Gfroerer, J.C. *Survey Measurement of Drug Use: Methodological Studies.* DHHS Pub. No. (ADM)92-1929. Rockville, MD: National Institute on Drug Abuse, 1992.

Warner, S.L. Randomized response: A survey technique for eliminating evasive answer bias. *J Am Stat Assoc* 60:63-69, 1965.

Wish, E.D. Drug policy in the 1990s: Insights from new data on arrestees. *Int J Addict* 25:377-409, 1990-1991.

Woodward, J.A.; Bonett, D.G.; and Brecht, M.L. Estimating the size of a heroin abusing population using multiple-recapture census. In: Rouse, B.; Kozel, N.; and Richards, L., eds. *Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity.* National Institute on Drug Abuse Research Monograph 57. DHHS Pub. No. (ADM)85-1402. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985. pp. 158-171.

Zdep, S.M.; Rhodes, I.N.; Schwartz, R.M.; and Kilkenny, M.J. The validity of the randomized response technique. *Pub Opinion Q* 43:544-549, 1979.

AUTHORS

Douglas Wright, M.S.
Mathematical Statistician

Joseph Gfroerer
Chief, Prevalence Branch

Joan Epstein, M.A.
Health Statistician

Office of Applied Studies
Substance Abuse and Mental Health Services Administration
Room 16C-06
5600 Fishers Lane
Rockville, MD 20857

National Institute on Drug Abuse

Research MONOGRAPH SERIES

While limited supplies last, single copies of the following monographs may be obtained free of charge from the National Clearinghouse for Alcohol and Drug Information (NCADI). Please also contact NCADI for information about other publications of the National Institute on Drug Abuse relevant to drug abuse research.

Additional copies may be purchased from the U.S. Government Printing Office (GPO) and/or the National Technical Information Service (NTIS) as indicated. NTIS prices are for paper copy; add $3.00 handling charge for each order. Microfiche copies also are available from NTIS. Prices from either source are subject to change.

Addresses are:

NCADI
National Clearinghouse for Alcohol and Drug Information
P.O. Box 2345
Rockville, MD 20852
(301) 468-2600
(800) 729-6686

GPO
Superintendent of Documents
U.S. Government Printing Office
P.O. Box 371954
Pittsburgh, PA 15220-7954
(202) 738-3238
FAX (202) 512-2233

NTIS
National Technical Information Service
U.S. Department of Commerce
Springfield, VA 22161
(703) 487-4650

For information on availability of NIDA Research Monographs from 1975-1996 and those not listed, write to NIDA, Public Information Branch, Room 10A-39, 5600 Fishers Lane, Rockville, MD 20857.

26 THE BEHAVIORAL ASPECTS OF SMOKING.
Norman A. Krasnegor, Ph.D., ed. (Reprint from 1979 Surgeon General's Report on Smoking and Health.)
NCADI #M26 NTIS PB #80-118755/AS (A09) $27.00

42 THE ANALYSIS OF CANNABINOIDS IN BIOLOGICAL FLUIDS. Richard L. Hawks, Ph.D., ed.
NCADI #M42 NTIS PB #83-136044/AS (A07) $27.00

50 COCAINE: PHARMACOLOGY, EFFECTS, AND TREATMENT OF ABUSE. John Grabowski, Ph.D., ed.
NCADI #M50 NTIS PB #85-150381/AS (A07) $27.00

52 TESTING DRUGS FOR PHYSICAL DEPENDENCE POTENTIAL AND ABUSE LIABILITY. Joseph V. Brady, Ph.D., and Scott E. Lukas, Ph.D., eds.
NCADI #M52 NTIS PB #85-150373/AS (A08) $27.00

53 PHARMACOLOGICAL ADJUNCTS IN SMOKING CESSATION. John Grabowski, Ph.D., and Sharon M. Hall, Ph.D., eds.
NCADI #M53 NTIS PB #89-123186/AS (A07) $27.00

