Redesigning an Ongoing National Household Survey:

Methodological Issues

JOSEPH GFROERER, JOE EYERMAN, and JAMES CHROMY

EDITORS

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

Gfroerer, J., Eyerman, J., and Chromy, J., Eds. (2002). *Redesigning an ongoing national household survey: Methodological issues*. DHHS Publication No. SMA 03-3768. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Obtaining Additional Copies of Publication

This publication can be accessed electronically through the Internet at: http://www.DrugAbuseStatistics.SAMHSA.gov. Copies may also be obtained, free of charge, from the National Clearinghouse for Alcohol and Drug Information (NCADI), a service of SAMHSA. Write or call NCADI at:

National Clearinghouse for Alcohol and Drug Information
P.O. Box 2345, Rockville, MD 20847-2345
1-301-468-2600, 1-800-729-6686, TDD 1-800-487-4889

Library of Congress Cataloging-in-Publication Data

Redesigning an ongoing national household survey: methodological issues / Joseph Gfroerer, Joe Eyerman, and James Chromy, editors.
p. cm. -- (DHHS Publication; no. SMA 03-3768)
Includes bibliographical references.
1. National Household Survey on Drug Abuse (U.S.) 2. Drug abuse surveys--Methodology. 3. Drug abuse--United States--Statistics. 4. Substance abuse--United States--Statistics. I. Gfroerer, Joseph C., 1954- II. Eyerman, Joe. III. Chromy, James R. IV. United States. Substance Abuse and Mental Health Services Administration. Office of Applied Studies. V. Series.
HV5825 .R4154 2002
362.29′0973′021--dc21

2002042816

Contents

Authors

PEGGY BARKER, Substance Abuse and Mental Health Services Administration

DICY BUTLER, Substance Abuse and Mental Health Services Administration

RACHEL CASPAR, RTI International

JAMES CHROMY, RTI International

TERESA DAVIS, RTI International

JOE EYERMAN, RTI International

RALPH FOLSOM, JR, RTI International

JOSEPH GFROERER, Substance Abuse and Mental Health Services Administration

KATHERINE GIACOLETTI, RTI International

ERIC GRAU, RTI International

ARTHUR HUGHES, Substance Abuse and Mental Health Services Administration

LARRY KROUTIL, RTI International

LAWRENCE MYERS, RTI International

DAWN ODOM, RTI International

LISA PACKER, RTI International

MICHAEL PENNE, RTI International

LANNY PIPER, RTI International

AVINASH SINGH, RTI International

DOUG WRIGHT, Substance Abuse and Mental Health Services Administration

SHIYING WU, RTI International

1

Introduction

Joseph Gfroerer

The National Household Survey on Drug Abuse (NHSDA) is the Federal Government's primary source of information on the magnitude of substance use and abuse in the United States. Conducted since 1971, the survey collects data by administering questionnaires to a representative sample of persons aged 12 or older at their places of residence. Data from the survey are used extensively by policymakers and researchers to measure the prevalence and correlates of licit and illicit substance use, to identify and monitor trends in substance use, and to analyze differences in substance use patterns by population subgroups.

In 1999, a major redesign of the NHSDA was implemented involving both the sample design and the data collection method of the survey. The strictly national design was changed to a much larger, State-based design to meet the needs of policymakers for estimates of substance use prevalence for each State. The data collection method was changed from a paper-and-pencil interview (PAPI) method to a computer-assisted interview (CAI) method, primarily to improve the quality of NHSDA estimates. Implementation of these significant changes posed a number of difficult challenges involving a variety of methodological issues. These issues cover many aspects of the survey, including the management of fieldwork, the processing of data, and the reporting of results. This publication discusses several of the most critical issues encountered and describes how the research team that conducts the survey addressed them.

This report has two purposes. First, it provides information on the impact of the redesign on the estimates produced from the NHSDA. Researchers and other users of NHSDA data will find this information helpful in interpreting NHSDA estimates, particularly if they are interested in comparing data from the new design with data from the old design. The second purpose is to present research findings of interest to survey methodologists involved in designing and conducting surveys of all types, not just surveys of substance abuse. Although these findings taken as a whole could be considered a case study in the redesign of a major ongoing survey, several of the chapters in this report present important research findings that are applicable to many types of surveys.

1.1 Background and History of the NHSDA

The NHSDA is funded and directed by the Substance Abuse and Mental Health Services Administration (SAMHSA), with data collection done under contract. Since 1988, Research Triangle Institute (RTI) has been contracted to conduct the survey.

Since 1971, the NHSDA has undergone a variety of changes in its sample design as data priorities have changed. During the 1970s and 1980s, the NHSDA was a relatively small, periodic survey. Conducted every 2 or 3 years, the sample size grew gradually from about 3,000 respondents per year in the early 1970s to 8,814 in 1988. In the late 1980s, the Nation′s substance abuse problem became a major concern of the public and of politicians. As a result, Congress passed new legislation that increased funding for substance abuse data collection and research and created the White House Office of National Drug Control Policy (ONDCP). ONDCP began producing annual National Strategies that used NHSDA data extensively in setting goals and tracking the progress of substance abuse policies and programs. With the increase in funds and greater reliance on NHSDA data by policymakers and researchers, annual fielding of the survey began in 1990, and a significant expansion of the sample began in 1991. The basic national sample size throughout the 1990s was about 18,000 respondents per year.

Throughout the NHSDA′s history, interest in particular subpopulations has led to sample design changes and augmentations. Rural areas were oversampled in 1979 and 1994, and the survey oversampled blacks and Hispanics from 1985 through 1998. Supplemental samples of specific metropolitan areas were included from 1990 through 1993, and supplemental samples in two States, California and Arizona, were added in 1997 and 1998.

Changes in the survey methodology prior to 1999 were infrequent and relatively minor. In general, the NHSDA used the same data collection methodology from 1971 through 1998: a confidential, anonymous, face-to-face interview conducted in households and employing self-administration of sensitive substance use items. However, some small but important changes were made in the survey procedures that affected NHSDA estimates. In 1982, questions on nonmedical use of psychotherapeutic drugs were converted from interviewer-administered to self-administered. Similarly, tobacco questions were shifted to self-administration in 1994. Machine editing procedures were incorporated into the NHSDA data processing for the first time in 1988. In 1994, following extensive research (Turner, Lessler, & Gfroerer, 1992), the NHSDA questionnaire and editing procedures were modified to provide more reliable substance use prevalence estimates (SAMHSA, 1996a).

1.2 Redesign of the NHSDA in 1999

Methodological research has demonstrated the benefits of audio-computer-assisted self-interviewing (ACASI) in collecting data on sensitive behaviors such as substance use in household surveys (O'Reilly, Hubbard, Lessler, Biemer, & Turner, 1994; Lessler, Weeks, & O'Reilly, 1994; Turner, Ku, Sorenstein, & Pleck, 1996; Duffer, Lessler, Weeks, & Mosher, 1996; Tourangeau & Smith, 1996; Turner, Ku, Rogers, Lindberg, Pleck, & Sonenstein, 1998). Most important, these studies indicate that respondents are more willing to report sensitive behaviors with ACASI than with other modes of data collection. Based on this research, SAMHSA decided in 1995 to initiate development and testing of CAI, including ACASI, in the NHSDA. The development of the CAI instrument was accomplished primarily under a contract awarded to RTI in early 1996. The testing protocol included a small (n = 400) initial field test in the fourth quarter of 1996, cognitive laboratory testing, a second larger (n = 1,982) field test in the fourth quarter of 1997, and a final pretest conducted in August 1998. A detailed report on the development of the CAI instrument for the NHSDA was recently published (SAMHSA, 2001).

At the same time that the new NHSDA CAI was being developed, a long-standing interest in State-level substance use prevalence data was culminating in legislation that would result in the redesign of the NHSDA sample. With the passage in 1996 of voter initiatives legalizing marijuana use for medical purposes in California and Arizona, and the substantial role of Federal block grant funds given to States for substance abuse prevention and treatment, Congress and the Clinton administration concluded it would be useful to have these State-level estimates. The House Appropriations

Committee Report accompanying the Department of Health and Human Services (DHHS) FY1997 appropriations bill called for SAMHSA to expand the NHSDA and use the new State-level data to improve the provision of treatment and prevention services in States with high levels of substance abuse. Based on the level of funding allocated by the legislation, this required the ability to make estimates for most States using only modest sample sizes in conjunction with small area estimation modeling techniques. SAMHSA determined that this was feasible based on the results of earlier modeling for selected States using the 1991–1993 NHSDA, as well as a sampling plan that would facilitate estimation for all States (SAMHSA, 1996b, 1997).

Although these two major redesign plans evolved essentially independently and for different purposes, SAMHSA ultimately decided that the most effective way to incorporate both these changes into the NHSDA was to implement them at the same time. Initial plans had called for the sample redesign in 1999 and a phase-in of the CAI implementation during 2000 and 2001. However, legislation proposed in the spring of 1998 required precise measurement of future trends in the prevalence of youth cigarette use beginning with the 1999 NHSDA. Because the CAI development and testing had progressed well at this point, SAMHSA decided to accelerate implementation to meet the requirements of the proposed legislation (which actually was never enacted). Thus, the expanded, State-based sample design and the new CAI instrument were both implemented in January 1999.

1.3 Content of this Report

Chapter 2, prepared by Doug Wright, Peggy Barker, Joseph Gfroerer, and Lanny Piper, contains a description of the 1999 redesign. The chapter describes the new State-based sample design, contrasting it with the national sample employed from 1971 through 1998. It discusses the change in the household screening procedure in 1999, moving from a complex paper screening form to the hand-held Apple Newton computer. The chapter also summarizes the development of the new CAI instrument and explains the differences between the PAPI and CAI instruments. The design of the supplemental PAPI sample is described. ***Chapter 2*** closes with a discussion of the difficulties faced by the NHSDA managers in expanding the field staff in 1999, increasing the number of interviewers from about 300 to about 1,200.