54 MECHANISMS OF TOLERANCE AND DEPENDENCE. Charles Wm. Sharp, Ph.D., ed.
NCADI #M54 NTIS PB #89-103279/AS (A19) $52.00

56 ETIOLOGY OF DRUG ABUSE: IMPLICATIONS FOR PREVENTION. Coryl LaRue Jones, Ph.D., and Robert J. Battjes, D.S.W., eds.
NCADI #M56 NTIS PB #89-123160/AS (A13) $36.50

61 COCAINE USE IN AMERICA: EPIDEMIOLOGIC AND CLINICAL PERSPECTIVES. Nicholas J. Kozel, M.S., and Edgar H. Adams, M.S., eds.
NCADI #M61 NTIS PB #89-131866/AS (A11) $36.50

62 NEUROSCIENCE METHODS IN DRUG ABUSE RESEARCH. Roger M. Brown, Ph.D., and David P. Friedman, Ph.D., eds.
NCADI #M62 NTIS PB #89-130660/AS (A08) $27.00

63 PREVENTION RESEARCH: DETERRING DRUG ABUSE AMONG CHILDREN AND ADOLESCENTS. Catherine S. Bell, M.S., and Robert J. Battjes, D.S.W., eds.
NCADI #M63 NTIS PB #89-103287/AS (A11) $36.50

64 PHENCYCLIDINE: AN UPDATE. Doris H. Clouet, Ph.D., ed.
NCADI #M64 NTIS PB #89-131858/AS (A12) $36.50

65 WOMEN AND DRUGS: A NEW ERA FOR RESEARCH. Barbara A. Ray, Ph.D., and Monique C. Braude, Ph.D., eds.
NCADI #M65 NTIS PB #89-130637/AS (A06) $27.00

69 OPIOID PEPTIDES: MEDICINAL CHEMISTRY. Rao S. Rapaka, Ph.D.; Gene Barnett, Ph.D.; and Richard L. Hawks, Ph.D., eds.
NCADI #M69 NTIS PB #89-158422/AS (A17) $44.50

70 OPIOID PEPTIDES: MOLECULAR PHARMACOLOGY, BIOSYNTHESIS, AND ANALYSIS. Rao S. Rapaka, Ph.D., and Richard L. Hawks, Ph.D., eds.
NCADI #M70 NTIS PB #89-158430/AS (A18) $52.00

72 RELAPSE AND RECOVERY IN DRUG ABUSE. Frank M. Tims, Ph.D., and Carl G. Leukefeld, D.S.W., eds.
NCADI #M72 NTIS PB #89-151963/AS (A09) $36.50

74 NEUROBIOLOGY OF BEHAVIORAL CONTROL IN DRUG ABUSE. Stephen I. Szara, M.D., D.Sc., ed.
NCADI #M74 NTIS PB #89-151989/AS (A07) $27.00

77 ADOLESCENT DRUG ABUSE: ANALYSES OF TREATMENT RESEARCH. Elizabeth R. Rahdert, Ph.D., and John Grabowski, Ph.D., eds.
NCADI #M77 NTIS PB #89-125488/AS (A0) $27.00

78 THE ROLE OF NEUROPLASTICITY IN THE RESPONSE TO DRUGS. David P. Friedman, Ph.D., and Doris H. Clouet, Ph.D., eds.
NCADI #M78 NTIS PB #88-245683/AS (A10) $36.50

79 STRUCTURE-ACTIVITY RELATIONSHIPS OF THE CANNABINOIDS. Rao S. Rapaka, Ph.D., and Alexandros Makriyannis, Ph.D., eds.
NCADI #M79 NTIS PB #89-109201/AS (A10) $36.50

80 NEEDLE SHARING AMONG INTRAVENOUS DRUG ABUSERS: NATIONAL AND INTERNATIONAL PERSPECTIVES. Robert J. Battjes, D.S.W., and Roy W. Pickens, Ph.D., eds.
NCADI #M80 NTIS PB #88-236138/AS (A09) $36.50