The remaining chapters discuss specific methodological issues confronted during the implementation of the new NHSDA design. In

Chapter 3, Joe Eyerman, Dawn Odom, Shiying Wu, and Dicy Butler report on the unexpected drop in response rates in conjunction with the 1999 NHSDA redesign. They analyze several factors that could have caused the drop and conclude that the large number of inexperienced interviewers was a major reason; however, this factor does not fully explain the lower response rate in 1999. The results also indicate that the response rate was slightly higher with CAI than with PAPI. The switch from a paper to an electronic household screening procedure is shown to have reduced sample bias by removing interviewer effects from the screening routine, although this had a small negative effect on the response rate.

In ***Chapter 4***, Rachel Caspar and Michael Penne analyze the performance of the CAI instrument and demonstrate improvements in data quality. The chapter shows that the level of missing data for substance use items in the NHSDA has been reduced with the CAI instrument through the use of data checks and respondent probing. Data are also presented showing that interviewers perceive that respondents have less difficulty completing the CAI than the PAPI interview, and that CAI offers a more private interview setting.

The new questionnaire required development of new editing rules for the CAI substance use data, described by Larry Kroutil and Lawrence Myers in ***Chapter 5***. They discuss analyses done on the first 6 months of data from the 1999 NHSDA to assess several alternative editing rules. Based on these analyses, a "flag and impute" rule was adopted for editing key substance use measures. This new rule is less likely than the previous editing rules used in the PAPI to assign particular values to substance use variables when respondents provide inconsistent information. With the new editing rules, inconsistent responses are generally assigned values that require further imputation based on statistical models. Overall, the new instrument and editing rules greatly reduce the impact of editing on NHSDA substance use prevalence estimates.

With greater emphasis on statistical imputation relative to editing in the redesigned NHSDA, an improved imputation procedure for NHSDA substance use variables was also developed. This new method, called Predictive Mean Neighborhood (PMN) imputation, is described by Avinash Singh, Eric Grau, and Ralph Folsom, Jr., in ***Chapter 6***. PMN uses modeling to identify respondents with data who are "similar" to respondents without data and then randomly chooses one of them to be a "donor" of substance use variables to the case with missing data. The chapter points out several advantages of PMN over the unweighted hot deck method used in past

NHSDAs. These include the greater ability to use covariates to determine donors, the easy incorporation of sampling weights into the models, and reduced bias due to the more controlled random selection of "similar" donors.

Of great interest to survey methods researchers as well as policymakers and analysts who have been using NHSDA data to monitor trends in substance use is the impact of the change in data collection mode on NHSDA estimates. Although the dual sample in 1999 was not designed as an experiment to specifically test the impact of interview mode (CAI vs. PAPI) on the reporting of substance use behaviors, the large sample provides an opportunity to add considerable knowledge on this important topic. In ***Chapter*** *7*, James Chromy, Teresa Davis, Lisa Packer, and Joseph Gfroerer present an analysis of the 1999 CAI and PAPI substance use prevalence estimates. They show that the CAI tended to produce higher reporting of substance use behaviors, based on matched samples and comparable questions asked in the two modes. However, editing and imputation had a greater impact on the PAPI estimates, so comparisons between final PAPI and CAI estimates showed mixed results depending on the samples compared, drug type, and age group. With the simultaneous changes in mode, question wording, editing, and imputation in 1999, it is not possible to fully isolate the effect of mode, but the combined effect of these changes is clearly substantial and prevents valid comparisons of estimates from the redesigned NHSDA with earlier estimates based on the old methods.

In ***Chapter 8***, Arthur Hughes, James Chromy, Katherine Giacoletti, and Dawn Odom report on the surprising discovery that the experience level of field interviewers affected the reporting of drug use by respondents in the 1999 NHSDA. This unfortunate but interesting finding should be of concern to survey researchers. It causes difficulties for NHSDA analysts attempting to understand trends in substance use prevalence. In this chapter, results of a series of regression models isolating the effect of interviewer experience are presented demonstrating that increased experience is associated with lower reporting of substance use by NHSDA respondents. This effect, in conjunction with the large number of new, inexperienced interviewers in 1999, tended to inflate the 1999 PAPI estimates. This led to SAMHSA's decision to revise the analytic weights used for the 1999 PAPI data and to limit the set of estimates released from this dataset. Analysis of CAI data shows a similar, but much smaller, effect for interviewer experience.

Finally, ***Chapter 9***, by Lisa Packer, Dawn Odom, James Chromy, Teresa Davis, and Joseph Gfroerer discusses how the NHSDA redesign facilitated an improvement in the method used to estimate the incidence (i.e., initiation) of drug use. The new CAI instrument includes questions about year and month of first substance use. In addition, statistical imputation of incidence-related variables was not done with PAPI data but is used for CAI data. These data improvements allowed a more accurate computation of incidence rates. This chapter describes the new method of computing rates and compares incidence estimates from the CAI with estimates previously generated from PAPI data.

1.4 References

Duffer, A., Lessler, J., Weeks, M., & Mosher, W. (1996). Impact of incentives and interviewing modes: Results from the National Survey of Family Growth Cycle V Pretest. *Proceedings of the Health Survey Research Methods Conference*, 147-152.

Lessler, J.T., Weeks, M., & O'Reilly, J. (1994). Results from the National Survey of Family Growth Cycle V Pretest. *Proceedings of the Section on Survey Research Methods*, 64-70. Alexandria, VA: American Statistical Association.

O'Reilly, J.M., Hubbard, M.L., Lessler, J.T., Biemer, P.P., & Turner, C.F. (1994). Audio and video computer-assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics*, *10* (2), 197-214.

Substance Abuse and Mental Health Services Administration. (1996a). *The development and implementation of a new data collection instrument for the 1994 NHSDA* (DHHS Pub. No. SMA 96-3084). Rockville, MD: Author.

Substance Abuse and Mental Health Services Administration. (1996b). *Substance abuse in states and metropolitan areas*: *Model based estimates from the 1991-1993 NHSDA summary report* (DHHS Pub. No. SMA 96-3095). Rockville, MD: Author.

Substance Abuse and Mental Health Services Administration. (1997). *Substance abuse in states and metropolitan areas*: *Model based estimates from the 1991-1993 NHSDA methodology report*. Office of Applied Studies, Methodological Series M-1 (DHHS Pub. No. SMA 97-3140). Rockville, MD: Author.

Substance Abuse and Mental Health Services Administration. (2001). *Development of computer-assisted interviewing procedures for the NHSDA*. Office of Applied Studies, Methodological Series M-3 (DHHS Pub. No. SMA 01-3514). Rockville, MD: Author.

Tourangeau, R., & Smith, T.W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60*, 275-304.

Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., & Sonenstein, F.L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, *280*, 867-873.

Turner, C., Ku, L., Sonenstein, F., & Pleck, J. (1996). Impact of ACASI on reporting of male-male sexual contacts: Preliminary results from the 1995 National Survey of Adolescent Males. In *Proceedings of the Health Survey Research Methods Conference,* 171-176.

Turner, C.F., Lessler, J.T., & Gfroerer, J.C., Eds. (1992). *Survey measurement of drug use: Methodological studies* (DHHS Pub. No. ADM 92-1929). Rockville, MD: National Institutes of Health.

2

Summary of NHSDA Design Changes in 1999

Doug Wright, Peggy Barker, Joseph Gfroerer, and Lanny Piper

An entirely new sample design and a state-of-the-art data collection methodology were implemented with the 1999 National Household Survey on Drug Abuse (NHSDA). The sample design changed from a national, stratified, multistage, area probability sample to a 50-State design, with independent stratified, multistage, area probability samples selected in each State. The sample size increased from about 25,500 interviews in 1998 to about 67,000 interviews in 1999. For the first time in NHSDA history, the 1999 survey administered the interview using computer-assisted interviewing (CAI) technology exclusively, including both computer-assisted personal interviewing (CAPI) and audio computer-assisted self-interviewing (ACASI).

Because this new methodology was being implemented, an additional national sample was selected, and about 14,000 interviews administered using the previous paper-and-pencil interviewing (PAPI) methodology. Together, this PAPI sample and the CAI sample served three purposes. First, the PAPI samples for 1998 and 1999 provided a way to continue to measure the trend in substance use for that period. Second, with both representative samples for 1999, the effect of the change in data collection from PAPI to CAI could be measured without being confounded with the measurement of trend. Third, with a measurement of the impact of the switch to the CAI methodology, estimates for 1998 and earlier years could be adjusted to be comparable to CAI estimates for 1999 and later so that long-term trends in substance use could be estimated. The CAI and PAPI samples for 1999 together resulted in 81,000 completed interviews.

The 1999 NHSDA fully employed another technological innovation: use of a hand-held computer at each sample dwelling unit to conduct household screening and to select the sample person(s) for the interview. With this new design, technology, and markedly increased sample size, the structure of the data collection staff also had to be modified significantly for 1999. This chapter presents details of these changes.

2.1 Sample Design for State-Level Estimates

Since 1991, the NHSDA target population has been the entire civilian, noninstitutionalized population aged 12 or older living in the United States. Excluded from the sample were the homeless who never use shelters, active military personnel, and residents of institutional group quarters, such as jails and long-term hospitals. This target population remained the same in the 1999 NSHDA. To provide State-representative estimates using newly designed survey-weighted hierarchical Bayes estimation methodology, it was necessary to change from a national sample design to one that provided a representative sample in each of the 50 States and the District of Columbia.