82 OPIOIDS IN THE HIPPOCAMPUS. Jacqueline F. McGinty, Ph.D., and David P. Friedman, Ph.D., eds.
NCADI #M82 NTIS PB #88-245691/AS (A06) $27.00

83 HEALTH HAZARDS OF NITRITE INHALANTS. Harry W. Haverkos, M.D., and John A. Dougherty, Ph.D., eds.
NCADI #M83 NTIS PB #89-125496/AS (A06) $27.00

84 LEARNING FACTORS IN SUBSTANCE ABUSE. Barbara A. Ray, Ph.D., ed.
NCADI #M84 NTIS PB #89-125504/AS (A10) $36.50

85 EPIDEMIOLOGY OF INHALANT ABUSE: AN UPDATE. Raquel A. Crider, Ph.D., and Beatrice A. Rouse, Ph.D., eds.
NCADI #M85 NTIS PB #89-123178/AS (A10) $36.50

86 COMPULSORY TREATMENT OF DRUG ABUSE: RESEARCH AND CLINICAL PRACTICE. Carl G. Leukefeld, D.S.W., and Frank M. Tims, Ph.D., eds.
NCADI #M86 NTIS PB #89-151997/AS (A12) $36.50

87 OPIOID PEPTIDES: AN UPDATE. Rao S. Rapaka, Ph.D., and Bhola N. Dhawan, M.D., eds.
NCADI #M87 NTIS PB #89-158430/AS (A11) $36.50

88 MECHANISMS OF COCAINE ABUSE AND TOXICITY. Doris H. Clouet, Ph.D.; Khursheed Asghar, Ph.D.; and Roger M. Brown, Ph.D., eds.
NCADI #M88 NTIS PB #89-125512/AS (A16) $44.50

89 BIOLOGICAL VULNERABILITY TO DRUG ABUSE. Roy W. Pickens, Ph.D., and Dace S. Svikis, B.A., eds.
NCADI #M89 NTIS PB #89-125520/AS (A09) $27.00

92 TESTING FOR ABUSE LIABILITY OF DRUGS IN HUMANS. Marian W. Fischman, Ph.D., and Nancy K. Mello, Ph.D., eds.
NCADI #M92 NTIS PB #90-148933/AS (A17) $44.50

93 AIDS AND INTRAVENOUS DRUG USE: FUTURE DIRECTIONS FOR COMMUNITY-BASED PREVENTION RESEARCH. Carl G. Leukefeld, D.S.W.; Robert J. Battjes, D.S.W.; and Zili Amsel, D.S.C., eds.
NCADI #M93 NTIS PB #90-148933/AS (A14) $44.50

94 PHARMACOLOGY AND TOXICOLOGY OF AMPHETAMINE AND RELATED DESIGNER DRUGS. Khursheed Asghar, Ph.D., and Errol De Souza, Ph.D., eds.
NCADI #M94 NTIS PB #90-148958/AS (A16) $44.50

95 PROBLEMS OF DRUG DEPENDENCE, 1989. PROCEEDINGS OF THE 51st ANNUAL SCIENTIFIC MEETING. THE COMMITTEE ON PROBLEMS OF DRUG DEPENDENCE, INC.
Louis S. Harris, Ph.D., ed.
NCADI #M95 NTIS PB #90-237660/AS (A99) $67.00

96 DRUGS OF ABUSE: CHEMISTRY, PHARMACOLOGY, IMMUNOLOGY, AND AIDS. Phuong Thi Kim Pham, Ph.D., and Kenner Rice, Ph.D., eds.
NCADI #M96 NTIS PB #90-237678/AS (A11) $36.50

97 NEUROBIOLOGY OF DRUG ABUSE: LEARNING AND MEMORY. Lynda Erinoff, Ph.D., ed.
NCADI #M97 NTIS PB #90-237686/AS (A11) $36.50

98 THE COLLECTION AND INTERPRETATION OF DATA FROM HIDDEN POPULATIONS.
Elizabeth Y. Lambert, M.S., ed.
NCADI #M98 NTIS PB #90-237694/AS (A08) $27.00

99 RESEARCH FINDINGS ON SMOKING OF ABUSED SUBSTANCES. C. Nora Chiang, Ph.D., and Richard L. Hawks, Ph.D., eds.
NCADI #M99 NTIS PB #91-141119 (A09) $27.00

100 DRUGS IN THE WORKPLACE: RESEARCH AND EVALUATION DATA. VOL II. Steven W. Gust, Ph.D.; J. Michael Walsh, Ph.D.; Linda B. Thomas, B.S.; and Dennis J. Crouch, M.B.A., eds.
NCADI #M100 GPO Stock #017-024-01458-3 $8.00

101 RESIDUAL EFFECTS OF ABUSED DRUGS ON BEHAVIOR.
John W. Spencer, Ph.D., and John J. Boren, Ph.D., eds.
NCADI #M101 NTIS PB #91-172858/AS (A09) $27.00

102 ANABOLIC STEROID ABUSE. Geraline C. Lin, Ph.D., and Lynda Erinoff, Ph.D., eds.
NCADI #M102 NTIS PB #91-172866/AS (A11) $36.50

103 DRUGS AND VIOLENCE: CAUSES, CORRELATES, AND CONSEQUENCES. Mario De La Rosa, Ph.D.; Elizabeth Y. Lambert, M.S.; and Bernard Gropper, Ph.D., eds.
NCADI #M103 NTIS PB #91-172874/AS (A13) $36.50

104 PSYCHOTHERAPY AND COUNSELING IN THE TREATMENT OF DRUG ABUSE. Lisa Simon Onken, Ph.D., and Jack D. Blaine, M.D., eds.
NCADI #M104 NTIS PB #91-172874/AS (A07) $27.00

106 IMPROVING DRUG ABUSE TREATMENT.
Roy W. Pickens, Ph.D.; Carl G. Leukefeld, D.S.W.; and Charles R. Schuster, Ph.D., eds.
NCADI #M106 NTIS PB #92-105873(A18) $50.00

107 DRUG ABUSE PREVENTION INTERVENTION RESEARCH: METHODOLOGICAL ISSUES. Carl G. Leukefeld, D.S.W., and William J. Bukoski, Ph.D., eds.
NCADI #M107 NTIS PB #92-160985 (A13) $36.50

108 CARDIOVASCULAR TOXICITY OF COCAINE: UNDERLYING MECHANISMS. Pushpa V. Thadani, Ph.D., ed.
NCADI #M108 NTIS PB #92-106608 (A11) $36.50

109 LONGITUDINAL STUDIES OF HIV INFECTION IN INTRAVENOUS DRUG USERS: METHODOLOGICAL ISSUES IN NATURAL HISTORY RESEARCH. Peter Hartsock, Dr.P.H., and Sander G. Genser, M.D., M.P.H., eds.
NCADI #M109 NTIS PB #92-106616 (A08) $27.00

111 MOLECULAR APPROACHES TO DRUG ABUSE RESEARCH: RECEPTOR CLONING, NEUROTRANSMITTER EXPRESSION, AND MOLECULAR GENETICS: VOLUME I. Theresa N.H. Lee, Ph.D., ed.
NCADI #M111 NTIS PB #92-135743 (A10) $36.50

112 EMERGING TECHNOLOGIES AND NEW DIRECTIONS IN DRUG ABUSE RESEARCH. Rao S. Rapaka, Ph.D.; Alexandros Makriyannis, Ph.D.; and Michael J. Kuhar, Ph.D., eds.
NCADI #M112 NTIS PB #92-155449 (A15) $44.50