2.1.1 Old Design

Prior to 1999, the NHSDA sample was generally a national, stratified, multistage, area probability sample. The building blocks for the sampling frame were counties or groups of counties called primary sampling units (PSUs). A number of these PSUs were selected with certainty, and the remainder were placed into strata and sampled to provide a representative national sample. The sample PSUs selected prior to 1999 fell primarily into densely populated metropolitan areas across the nation. Typically, about 40 States had sample cases in any given year.

Each PSU was further divided into smaller areas (hereafter called segments) that comprised one or more U.S. Bureau of the Census blocks and were designed to include approximately 90 households. Within PSUs that had been selected with certainty, the segment became the first stage of sampling. The segments were stratified to provide sufficient and representative samples across the range of socioeconomic status of Hispanics, blacks, youths aged 12 to 17, and young adults aged 18 to 25. Within each segment, all households were listed and screened, and a subsample of 0, 1, or 2 persons was selected for each. Sample persons were selected using a simple matrix of information from the screening interview on age and race/ethnicity.

The change between 1998 and 1999 to produce State-level estimates as well as national estimates was due largely to the devolvement of responsibility, authority, and funding from the Federal Government to State and local governments. Many programmatic decisions were being made at the local level, so it became logical to provide estimates for States to track their progress.

2.1.2 New Design

In 1999, therefore, the NHSDA sample was redesigned to be representative of each of the 50 States and the District of Columbia. In the new design, each State was first divided into field interviewer (FI) regions (an area for which a single field interviewer would have responsibility). The FI regions were designated based on similar total population sizes. In more densely populated areas, the FI regions were smaller. The FI regions in rural areas consisted mostly of one or more contiguous counties; they comprised combinations of contiguous U.S. Bureau of the Census tracts and counties or both.

To enhance the precision of the national estimates, oversampling was used in each of the eight most populous States (California, New York, Texas, Florida, Pennsylvania, Illinois, Michigan, and Ohio), which together accounted for about 50 percent of the total U.S. population. Sample sizes were designed to yield about 3,600 to 4,630 respondents in each of those States. In the remaining 42 States and the District of Columbia, the goal was to yield 900 to 1,030 completed interviews. The sample size for youths was increased by 2,500 to enhance precision for the national estimates from the Tobacco Supplement for youths aged 12 to 17 years old. Because the design of the national sample used roughly equal sample sizes (to achieve equal precision) separately within the large and small States, the supplemental sample was implemented by increasing the samples for the relatively larger States within each group. This was done to reduce the impact of unequal weights and nonproportional (to the number of youths in each State) allocation of samples among States. The overall sample size and allocation was designed to produce national estimates for subdomains of interest with precision similar to, or better than, prior years while also producing sufficiently precise State estimates.

In States with smaller populations and samples, 12 FI regions (geographic strata) were defined. In the eight large States, 48 FI regions were created. Within each FI region, a sample of segments was selected at the first stage. These segments were designed to be somewhat larger than the segments used in prior years, encompassing about 175 households compared with the earlier segments of about 90 households. The segments were first sorted by a socioeconomic indicator and by the percentage of population that is non-Hispanic and white. In each FI region, eight segments were selected for a given year and two were assigned randomly to each of the four quarters of data collection. Within each sample segment, all addresses were listed and a subsample of the addresses was selected. Nationally, approximately 224,000 addresses were sampled and approximately 188,000 were determined to be eligible.

For the 1999 NHSDA, the target number of interviews for the basic 50-State design was 67,500. This was augmented with an additional sample of 2,500 youths aged 12 to 17 years old, bringing the total targeted sample to 70,000. The added youth sample was allocated among States roughly proportionate to the size of their youth population (12 to 17 years old) in order to reduce the impact of unequal State youth weighting effects on youth estimates at the national level.

Although oversampling of blacks and Hispanics had been incorporated into the sample design since 1985, the move to the 50-State design and much larger sample sizes rendered the expected numbers of blacks and Hispanics in the sample population sufficiently large so that oversampling of these populations was no longer required in 1999 and later NHSDA surveys.

Since its inception, the target population of the NHSDA has been all persons aged 12 or older. However, there has been special interest in the younger populations aged 12 to 25 because these are the ages at which most initiation of substance use takes place. In addition, substance prevention programs historically have focused on youths aged 12 to 17, and especially the younger ages (Cazares & Beatty, 1994). For these reasons, one of the goals of the 1999 survey was to make separate State estimates for ages 12 to17, 18 to 25, and 26 or older. Therefore, approximately one-third of the total initial sample was allocated to each age group. Within the 26 or older age group, the sample was further allocated to ages 26 to 34, 35 to 49, and 50+. Relatively more of the sample was assigned to the younger age groups than their proportionate representation would indicate.

Locating sufficient eligible sample persons in the 12- to 17-year-old age group required a two-stage process of first oversampling and listing addresses, yielding a national sample of 188,000 eligible addresses, then screening them to provide household information for use in the final stage of selection. The second stage involved subsampling to achieve the designated sample sizes for individuals by age group. This selection implicitly resulted in a subsample of households in which zero persons were selected. The allocation of the national sample to the three age groups was carried down to the State level to provide approximately one-third of the sample in each of the age groups: 12 to 17, 18 to 25, and 26 or older.

2.2 Electronic Screener

In the final stage of sampling, a resident of each selected dwelling unit was contacted to complete the screening interview using the following steps:

1. A roster of household members was compiled during the screening interview.
2. A series of questions was asked to obtain a list of all persons who were members of the household, along with each person's relationship to the head of household (the person or one of the persons who owns or rents the home) and each person's gender, age, race, ethnic origin (Hispanic or non-Hispanic), and current military status (active duty or not). Persons under the age of 12 and those on active military duty are out-of-scope for the NHSDA.

2.2.1 Old Design

Prior to 1999, all of the screening and oversampling by age and race/ethnicity had been completed on an oversized, paper screening form that was difficult to manage. When the household roster was completed, the interviewer executed a complex respondent selection algorithm by consulting a set of preprinted selection tables to select zero, one, or two household members for the interview. This process was difficult to manage, prone to error, expensive to process, and limited in terms of the sample selection algorithms that could be implemented.

2.2.2 New Design

In 1999, screening in both English and Spanish and case management were converted to a computer-based application, using the hand-held, pen-based Apple Newton computer. All cases (addresses) assigned to an FI were transmitted over telephone lines to his or her Newton. The Newton was programmed with the person-selection algorithm and the screening interview. Once an address was located, the FI would select the address line on the Newton, then have the Newton bring up the screening interview. If an appropriate household respondent were contacted, the FI would ask the screening questions, enter responses, and the Newton would automatically select zero, one, or two respondents for interviewing. Whether or not a household respondent was contacted, the FI would enter the result code for the visit and any relevant notes for future contact. Each day, the FI would transmit all data from the Newton to RTI by telephone. Any necessary revisions to the Newton program or data, such as changing the algorithm, or adding or deleting cases, were made during transmission.

The method used to subsample the eligible persons for the interview and to obtain the required sample sizes in each age group was a modified version of Brewer's method of selection for samples of size two. This method also enabled the selection of a random sample of two persons per household in a subset of the households. The random household sample of pairs was needed to analyze issues such as the impact of parents on their children's use of substances.

2.3 Sample Disposition

The total CAI sample after data collection included 66,706 completed person interviews. The total was fewer than the target sample of 70,000 because of lower than expected screening and interview response rates. The weighted response rates for household screening and for interviewing were 89.6 percent and 68.6 percent, respectively. Weighted response rates for individual States for household screening ranged from 79.9 to 96.1 percent. For interviewing, the response rates for the States ranged from 58.4 to 82.8 percent.

2.4 Weights and Estimation

Sampling weights were developed to reflect the probability of selection as well as numerous other adjustments made during the survey process. The primary weight was developed for analysis of responses from individuals; however, other weights were developed for analysis at the household level and for selected pairs of individuals representing different topics and subdomains of interest. The primary weight reflected the three stages of sample selection (segment, household, and person level) plus any adjustments due to counting and listing or subsampling during the collection phase. In addition, individual records could be further adjusted for outliers, nonresponse, and poststratification at the household or person level, or both. The data also were subject to adjustments due to editing rules and imputation for missing data.

Because the design involved a complex stratified multistage sampling approach, this needed to be reflected in the estimates of precision as well. Within each FI region, one member of each pair of segments assigned to one of the four quarters was designated to be in replicate 1 and the other in replicate 2. Together across the four quarters, there were two replicates, each with four segments. For small States there were 12 variance strata, and for large States there were 48. All selected households and individuals in a sample segment were assigned to the same sample replicate. The variance estimation thus reflected the stratification and clustering present within each State.

2.5 Data Collection Mode Change to CAI

Considerable research has been conducted over the years to continually improve the quality of data collected in the NHSDA. This research has focused on content, sample design, questioning strategies, interview mode, editing methods, and estimation procedures. The sensitivity of the information sought in the NHSDA has made it clear to the questionnaire designers that privacy is essential to maximizing respondents′ willingness to respond truthfully. Therefore, a self-administered interview method is preferred (Aquilino, 1993; Turner, Lessler, & Gfroerer, 1992; Harrison & Hughes, 1997). However, it is difficult to design written questions with a level of literacy appropriate for children as young as 12 years old and for older persons with varying reading abilities. The complexity of the instrument led to the desire to have greater ability to perform edit checks of responses during the interview. The advances of computer technology in the survey research field have enabled CAI to improve these areas over the traditional PAPI mode of interviewing. ACASI is a further advancement of the CAI technology that addresses both the privacy and literacy issues. Privacy is addressed in that the respondent uses a self-administered computerized questionnaire that the field interviewer cannot see. Literacy is addressed with the audio enhancement—the respondent can listen to the questions through headphones as well as read them from the computer screen, thus a limited ability to read does not pose a problem.