113 ECONOMIC COSTS, COST EFFECTIVENESS, FINANCING, AND COMMUNITY-BASED DRUG TREATMENT. William S. Cartwright, Ph.D., and James M. Kaple, Ph.D., eds.
NCADI #M113 NTIS PB #92-155795 (A10) $36.50

114 METHODOLOGICAL ISSUES IN CONTROLLED STUDIES ON EFFECTS OF PRENATAL EXPOSURE TO DRUG ABUSE.

M. Marlyne Kilbey, Ph.D., and Khursheed Asghar, Ph.D., eds.
NCADI #M114 NTIS PB #92-146216 (A16) $44.50

115 METHAMPHETAMINE ABUSE: EPIDEMIOLOGIC ISSUES AND IMPLICATIONS. Marissa A. Miller, D.V.M., M.P.H., and Nicholas J. Kozel, M.S., eds.
NCADI #M115 NTIS PB #92-146224/ll (AO7) $27.00

116 DRUG DISCRIMINATION: APPLICATIONS TO DRUG ABUSE RESEARCH. R.A. Glennon, Ph.D.; Toubjörn U.C. Järbe, Ph.D.; and J. Frankenheim, Ph.D., eds.
NCADI #M116 NTIS PB #94-169471 (A20) $52.00

117 METHODOLOGICAL ISSUES IN EPIDEMIOLOGY, PREVENTION, AND TREATMENT RESEARCH ON DRUG-EXPOSED WOMEN AND THEIR CHILDREN.
M. Marlyve Kilbey, Ph.D., and Kursheed Asghar, Ph.D., eds.
GPO Stock #O17-024-01472-9 $12.00
NCADI #M117 NTIS PB #93-102101/LL (A18) $52.00

118 DRUG ABUSE TREATMENT IN PRISONS AND JAILS.
Carl G. Leukefeld, D.S.W., and Frank M. Tims, Ph.D., eds.
GPO Stock #O17-024-01473-7 $16.00
NCADI #M118 NTIS PB #93-102143/LL (A14) $44.50

120 BIOAVAILABILITY OF DRUGS TO THE BRAIN AND THE BLOOD-BRAIN BARRIER. Jerry Frankenheim, Ph.D., and Roger M. Brown, Ph.D., eds.
GPO Stock #017-024-01481-8 $10.00
NCADI #M120 NTIS PB #92-214956/LL (A12) $36.50

121 BUPRENORPHINE: AN ALTERNATIVE TREATMENT FOR OPIOID DEPENDENCE. Jack D. Blaine, Ph.D., ed.
GPO Stock #017-024-01482-6 $5.00

NCADI #M121 NTIS PB #93-129781/LL (A08) $27.00

123 ACUTE COCAINE INTOXICATION: CURRENT METHODS OF TREATMENT. Heinz Sorer, Ph.D., ed.
GPO Stock #017-024-01501-6 $6.50
NCADI #M123 NTIS PB #94-115433/LL (A09) $27.00

124 NEUROBIOLOGICAL APPROACHES TO BRAIN-BEHAVIOR INTERACTION. Roger M. Brown, Ph.D., and Joseph Fracella, Ph.D., eds.
GPO Stock #017-024-01492-3 $9.00
NCADI #M124 NTIS PB #93-203834/LL (A12) $36.50

125 ACTIVATION OF IMMEDIATE EARLY GENES BY DRUGS OF ABUSE. Reinhard Grzanna, Ph.D., and Roger M. Brown, Ph.D., eds.
GPO Stock #017-024-01503-2 $7.50
NCADI #M125 NTIS PB #94-169489 (A12) $36.50

126 MOLECULAR APPROACHES TO DRUG ABUSE RESEARCH VOLUME II: STRUCTURE, FUNCTION, AND EXPRESSION. Theresa N.H. Lee, Ph.D., ed.
NCADI #M126 NTIS PB #94-169497 (A08) $27.00