After several years of testing the CAI and ACASI methodology and making improvements to the instrument to take advantage of the methodology, full implementation of this technology was initiated in the 1999 NHSDA. The interview was programmed in Blaise 4.0 for Windows in both English and Spanish and loaded on Gateway Solo 5100 Multimedia Notebook laptop computers. The CAI begins when the FI enters a unique sample person questionnaire identification number generated by the Newton. The FI selects the language to be used (English or Spanish), conducts the initial CAPI portion of the interview, then turns the laptop over to the respondent, pointing out the keys the respondent will use and giving instructions in use of the headphones. The respondent is presented with a short, interactive ACASI tutorial that provides basic instructions and practice in entering responses to different types of questions, changing responses, and having questions repeated. Certain function keys are labeled with a template to aid respondents in backing up one question at a time, turning off the sound, replaying the sound, and entering "don't know" and "refused" responses.

The FI makes every effort to ensure that the interview is conducted in private and that no person other than the respondent can see or hear the questions and answers. After the respondent completes the ACASI portion, the FI administers the remaining CAPI portion of the interview. Finally, the FI enters responses to a few questions about his or her impressions of the interview. These FI debriefing items include an assessment of the effect of the computer on the respondent's decision to participate and on the privacy of the interview. The interview takes about 1 hour on average.

Each day, FIs transmit all screening work from the Newton and interviews from the laptop to RTI headquarters over telephone lines. Screening data are updated daily and made available on the project website to RTI and SAMHSA staff for monitoring field progress.

2.6 PAPI National Sample

Because the goal was to use the CAI methodology in the main NHSDA sample in 1999 and later years, it was necessary to estimate long-term trends between 1999 and earlier years that were based on the PAPI methodology. Therefore, a supplemental national sample was selected and fielded in 1999 using the same PAPI method of data collection and question wording and a sample design similar to that of previous years. In addition to the goal of estimating the trend between 1998 and 1999 using the same PAPI methodology, it was also desired to estimate the impact of the change from PAPI to CAI using the 1999 PAPI and 1999 CAI. The 1999 PAPI sample, therefore, was designed like recent NHSDA samples to oversample blacks, Hispanics, and the younger populations aged 12 to17 years old and 18 to 25 years old.

Prior to 1999, the design had used certainty and noncertainty PSUs for the first stage. The PSUs comprised counties or groups of counties, and a probability-proportionate-to-size sample of the noncertainty PSUs was selected. For PSUs designated as certainty PSUs, the first stage of selection was at the block group level.

Because the new State-representative CAI sample was based on 900 FI regions within the 50 States and the District of Columbia, it became convenient to use these FI regions as a basis for the first stage of selection of the PAPI sample. For the 1999 PAPI sample, four stages of selection were used:

- Stage 1: FI regions
- Stage 2: segments within FI regions
- Stage 3: dwelling units within segments
- Stage 4: persons within dwelling units.

The FI regions were stratified into six strata based on the percentage concentration of Hispanics, blacks, and whites. Required sample sizes were further determined for five age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 plus) by three race/ethnicity groups (Hispanic, black, and white) within each stratum. Once the allocations to substrata were determined, they were used to specify a composite measure of size that would result in the desired sample sizes. To maximize the selection of a racially diverse set of FI regions across a variety of States, FI regions were sorted within strata by State and by percent minority within State. A total of 250 FI regions were selected with probability proportionate to a composite measure of size. This selection was intended to "mimic" the first-stage PSU samples in prior survey years, constrained by the desire to minimize field collection costs.

Within each of the selected FI regions, all segments that had been sampled for the main CAI sample were also designated for the national PAPI sample. The purpose of this "pairing" was to create two samples of households and persons as similar as possible in every aspect except the mode of data collection: one with the CAI mode and one with the PAPI mode. This was done to better isolate and measure the effect of just the change in collection mode. Households selected for the CAI sample were made ineligible for the PAPI sample. Given eight segments per year (two per quarter), the total segment sample size was 2,000, with 500 assigned to each quarter. The PAPI sample used the Newton hardware and the modified

Brewer's method for the selection of persons. Household and person sampling were designed to yield a sample of 10 completed interviews by PAPI in each segment, for a total of 5,000 respondents per quarter.

A sample size of 20,000 persons, comparable to recent years' national samples, was originally planned for the PAPI sample (5,000 respondents per quarter). However, the sample was reduced to 15,000 during data collection due to budgetary constraints. This modification of sample size occurred in stages. First, Quarter 1 of the 1999 survey was yielding a larger-than-expected number of interviews (about 5,500 instead of 5,000), thus the targeted sample size was adjusted to 4,500 in Quarter 2. Then, after two quarters of data collection, it was decided to decrease the total targeted sample size from 20,000 to 15,000. Thus, in Quarters 3 and 4, the target sample size was set at 2,500.

Due to a lower-than-expected response rate, the final PAPI sample size was 13,809 completed interviews. Weighted response rates for screening and interviewing were 83.8 percent and 66.6 percent, respectively. Overall, 46,328 dwelling units were selected and 40,584 were determined to be eligible.

Weights were developed for the PAPI sample that reflected the segment, household, and person level sampling as well as the first-stage selection of FI regions. Adjustments were made for nonresponse, outliers, and poststratification.

2.7 Major Differences Between the PAPI and CAI Instruments

Significant changes were made in the NHSDA instrument to improve data quality and reduce respondent burden through the new technology. These include changes in the methodology for obtaining information on recency and frequency of substance use, implementation of data quality checks throughout the instrument, and new or revised questions in response to a need for more detailed data on tobacco use and new psychotherapeutic drugs. In response to new Office of Management and Budget (OMB) regulations, the CAI instrument allows respondents to report more than one racial group and more than one Hispanic subgroup. None of these changes were made to the 1999 PAPI instrument.

The most basic difference between the CAI and PAPI instruments reflects one of the main advantages of using the new technology, namely, the extensive use of skip patterns. To maximize respondent privacy and to minimize respondent confusion, no skip patterns were used in most of the self-administered PAPI answer sheets. Thus, respondents were presented with every question regardless of whether they had used a particular substance. Appropriate answer categories were included so that respondents who had not used the substance at all or had not used the substance in the time period of interest could indicate that the question was not applicable. Because of this procedure, persons responding to the PAPI questionnaire had more than one opportunity to report use of a substance in a given period. Based on results from the 1997 Field Experiment, persons selected for the CAI questionnaire were given only a single opportunity to report use of a substance for a given period, and extensive skip patterns were used to achieve this.

Response categories for the recency-of-use question were shortened from four in PAPI to three in CAI. In PAPI, all substances had one recency question, "How long has it been since you last [used the substance]?" and four response categories: "within the past 30 days," "more than 30 days ago but within the past 12 months," "more than 12 months ago but within the past 3 years," and "more than 3 years ago." In CAI, the same recency question is asked for all substances except tobacco, with three response categories: "within the past 30 days," "more than 30 days ago but within the past 12 months," and "more than 12 months ago." For tobacco, two recency questions are asked in CAI. First, a 30-day question is asked, and, if the respondent has not used the tobacco product within the past 30 days, an additional recency question is asked, with response categories including "more than 30 days ago but within the past 12 months," "more than 12 months ago but within the past 3 years," and "more than 3 years ago." The separate question for 30-day recency and the detailed information on recency of more than 12 months ago was recommended by a panel of experts in measurement of tobacco use (SAMHSA, 2001).

Another change that took advantage of the new technology was the extensive use of "fills" in the CAI instrument. Information given in previous questions is automatically filled in in subsequent questions to improve clarity, and, in questions about specific time periods, the reference date is automatically filled.

In the PAPI instrument, a single question was used to measure frequency of use in the past 12 months. Respondents were asked to select one of nine categories to report a range of days in the past 12 months that the substance was used. After extensive testing, a two-stage question was implemented in the CAI, in which respondents first chose the time unit that was easiest for them (days per year, days per month, or days per week) and then reported the number of days.

Data quality checks were implemented in the CAI questionnaire. These consisted of range edits, which triggered an error message if an out-of-range response was entered; verify checks, for unusual or inconsistent responses; and inconsistency resolution edits, which asked the respondent to choose the correct response from two inconsistent entries. ***Chapter 4*** of this report describes the new method for measuring frequency of use and the resolution methodology and their effects on the 1999 CAI data. Changes in interview questions are summarized below.

The tobacco module in CAI is significantly changed from PAPI in the following ways:

- wording for cigarettes and cigars was changed from "even a few puffs" to "part or all of a ___"
- snuff and chewing tobacco are asked about separately
- new questions on usual (past 30-day) brand of cigarettes, cigars, snuff, and chewing tobacco
- new questions on use of pipe tobacco and "roll-your-own" cigarettes
- a new set of questions for adolescents who do not smoke, asking about the likelihood of initiating smoking in the next year.

Questions about nonmedical use of prescription drugs were revised. The term "pain killer" for analgesics was changed to "pain reliever," and all pill cards were updated to include new drugs in each category. In PAPI, certain items on each pill card were asked about separately, then an "any other" question was asked, and the respondent could write in the name of other drugs. In CAI, the pill cards were designed with similar drugs grouped into boxes above and below a red line. The questions first ask about lifetime use of any of the specific drugs in the first box, then the second box, and so on. Then the respondent is asked if he/she has used any of the drugs below the red line, and, if so, the list of drugs appears and the respondent selects the one(s) he/she has used from the list.

New questions to elicit month and year of first use were added to the CAI instrument for all core substances to obtain better incidence data on persons recently initiating substance use. That is, for all respondents who reported that their age at first use was their current age or 1 year less than their current age, two questions appeared: first, a question that asks what year they first used the substance, and then a question that asks what month in that year they first used it.

Two questions traditionally asked in the PAPI core substance sections were deleted as a result of field and laboratory testing: lifetime number of days a substance was used and number of days in the past 12 months that a respondent got very high or drunk. The former question was determined to be too difficult for most respondents to answer accurately, and the latter was determined to be poorly understood based on laboratory testing and on data from consistency checks in the field test.