127 PROGRESS AND ISSUES IN CASE MANAGEMENT. Rebecca S. Ashery, D.S.W., ed.
NCADI #M127 NTIS PB #94-169505 (A18) $52.00

128 STATISTICAL ISSUES IN CLINICAL TRIALS FOR TREATMENT OF OPIATE DEPENDENCE. Ram B. Jain, Ph.D., ed.
NCADI #M128 NTIS PB #93-203826/LL (A09) $27.00

129 INHALANT ABUSE: A VOLATILE RESEARCH AGENDA. Charles W. Sharp, Ph.D.; Fred Beauvais, Ph.D.; and Richard Spence, Ph.D., eds.
GPO Stock #017-024-01496-6 $12.00
NCADI #M129 NTIS PB #93-183119/LL (A15) $44.50

130 DRUG ABUSE AMONG MINORITY YOUTH: ADVANCES IN RESEARCH AND METHODOLOGY. Mario De La Rosa, Ph.D., and Juan-Luis Recio Adrados, Ph.D., eds.
GPO Stock #017-024-01506-7 $14.00
NCADI #M130 NTIS PB #94-169513 (A15) $44.50

131 IMPACT OF PRESCRIPTION DRUG DIVERSION CONTROL SYSTEMS ON MEDICAL PRACTICE AND PATIENT CARE. James R. Cooper, Ph.D.; Dorynne J. Czechowicz, M.D.; Stephen P. Molinari, J.D., R.Ph.; and Robert C. Peterson, Ph.D., eds.
GPO Stock #017-024-01505-9 $14.00
NCADI #M131 NTIS PB #94-169521 (A15) $44.50

132 PROBLEMS OF DRUG DEPENDENCE, 1992: PROCEEDINGS OF THE 54TH ANNUAL SCIENTIFIC MEETING OF THE COLLEGE ON PROBLEMS OF DRUG DEPENDENCE. Louis Harris, Ph.D., ed.
GPO Stock #017-024-01502-4 $23.00
NCADI #M132 NTIS PB #94-115508/LL (A99)

133 SIGMA, PCP, AND NMDA RECEPTORS. Errol B. De Souza, Ph.D.; Doris Clouet, Ph.D., and Edythe D. London, Ph.D., eds.
NCADI #M133 NTIS PB #94-169539 (A12) $36.50

134 MEDICATIONS DEVELOPMENT: DRUG DISCOVERY, DATABASES, AND COMPUTER-AIDED DRUG DESIGN. Rao S. Rapaka, Ph.D., and Richard L. Hawks, Ph.D., eds.
GPO Stock #017-024-01511-3 $11.00
NCADI #M134 NTIS PB #94-169547 (A14) $44.50

135 COCAINE TREATMENT: RESEARCH AND CLINICAL PERSPECTIVES. Frank M. Tims, Ph.D., and Carl G. Leukefeld, D.S.W., eds.
GPO Stock #017-024-01520-2 $11.00
NCADI #M135 NTIS PB #94-169554 (A13) $36.50

136 ASSESSING NEUROTOXICITY OF DRUGS OF ABUSE. Lynda Erinoff, Ph.D., ed.
GPO Stock #017-024-01518-1 $11.00
NCADI #M136 NTIS PB #94-169562 (A13) $36.50

137 BEHAVIORAL TREATMENTS FOR DRUG ABUSE AND DEPENDENCE. Lisa Simon Onken, Ph.D.; Jack D. Blaine, M.D.; and John J. Boren, Ph.D., eds.
GPO Stock #017-024-01519-9 $13.00
NCADI #M137 NTIS PB #94-169570 (A15) $44.50

138 IMAGING TECHNIQUES IN MEDICATIONS DEVELOPMENT: CLINICAL AND PRECLINICAL ASPECTS. Heinz Sorer, Ph.D., and Rao S. Rapaka, Ph.D., eds.
NCADI #M138