2.8 Data Collection Staff

Major changes to the organizational structure and the size of the data collection staff were required for the transition from the 1998 to 1999 NHSDA. These changes were necessitated primarily by the dramatic increase in sample size and the sample design change from a national sample to a 50-State sample. Training of field staff also became more involved because of the change in interview methodology from a totally PAPI administration to a primarily CAI administration with a PAPI supplement and the introduction of the electronic screening technology using the Newton.

From 1998 to 1999, the number of interviewers increased from about 300 to about 1,230. This increase and the design change requiring interviewers in every State and the District of Columbia resulted in a majority of the interviewing staff (approximately 87 percent) having no experience with the NHSDA interview. Approximately 40 percent of the staff had no prior interviewing experience before being hired for the 1999 NHSDA.

The increased number of interviewers and the decreased level of experience necessitated much more personal supervision than in 1998. The number of field supervisors (FSs) was increased from 15 in 1998 to 80 in 1999. Again, the great majority of these FSs had no NHSDA supervisory experience, and a majority had no experience as an FS.

This, in turn, required closer and more focused supervision of the FSs. The number of regional supervisors (RSs) who are responsible for supervising the FSs increased from 5 in 1998 to 19 in 1999. All but 2 of the RSs had no experience in their role as an RS.

To supervise this large and inexperienced number of RSs, a new position—regional director (RD)—was created. Seven RDs were responsible for managing the 19 RSs. This was a new management position, so none of these staff had experience in this role. These seven RDs were supervised by the national field director, who had been in that role for 5 years with the NHSDA prior to 1999.

The staffing problems resulted in considerable variations in the State-level response rates and the quality of the data collection process. These issues are discussed further in ***Chapters 3*** and ***8***.

2.9 References

Aquilino, W.S. (1993). Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly, 57*, 358–376.

Cazares, A., & Beatty, L.A., Eds. (1994). *Scientific methods for prevention intervention research.* NIDA Research Monograph 139 (DHHS Pub. No. 94-3631). Rockville, MD: National Institutes of Health.

Harrison, L., & Hughes, A., Eds. (1997). *The validity of self-reported drug use: Improving the accuracy of survey estimates*. NIDA Research Monograph 167 (NIH Pub. No. 96-4147). Rockville, MD: National Institutes of Health.

Substance Abuse and Mental Health Services Administration. (2001). *Development of computer-assisted interviewing procedures for the NHSDA*. Office of Applied Studies, Methodological Series M-3 (DHHS Pub. No. SMA 01-3514). Rockville, MD: Author.

Turner, C.F., Lessler, J.T., & Gfroerer, J.C., Eds. (1992). *Survey measurement of drug use: Methodological studies* (DHHS Pub. No. ADM 92-1929). Rockville, MD: National Institutes of Health.

3

Nonresponse in the 1999 NHSDA

Joe Eyerman, Dawn Odom, Shiying Wu, and Dicy Butler

The redesign of the National Household Survey on Drug Abuse (NHSDA) in 1999 resulted in major changes in many aspects of the data collection procedures. This raised concerns that the response rate could be affected. In particular, the increased sample size, reduced clustering of sample segments, transition from a paper to a computerized screening instrument, and the transition from paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI) all had the potential to change the response rate. During the first quarter of 1999, an assessment of the progress in completing the fieldwork indicated a reduction in response rate, relative to response rates achieved historically in the NHSDA. To address this problem, several management actions were immediately implemented. Although the response rate improved steadily throughout the remainder of the year, the result was a significantly lower response rate for the 1999 NHSDA than for prior NHSDAs.

Extensive analysis was undertaken in an attempt to understand the reasons for the drop in response rate and how it was related to each of the design changes. This chapter summarizes this analysis. It also discusses the management actions implemented during 1999 to improve the response rate and assesses the effectiveness of these actions.

3.1 Background

The NHSDA has historically employed a variety of protocols designed to reduce nonresponse at the screening and interview stages of the survey. For example, the project has always used lead letters, controlled access procedures, a minimum of five screening calls, and focused training and field efforts to prevent and convert refusals. In past years, these approaches have been effective, and the response rates were high and stable until the design changes in 1999. From 1994 to 1998, the annual screening rate was about 93 percent, and the interview response rate was about 77 percent, as demonstrated by ***Figure 3.1***.[1] The highest quarterly screening rate for this period was 95 percent in Quarter 2 of 1995. The highest interview rate for this period was 83 percent in Quarter 3 of 1995.

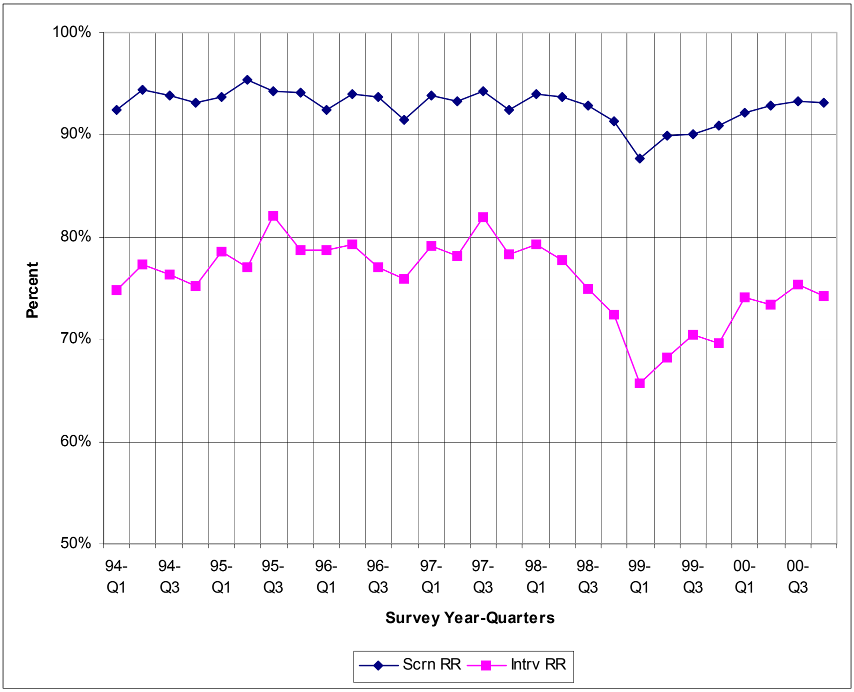

Figure 3.1 Screening and interview rates by quarter, 1994 to 2000.

The impact of the design changes can be seen starting in the fourth quarter of 1998, coinciding with the increase in recruiting and training in preparation for 1999. The screening and interview rates for that quarter were the lowest for the period 1994 to 1998 (92 and 72 percent, respectively). The downward trend continued through the first quarter of 1999. The rates began to improve in the second quarter of 1999, but did not rise

[1]All response rates are weighted.

above the fourth quarter 1998 rates until the first quarter of 2000. The annual response rates for the period 1994 to 1999 are presented in ***Table 3.1***.

Table 3.1 NHSDA Weighted Screening and Interview Rates: 1994 to 1999 (%)

Year	Screening	Interview	Overall
1994	93.4	75.9	70.9
1995	94.4	79.1	74.6
1996	92.9	77.7	72.2
1997	93.5	79.4	74.2
1998	92.9	76.0	70.6
1999 CAI	89.6	68.6	61.4
1999 PAPI	83.8	66.6	55.8

3.1.1 Management Actions

When the reduced response rate became evident during the first quarter of 1999, an action plan for addressing the problem was developed and quickly put in place. Although a definitive assessment of the causes of the drop in response rate was incomplete, the information available at that time clearly indicated that an insufficient staff (in terms of size and experience) of field interviewers (FIs) was a primary cause. ***Table 3.2*** summarizes the management actions taken early in 1999 to address nonresponse.

The impact of these changes cannot be quantified accurately for several reasons. First, because the changes reflect adjustments in the survey management design, they were applied to the full study and implemented as soon as possible. This prevents evaluation of the impact of the presence or absence of the management change. Second, in most cases, the changes were designed to address specific problems identified by the field staff. These changes were often implemented in groups and therefore cannot be isolated for analysis. Third, many of the changes were designed to improve the overall management design of the project; therefore, they have no clear direct relationship with response rates. Perhaps the best evaluation of the 1999 management adjustments can be seen in the quarterly response rates for the screening and the interview, plotted in ***Figure 3.1***. The rates reached an all-time low in the first quarter of 1999. The screening and interview rates improved in the second quarter and have continued on an overall upward trend since. Although this cannot be directly associated with the management adjustments made during the survey year, it is reasonable to assume that the management adjustments contributed to the improvement in the following quarters.

Table 3.2 Summary of Management Changes

Action	Description
Decreased workload per interviewer	Many interviewers were unable to work the required 15 to 20 hours per week. More interviewers were hired and allowed to work fewer hours per week.
Mini-training sessions	Region-specific mini-trainings were tailored to the needs of a poorly performing region. The 1-day mini-sessions were conducted by the regional directors, regional supervisors, and field supervisors (FSs).
Increased management-to-staff ratio	Two regional directors, two regional supervisors, seven field supervisors, and more support staff for regional directors were added in Quarter 2. Six super FSs were added in Quarter 3.
Increased field interviewer pay rate	The starting pay rate was increased to compete with the U.S. Bureau of the Census and private survey firms for new hires. Equitable increases were given to retain existing staff.
Traveling field interviewers	The traveling field interviewers (TFIs) were assigned to one of six regional teams to assist with productivity problems. TFIs were distributed regionally based on staffing needs and were not given a standing local work assignment.
Realignment of field supervisor regions	Problem regions identified in Quarter 1 were divided and reassigned in Quarter 2 to more than one FS. This provided a more equitable distribution of the hard-to-contact workload.
Sample waves	The quarterly sample was partitioned into waves to allow for tighter control of work effort. Only the mandatory sample was released in the first wave; additional sample was released once sufficient progress was made.
Site visits	Regional supervisors and field supervisors visited field interviewers to provide detailed guidance and management of problem regions.
Quarterly pay increases	The annual pay increase schedule was changed to quarterly to more closely associate performance with income and to reduce attrition in the field staff.

3.1.2 Staffing

One of the persistent problems facing the NHSDA management team was maintaining sufficient field staff to complete 66,706 CAI and 13,809 PAPI interviews. This required a large expansion of the field interviewer and field supervisor staff to respond to the increased workload. The recruitment, hiring, and training of the additional staff began in the last quarter of 1998 and continued through the end of the data collection year. This process is summarized in the training and attrition schedule ***(Table 3.3)***.

The effort to recruit new interviewers revealed several interesting challenges. First, there appeared to be a threshold for the number of qualified and capable field interviewers that could be hired in a short period of time. Second, the qualified field interviewers were not necessarily distributed in the same States and cities as the sample elements. Third, prior experience with field surveys is a very important measure of interviewer productivity,

Table 3.3 Summary of Interviewer Training and Attrition

Month	FI Training Sessions & Locations	FIs Trained	Cum. No. of FIs	Attrited FIs	Cum. No. of Attrited FIs
Initial Training Session at 14 sites; Makeup Training Session at Raleigh (NC) and Los Angeles					
Jan		1,073	1,073	40	40
Replacement Training Session Locations					
Feb	Cary (NC) and Los Angeles	62	1,135	46	86
Mar	Raleigh and Los Angeles	86	1,221	54	140
Apr	Atlanta, Chicago, Minneapolis, Pittsburgh, Stamford (CT)	56	1,277	66	206
May	Los Angeles and Minneapolis	67	1,344	58	264
June	Los Angeles, Minneapolis, Pittsburgh, Stamford	258	1,602	73	337
July	Durham (NC), Los Angeles, Stamford	55	1,657	81	418
Aug	Los Angeles and Raleigh	71	1,728	54	472
Sept	Cincinnati, Los Angeles, Raleigh, Stamford	152	1,880	96	568
Oct	Durham	51	1,931	59	627
Nov	Raleigh	66	1,997	69	696
Dec	No training session	0	1,997	60	754

and the number of experienced field interviewers was limited. Fourth, a tight labor market made the competition for hourly wage and part-time workers intense, particularly in cities with many alternative employers offering higher pay rates or more flexible hours. This resulted in a high rate of attrition.

3.2 Preliminary Analysis

Initial attempts to explain the decline in the response rates were made with descriptive statistics using the set of available data. Auxiliary data about the survey process had been collected as part of the NHSDA protocol and in anticipation of the 1999 redesign. Many of these data had been routinely collected as part of the standard NHSDA procedures; others were collected specifically to evaluate the impact of the design change on response rates.

The data can be grouped into the conceptual framework presented in Groves and Couper (1998). This framework describes the process of survey participation as the interaction between the respondent and the interviewer, which is influenced by four factors: (1) social environment, (2) householder (respondent) characteristics, (3) interviewer characteristics, and (4) survey design. The data sources used for the preliminary analysis included segment-level U.S. Bureau of the Census data and other government data sources, a survey of field interviewers who worked on the 1999 NHSDA, design characteristics such as the survey mode or the number of persons selected in a household, and householder characteristics collected during the screening stage of the NHSDA data collection.

The wide range of data led to a rich analysis of the nonresponse patterns in 1999 (RTI, 2000). This analysis included extensive tables with summaries of the sources of nonresponse (refusals, noncontacts, etc.), geographic distribution of nonresponse, and historical trends. Most of these were evaluated with contingency tables and regression models along with measures of social environment, respondent characteristics, interviewer characteristics, and survey design (Snijkers, Hox, & De Leeuw, 1999; Laurie, Smith, & Scott, 1999).

Unfortunately, many of the measures used in the preliminary analysis were limited in scope and coverage. For example, the survey of field interviewers provided a rich source of data to explore the traits of successful field staff, but it was administered only to the subset of field interviewers who returned to the project in 2000. Analysis of these data provides considerable detail in specific situations, but cannot be used to generalize to the full project because they apply only to a subset. For this reason, the set of descriptive statistics reviewed in this chapter and used for the models is limited to those measures that are available for 1998 and for both CAI and PAPI in 1999.

The two major sources of the data for these measures were the sample design and data collected at screening. The information used to design the national probability sample for the study was used to code the environmental characteristics for each eligible case. The respondent characteristics were taken from the data collected during the eligibility screening and were available for both interview respondents and nonrespondents. The interviewer characteristics were taken from a composite of auxiliary data collected during the hiring and training process, during the data collection year, and during their entire tenure on the NHSDA. Two measures of experience were generated. The first, NHSDA experience, is a dichotomous measure scored 1 if the interviewer ever worked on the NHSDA as an employee of RTI. The second, on-the-job experience, is a measure of the amount of each employee′s experience with the current data collection instrument and design. It is a count of the number of screenings completed by the interviewer prior to the interview in question and is grouped into five categories: 0 to 59, 60 to 119, 120 to 179, 180 to 299, and 300+ completed screenings.[2]

3.2.1 Patterns of Nonresponse

In general, the preliminary analysis was consistent with expectations drawn from the Groves and Couper (1998) conceptual model and in the general literature (Gfroerer, Lessler, & Parsley, 1997; Couper & Groves, 1996; Huisman, Krol, & Van Sonderen, 1998; Caspar, 1992). Results are shown in ***Table 3.4***.

- **Environment**. As expected, eligible respondents in the Northeast were less likely to complete the interview than those in other parts of the country. This is consistent with the conventional wisdom about the cluster of urban centers in the Northeast and is supported by the population density measure, which shows a negative relationship between population density and response rates.
- **Respondent Characteristics**. Consistent with findings in the literature and with past experiences on the NHSDA, the most cooperative respondents were females and youths ages 12 to 17.
- **Field Interviewer Characteristics**. Consistent with the literature and past experience, the most successful interviewers were in the older age groups. Also as expected, interviewers with experience were much more successful.
- **Survey Design**. The preliminary analysis showed a pattern of lower response rates for the PAPI mode than for the CAI. This relationship is evaluated in more detail later in the chapter.

3.2.2 Trends in Type of Nonresponse

A closer examination of the sources of nonresponse provides some insight into the declining response rate in 1999. The final disposition of

---

[2] The count of completed screenings was expected to be correlated with the response gains yielded by the management adjustments made over the survey year. An indicator for the quarter in which the survey was attempted was added to adjust for this.

Table 3.4 Weighted Interview Response Rates by Environmental Factors and Respondent and Interviewer Characteristics

		1998 (%)	1999 PAPI (%)	1999 CAI (%)
Total		76.0	66.6	68.6
Environmental Factors				
Quarter	1	79.3	60.3	65.7
	2	77.7	69.0	68.3
	3	74.9	69.0	70.5
	4	72.3	67.9	69.6
Census region	Northeast	69.2	59.1	64.0
	North-central	77.5	72.5	70.9
	South	80.7	64.9	69.6
	West	73.0	65.6	67.5
Population density	Segment in MSA, >1 mil	72.5	62.9	65.0
	Segment in MSA, 250K-1 mil	75.5	66.4	67.9
	Segment in MSA, <250K	79.6	70.3	72.8
	Segment not in MSA, not rural	80.6	75.6	75.5
	Segment not in MSA, rural	81.8	70.2	74.1
Respondent Characteristics				
Age	12-17	81.3	76.3	78.1
	18-25	76.9	67.8	71.2
	26-34	75.1	65.0	69.4
	35+	75.1	65.0	66.1
Gender	Male	73.5	64.6	67.1
	Female	78.1	68.3	69.8
Race	Hispanic	78.7	72.1	74.6
	Black	78.4	72.1	70.4
	White and other	75.3	65.0	67.5
Interviewer Characteristics				
NHSDA experience	Prior	77.9	71.1	69.1
	None	70.3	65.2	68.3
Age	Missing	64.5	70.0	69.7
	0-40	77.7	63.8	67.1
	41-50	75.7	64.0	68.8
	51-60	77.9	69.7	68.9
	61+	74.4	67.7	68.3
Gender	Male	78.2	62.2	67.1
	Female	75.6	67.6	68.9
Race	Asian/Pacific Islander	75.1	65.9	73.5
	Black	72.3	65.4	68.8
	White	76.8	66.7	68.6
	Other	69.2	66.7	68.0
On-the-job experience (number of screenings completed)	0-59	69.1	63.4	67.1
	60-119	75.2	64.2	67.8
	120-179	74.4	70.2	69.7
	180-299	77.7	70.1	71.0
	300+	78.8	70.3	72.0

eligible and selected cases from 1994 to 1999 is presented in ***Table 3.5***. In general, the percentage of noncontacts remained stable for the period, including 1999. However, the percentage of refusals for both the screening and the interview were highest in 1999, which appears to account for most of the nonresponse in 1999. This suggests that the relatively low response rate in 1999 was a result of the interviewers' inability to convince sample members to participate once they were contacted. One possible explanation for this can be drawn from the characteristics of the sampled persons. It is possible that the U.S. population is becoming more hostile toward surveys and more likely to refuse. This explanation would account for the small upward trend in the percentage of refusals for the period, but would not account for the dramatic increase between 1998 and 1999.