139 SCIENTIFIC METHODS FOR PREVENTION INTERVENTION RESEARCH. Arturo Cazares, M.D., M.P.H., and Lula A. Beatty, Ph.D., eds.
NCADI #M139

140 PROBLEMS OF DRUG DEPENDENCE, 1993: PROCEEDINGS OF THE 55TH ANNUAL SCIENTIFIC MEETING, THE COLLEGE ON PROBLEMS OF DRUG DEPENDENCE, INC. VOLUME I: PLENARY SESSION SYMPOSIA AND ANNUAL REPORTS. Louis S. Harris, Ph.D., ed.
NCADI #M140

141 PROBLEMS OF DRUG DEPENDENCE, 1993: PROCEEDINGS OF THE 55TH ANNUAL SCIENTIFIC MEETING, THE COLLEGE ON PROBLEMS OF DRUG DEPENDENCE, INC. VOLUME II: ABSTRACTS. Louis S. Harris, Ph.D., ed.
NCADI #M141

142 ADVANCES IN DATA ANALYSIS FOR PREVENTION INTERVENTION RESEARCH. Linda M. Collins, Ph.D., and Larry A. Seitz, Ph.D., eds.
NCADI #M142

143 THE CONTEXT OF HIV RISK AMONG DRUG USERS AND THEIR SEXUAL PARTNERS. Robert J. Battjes, D.S.W.; Zili Sloboda, Sc.D.; and William C. Grace, Ph.D., eds.
NCADI #M143

144 THERAPEUTIC COMMUNITY: ADVANCES IN RESEARCH AND APPLICATION. Frank M. Tims, Ph.D.; George De Leon, Ph.D.; and Nancy Jainchill, Ph.D., eds.
NCADI #M144

145 NEUROBIOLOGICAL MODELS FOR EVALUATING MECHANISMS UNDERLYING COCAINE ADDICTION. Lynda Erinoff, Ph.D., and Roger M. Brown, Ph.D., eds.
NCADI #M145

146 HALLUCINOGENS: AN UPDATE. Geraline C. Lin, Ph.D., and Richard A. Glennon, Ph.D., eds.
NCADI #M146

147 DISCOVERY OF NOVEL OPIOID MEDICATIONS. Rao S. Rapaka, Ph.D., and Heinz Sorer, Ph.D., eds.
NCADI #M147

148 EPIDEMIOLOGY OF INHALANT ABUSE: AN INTERNATIONAL PERSPECTIVE. Nicholas J. Kozel, M.S.; Zili Sloboda, Sc.D.; and Mario R. De La Rosa, Ph.D., eds.
NCADI #M148

149 MEDICATIONS DEVELOPMENT FOR THE TREATMENT OF PREGNANT ADDICTS AND THEIR INFANTS. C. Nora Chiang, Ph.D., and Loretta P. Finnegan, M.D., eds.
NCADI #M149

150 INTEGRATING BEHAVIORAL THERAPIES WITH MEDICATIONS IN THE TREATMENT OF DRUG DEPENDENCE. Lisa Simon Onken, Ph.D.; Jack D. Blaine, M.D.; and John J. Boren, Ph.D., eds.
NCADI #M150

151 SOCIAL NETWORKS, DRUG ABUSE, AND HIV TRANSMISSION. Richard H. Needle, Ph.D., M.P.H.; Susan L. Coyle, Ph.D.; Sander G. Genser, M.D., M.P.H.; and Robert T. Trotter II, Ph.D., eds.
NCADI #M151

152 PROBLEMS OF DRUG DEPENDENCE 1994: PROCEEDINGS OF THE 56TH ANNUAL SCIENTIFIC MEETING, THE COLLEGE ON PROBLEMS OF DRUG DEPENDENCE, INC. VOLUME I: PLENARY SESSION SYMPOSIA AND ANNUAL REPORTS. Louis S. Harris, Ph.D., ed.
NCADI #M152