Table 3.5 Final Disposition of NHSDA Cases: 1994 to 1999

Year	Screening (S) or Inter-viewing (I)	Total Eligible Cases	Unable to Contact†		Refusal		Other Incomplete‡		Complete	
			N	%	N	%	N	%	N	%
1994§	S	58,150	1,707	2.9	1,558	2.7	357	0.6	54,528	93.8
	I	22,785	1,693	7.4	2,276	10.0	1,007	4.4	17,809	78.1
1995	S	56,469	1,427	2.5	1,521	2.7	320	0.6	53,201	94.2
	I	22,016	1,394	6.3	2,004	9.1	871	4.0	17,747	80.6
1996	S	56,912	1,876	3.3	1,727	3.0	539	1.0	52,770	92.7
	I	23,240	1,795	7.7	2,300	9.9	876	3.8	18,269	78.6
1997	S	81,068	2,635	3.3	2,379	2.9	918	1.1	75,136	92.7
	I	31,290	2,198	7.0	3,364	10.8	1,223	3.9	24,505	78.3
1998	S	80,866	2,325	2.9	2,647	3.3	727	0.9	75,167	93.0
	I	33,128	2,300	6.9	3,937	11.9	1,391	4.2	25,500	77.0
1999	S	187,842	6,479	3.4	11,097	6.0	1,100	0.6	169,166	90.1
CAI	I	89,883	5,692	6.3	14,164	15.8	3,321	3.7	66,706	74.2

† Includes controlled-access situations.
‡ Includes language barriers, mental or physical incapabilities, and situations in which the screener was not returned from the field.
§ 1994 numbers are for the new questionnaire (SAMHSA, 1996).

The dramatic increase in 1999 is more likely a result of the changes to the survey design that year. It is likely that the large number of inexperienced interviewers used in 1999 produced a larger than expected percentage of refusals. Inexperienced interviewers are typically less confident when they approach the dwelling unit, resulting in more initial refusals. In addition, they are less able to overcome objections diplomatically and are more likely to accept initial refusals as final.

3.2.3 State Variations

State-level analysis supports the findings of past research about the role of urban density in survey participation (Groves & Couper, 1998). As expected, States with large urban centers generally had lower screening and interview rates than rural States. As can be seen in ***Table 3.6***, the States with the best screening rates were New Mexico (96.1%), Arkansas (95.7%), and West Virginia (95.6%). The worst screening rates were in Nevada (79.9%), Massachusetts (80.6%), and New York (84.3%). The

Table 3.6 1999 CAI Screening and Interview Response Rates by State (Weighted Percentages)

State	Screening Response Rate (%)	Interview Response Rate (%)	State	Screening Response Rate (%)	Interview Response Rate (%)
Overall	89.6	68.6	Missouri	91.3	73.6
Alabama	92.6	71.4	Montana	92.8	76.4
Alaska	91.1	77.2	Nebraska	90.0	72.1
Arizona	94.4	65.9	Nevada	79.9	63.1
Arkansas	95.7	80.5	New Hampshire	85.4	69.9
California	87.5	64.1	New Jersey	89.7	65.2
Colorado	91.6	65.8	New Mexico	96.1	77.8
Connecticut	85.6	58.6	New York	84.3	60.0
Delaware	87.1	58.4	North Carolina	92.9	71.8
District of Columbia	93.4	79.9	North Dakota	89.9	77.5
Florida	89.9	68.2	Ohio	90.4	67.8
Georgia	90.5	67.0	Oklahoma	91.6	67.8
Hawaii	89.1	67.6	Oregon	85.2	71.6
Idaho	92.9	75.5	Pennsylvania	92.3	69.0
Illinois	87.4	63.7	Rhode Island	86.7	66.7
Indiana	91.7	73.1	South Carolina	92.0	65.9
Iowa	92.4	69.7	South Dakota	94.4	76.1
Kansas	90.6	72.9	Tennessee	90.9	67.7
Kentucky	92.4	73.8	Texas	92.6	75.1
Louisiana	94.8	77.0	Utah	93.2	81.7
Maine	90.0	75.2	Vermont	90.3	74.5
Maryland	87.8	64.7	Virginia	89.8	66.3
Massachusetts	80.6	61.8	Washington	86.5	75.1
Michigan	88.2	66.5	West Virginia	95.6	74.3
Minnesota	89.5	77.7	Wisconsin	90.2	73.1
Mississippi	94.5	82.8	Wyoming	93.8	72.6

States with the best interview rates were Mississippi (82.8%), Utah (81.7%), and Arkansas (80.5%). The worst interview rates were in Delaware (58.4%), Connecticut (58.6%), and New York (60.0%).

3.2.4 Multivariate Models

The preliminary analyses reveal some important patterns of nonresponse, but do not address the design changes that occurred between 1998 and 1999. The multivariate models described in ***Sections 3.3*** through ***3.5*** draw on the information from the preliminary analyses to build multivariate models that isolate the impact of two of the major changes in the design. The first model series in ***Section 3.3*** evaluates the transition to a computerized screening instrument, the Newton, while controlling for the patterns of nonresponse revealed in preliminary analyses. The second model series in ***Section 3.4*** evaluates the transition from PAPI to CAI while controlling the same patterns of nonresponse. The final model evaluates the effectiveness of the set of control variables used to explain the changes between 1998 and 1999 by comparing the two years using only the cases collected with the PAPI instrument in both years.

3.3 Evaluation of Difference in Response Rate Patterns Between Paper and Computerized Screening Instrument

In past years, the field interviewers used printed tables to select the sampled persons within each household, as determined by the sample design. Interviewers could use the information contained in the printed tables to anticipate which persons would be selected from a household. In households with two or more persons in the same age group, the printed tables determined who would be selected based on the order in which they were recorded on the screening roster. Some of the interviewers may have manipulated the roster order to select the most convenient person within each age group. It was expected that this occurred most often in households with married adults. This would lead to a disproportionate selection of females because they were more likely to be home and screened than the males, as demonstrated in ***Table 3.7***.

Table 3.7 Weighted Percentage of Female Screening Respondents in Certain Domains

Age Group	Year/Mode of Survey		
	1998	1999 PAPI	1999 CAI
12-17	—†	—†	—†
18-25	54.00	54.82	53.62
26-34	57.20	55.55	54.82
35-49	58.75	53.55	55.30
50+	54.48	48.71	50.87

†—No estimate reported due to low precision (i.e., less than 10 observations in domain).

All of the 1999 screenings were conducted using the Newton hand-held computerized screening instrument. The Newton was programmed with the sample design and a screening script to standardize the process. The field interviewers did not have prior information about each household as they had with the paper screener; therefore they should not have been able to anticipate which persons would be selected. It was expected that the Newton prevented convenient sample selection manipulation. This should have resulted in a lower response rate in 1999 than in previous years because interviewers were not able to select the person who was readily available.

The following analysis evaluates the impact of the Newton in three ways. First, it examines the convenient-selection hypothesis by examining differences in selection rates between 1998 (paper screener) and 1999 (Newton). Second, it evaluates the robustness of the convenient-selection hypothesis by controlling for alternative explanations in a multivariate model. Third, it applies the results of the multivariate model to the response rates in 1999 to generate an adjusted response rate. The adjusted response rate approximates the rates that would have been produced if interviewers had been able to make convenient selections in 1999.

3.3.1 Analysis Design

The dataset used for this analysis consists of those households with more than one person in an age category used by the screening instrument (12 to 17, 18 to 25, 26 to 34, 35 to 49, 50+). These multiples are the households that are vulnerable to the effects of the convenient-selection process. In households with only one person per age group, the selection tables identified the appropriate person based on the sample design. In multiples, the appropriate age group was determined by the sample design, but the selected person within each age group was determined by the order in which the field interviewer listed the persons on the household roster. The definitions presented in ***Tables 3.8*** and ***3.9*** are used throughout this section of the chapter.

To isolate only those cases in which screener manipulation could occur, the data were subset in three ways. As stated above, only those cases in which a household contained two or more eligible persons in an age group were used in the analysis. In addition, the data were limited to cases in which the number of eligible persons in the age group did not equal the number of selected persons in the age group. For example, if two people were selected from an age group in which only two people were eligible, the FI would not have had the opportunity to manipulate the household roster. Finally, at least one male and one female had to be present in the age group because a roster with only one gender represented in an age group could not be manipulated based on gender.

The transition to the Newton was one of many changes to the NHSDA in 1999. The results in ***Table 3.10*** demonstrate that there was a difference between the selection process in 1998 and 1999. However, the difference might be explained by other changes between the two years.

Table 3.8 Definitions

Term	Definition
Multiples	Households in which there are more than one eligible person per age group. These are the critical cases for this analysis because they could be manipulated with the paper screener but not with the Newton.
Convenient household members	Persons at home when the interviewer conducts the screening.
Age group	The age groups used are 12-17, 18-25, 26-34, 35-49, and 50+.
Selection probability (P)	The theoretical conditional probability that a sample element has of being selected given that more than one person was selected in the age group. Because the analysis is limited to comparing males to females in multiples, the selection probability is determined by the number of persons within each age group. $$P = \frac{\#selected}{\#\,eligible}$$
Selection event (A)	The observed selection event for a sample element. If a person is selected, A = 1; otherwise, A = 0.
Difference (W)	The difference between the selection probability and the observed selection events. This is the dependent variable in the multivariate models presented in Tables 3.10-3.12. The expected value of W is 0; see Table 3.9. W = A - P

Table 3.9 Possible Values for W

Number of Persons in Age Group	Selection Event A	Selection Probability P	Difference W
2	0	.50	-.50
2	1	.50	.50
3	1	.33	.67
3	0	.33	-.33
3	0	.33	-.33

This possibility is addressed with a multivariate comparison of the selection probabilities and selection events for 1998 and 1999. The multivariate model controls for the other known correlates of response rates, as discussed in ***Section 3.2.***

Models were estimated for each age group within each year. The set of explanatory variables is the same as those used in the comparison of the 1998 and 1999 models discussed above.