153 PROBLEMS OF DRUG DEPENDENCE 1994: PROCEEDINGS OF THE 56TH ANNUAL SCIENTIFIC MEETING, THE COLLEGE ON PROBLEMS OF DRUG DEPENDENCE, INC. VOLUME II: ABSTRACTS. (1995) Louis S. Harris, Ph.D., ed.
NCADI #M153 GPO Stock #017-024-01564-4 $22.00

154 MEMBRANES AND BARRIERS: TARGETED DRUG DELIVERY. (1995) Rao S. Rapaka, Ph.D., ed.
NCADI #M154 GPO Stock #017-024-01583-1 $10.00

155 REVIEWING THE BEHAVIORAL SCIENCE KNOWLEDGE BASE ON TECHNOLOGY TRANSFER. (1995)
Thomas E. Backer, Ph.D.; Susan L. David; and Gerald Soucy, Ph.D., eds.
NCADI #M155 GPO Stock #017-024-01581-4 $12.00

156 ADOLESCENT DRUG ABUSE: CLINICAL ASSESSMENT AND THERAPEUTIC INTERVENTIONS. (1995)
Elizabeth Rahdert, Ph.D.; Zili Sloboda, Ph.D.; and Dorynne Czechowicz, M.D., eds.
NCADI #M156 GPO Stock #017-024-01585-7 $14.00

157 QUALITATIVE METHODS IN DRUG ABUSE AND HIV RESEARCH. (1995)
Elizabeth Y. Lambert, M.S.; Rebecca S. Ashery, D.S.W.; and Richard H. Needle, Ph.D., M.P.H., eds.
NCADI #M157 GPO Stock #017-024-01581-4

158 BIOLOGICAL MECHANISMS AND PERINATAL EXPOSURE TO DRUGS. (1995) Pushpa V. Thadani, Ph.D., ed.
NCADI #M158 GPO Stock #017-024-01584-9

159 INDIVIDUAL DIFFERENCES IN THE BIOBEHAVIORAL ETIOLOGY OF DRUG ABUSE. (1996) Harold W. Gordon, Ph.D., and Meyer D. Glantz, Ph.D., eds.
NCADI #M159

161 MOLECULAR APPROACHES TO DRUG ABUSE RESEARCH. VOLUME III: RECENT ADVANCES AND EMERGING STRATEGIES. (1996) Theresa N.H. Lee, Ph.D., ed.
NCADI #M161

162 PROBLEMS OF DRUG DEPENDENCE, 1995. PROCEEDINGS FROM THE 57TH ANNUAL SCIENTIFIC MEETING OF THE COLLEGE ON DRUG DEPENDENCE, INC. (1996)
Louis Harris, Ph.D., ed.
NCADI #M162

163 NEUROTOXICITY AND NEUROPATHOLOGY ASSOCIATED WITH COCAINE/STIMULANT ABUSE. (1996)
Dorota Majewska, Ph.D., ed.
NCADI #M163

164 BEHAVIORAL STUDIES OF DRUG-EXPOSED OFFSPRING: METHODOLOGICAL ISSUES IN HUMAN AND ANIMAL RESEARCH. (1996) Cora Lee Wetherington, Ph.D.; Vincent L. Smeriglio, Ph.D.; and Loretta P. Finnegan, Ph.D., eds.
NCADI #M164

165 BEYOND THE THERAPEUTIC ALLIANCE: KEEPING THE DRUG-DEPENDENT INDIVIDUAL IN TREATMENT. (1996)
Lisa Simon Onken, Ph.D.; Jack D. Blaine, M.D.; and John J. Boren, Ph.D.
NCADI #M165

166 TREATMENT FOR DRUG-EXPOSED WOMEN AND CHILDREN: ADVANCES IN RESEARCH METHODOLOGY. (1996) Elizabeth Rahdert, Ph.D., ed.
NCADI #M166

* U.S. GOVERNMENT PRINTING OFFICE: 1997-418-681/40004

NIH Publication No. 97-4147
Printed April, 1997