Table 3.10 p-Values from χ^2 Analysis under Certain Domains† Comparing Gender Distribution of Selected and Nonselected Persons per Age Group

	Year/Mode of Survey		
Age Group	**1998**	**1999 PAPI**	**1999 CAI**
12-17	.43	.84	.41
18-25	.01	.12	.48
26-34	.00	.23	.08
35-49	.00	.29	.92
50+	.06	.78	.09

†The domain consists of cases in which (1) the household has two or more eligible persons in the age group, (2) at least one was male and one female, (3) at least one person was selected, and (4) the number of persons selected in the age group did not equal the number of persons in the age group.

3.3.2 Convenient-Selection Hypothesis

The convenient-selection hypothesis holds that interviewers in the pre-Newton survey years were more likely to select the available person by manipulating the roster in households with multiple eligible persons in the same age group. Specifically, the hypothesis states

> *Interviewers are more likely to select females in households with more than one eligible person per age group if the group contains at least one male and one female.*

This is based on the assumption that most multiple adult households comprise married couples and that the male is more likely to be employed outside of the home (see ***Table 3.7***), making the female a more convenient selection. This hypothesis is evaluated in ***Table 3.10***, which summarizes a series of comparisons of the expected and actual distribution of males and females from multiple households within each age group. A weighted chi-quare statistic was calculated for each comparison; the cells of the table contain the probability associated with the chi-square test. That is, the cells represent the probability that the difference between the actual and expected distribution is zero.

The results support the convenient-selection hypothesis. With the exception of the youngest age group, the actual distribution was always statistically different from the expected distribution in 1998. This difference is not present in the comparisons for 1999 PAPI and 1999 CAI. This suggests that the selection process was biased on gender measures for adults in 1998 but not in 1999.

3.3.3 Regression Analysis

The results listed in ***Tables 3.11*** through ***3.13*** suggest that the difference between the actual and expected gender distribution shown in ***Table 3.10*** could not be explained by the other explanatory variables alone. This suggests that the implementation of the hand-held Newton is the most likely reason that the differences in selection by gender did not appear in 1999.

The cells under the male and female headings represent the difference between the expected probability of selection and the observed selection rates. The cells under the estimate heading represent the difference between the males and the females. For example, in 1998 the observed selection rate for males aged 18 to 25 was .036 lower than expected by the sample design, and the rate for females was .038 higher, for a combined difference of .074, which is statistically significant (p = .021). If this result were applied to a household with one male and one female aged 18 to 25, the probability of selecting a male would be .464 (.500 - .036 = .464), and the probability of selecting a female from the same household would be .538 (.500 + .038 = .538).

The difference between males and females is statistically significant at the level of $p < .05$ for all age groups in 1998, except for the group of 12-17 year olds. This is consistent with the convenient-selection hypothesis and mirrors the results in ***Table 3.10***. When the control measures are entered into the multivariate model, the difference remains significant for age groups 26 to 34 and 35 to 49 (marginal at p = .059). This suggests that the difference for the age groups 18 to 25 and 50+ may have been a function of the control variables, but that the difference for 26 to 34 and 35 to 49 is likely a product of the convenient-selection process.

The difference between males and females is not significant for any age groups in 1999 for PAPI (***Table 3.12***) or CAI (***Table 3.13***), except for the CAI 50+. The absence of relationship holds for models with and without the control variables. This suggests that the convenient-selection process did not occur in 1999 when the Newton was used.

3.3.4 Impact on Response Rates

As a result of the convenient selection of available respondents, field interviewers increased the number of females selected for the interview in 1998. This should have artificially inflated response rates in 1998 by reducing the risk of noncontact associated with follow-up calls and by

Table 3.11 Difference between Expected and Actual Probability of Selection for Males and Females by Age Groups for 1998—Weighted Linear Regression

	Change to Expected Probability (W)		Difference from Females to Males		
1998 Age Group	**Male**	**Female**	**Estimate**	**SE**	**p-Value**
12-17 year olds					
Unadjusted mean	-0.018	0.017	-0.035	0.040	.377
Adjusted by screener race			-0.035	0.040	.373
Adjusted by screener race, FI characteristics,† and geographic variables			-0.014	0.085	.866
18-25 year olds					
Unadjusted mean	-0.036	0.038	-0.074	0.032	.021[a]
Adjusted by screener race			-0.074	0.032	.022[a]
Adjusted by screener race, FI characteristics,† and geographic variables			0.036	0.065	.586
26-34 year olds					
Unadjusted mean	-0.056	0.058	-0.113	0.035	.001[a]
Adjusted by screener race			-0.115	0.035	.001[a]
Adjusted by screener race, FI characteristics,† and geographic variables			-0.160	0.072	.027
35-49 year olds					
Unadjusted mean	-0.088	0.089	-0.177	0.032	.000[a]
Adjusted by screener race			-0.177	0.032	.000[a]
Adjusted by screener race, FI characteristics,† and geographic variables			-0.118	0.062	.059
50+ year olds					
Unadjusted mean	-0.028	0.028	-0.056	0.028	.053
Adjusted by screener race			-0.048	0.026	.065
Adjusted by screener race, FI characteristics,† and geographic variables			0.075	0.052	.153

†FI = field interviewer. FI characteristics include experience.
[a] Difference is statistically significant at the 0.05 level.

reducing the risk of refusals associated with the generally less cooperative male respondents. However, the use of the electronic screening instrument on the Newton prevented the convenient selection of available respondents in 1999. Therefore, the interview response rate should have been lower in 1999 than in 1998 by the amount of the 1998 inflation, other things being equal.

Table 3.12 Difference between Expected and Actual Probability of Selection for Males and Females by Age Groups for 1999 PAPI—Weighted Linear Regression

	Change to Expected Probability (W)		Difference from Females to Males		
1999 PAPI Age Groups	Male	Female	Estimate	SE	p-Value
12-17 year olds					
Unadjusted mean	0.008	-0.008	0.017	0.049	.734
Adjusted by screener race			0.017	0.049	.734
Adjusted by screener race, FI characteristics,† and geographic variables			-0.012	0.052	.820
18-25 year olds					
Unadjusted mean	-0.024	0.027	-0.051	0.037	.171
Adjusted by screener race			-0.051	0.037	.170
Adjusted by screener race, FI characteristics,† and geographic variables			-0.020	0.047	.674
26-34 year olds					
Unadjusted mean	-0.023	0.026	-0.049	0.037	.192
Adjusted by screener race			-0.049	0.037	.185
Adjusted by screener race, FI characteristics,† and geographic variables			-0.052	0.045	.249
35-49 year olds					
Unadjusted mean	0.019	-0.022	0.041	0.037	.275
Adjusted by screener race			0.041	0.037	.274
Adjusted by screener race, FI characteristics,† and geographic variables			-0.017	0.033	.604
50+ year olds					
Unadjusted mean	0.004	-0.005	0.009	0.036	.805
Adjusted by screener race			0.009	0.036	.809
Adjusted by screener race, FI characteristics,† and geographic variables			0.007	0.042	.876

†FI = field interviewer. FI characteristics include experience.

Table 3.14 compares the actual 1999 response rate to the inflated rate that would have been realized if the Newton was not used and if the interviewers had been able to continue with the convenient selection of available respondents. The adjusted rate in Table 3.14 is calculated by inflating the actual 1999 rate to account for the additional females who would have been selected under convenient conditions and the higher

Table 3.13 Difference between Expected and Actual Probability of Selection for Males and Females by Age Groups for 1999 CAI—Weighted Linear Regression

1999 CAI Age Groups	Change to Expected Probability (W)		Difference from Females to Males		
	Male	Female	Estimate	SE	p-Value
12-17 year olds					
Unadjusted mean	0.009	-0.009	0.019	0.022	.382
Adjusted by screener race			0.019	0.022	.388
Adjusted by screener race, FI characteristics,† and geographic variables			0.011	0.025	.645
18-25 year olds					
Unadjusted mean	-0.007	0.007	-0.013	0.021	.523
Adjusted by screener race			-0.013	0.020	.520
Adjusted by screener race, FI characteristics,† and geographic variables			0.007	0.023	.750
26-34 year olds					
Unadjusted mean	-0.014	0.014	-0.028	0.017	.093
Adjusted by screener race			-0.028	0.017	.093
Adjusted by screener race, FI characteristics,† and geographic variables			-0.018	0.020	.358
35-49 year olds					
Unadjusted mean	0.001	-0.001	0.002	0.017	.901
Adjusted by screener race			0.002	0.017	.900
Adjusted by screener race, FI characteristics,† and geographic variables			-0.002	0.019	.923
50+ year olds					
Unadjusted mean	-0.017	0.017	-0.034	0.018	.060
Adjusted by screener race			-0.034	0.018	.060
Adjusted by screener race, FI characteristics,† and geographic variables			-0.052	0.020	.009[a]

†FI = field interviewer. FI characteristics include experience.
[a] Difference is statistically significant at the .05 level.

female response rate. No adjustment can be made for the reduction in non-contacts that would have occurred by selecting readily available males because there is no record of who is present in the dwelling unit during the screening.[3]

[3] The additional number of selected females is estimated using the 1998 selection rates ***(Table 3.11)*** and the 1999 screening data. The additional number of completed females is estimated using the actual 1999 female response rate by age.

Table 3.14 Adjusted 1999 Weighted Interview Response Rates Using 1998 Convenient-Selection Inflation Factor (%)

Age Groups	1999 Rate	Adjusted Rate	Difference
CAI			
12-17	78.1	78.1	0.00
18-25	71.2	71.2	0.01
26-34	69.5	69.6	0.14
35-49	67.8	68.1	0.33
50+	64.6	64.7	0.07
Total	68.6	68.7	0.14
PAPI			
12-17	76.3	76.2	-0.03
18-25	67.8	67.8	0.06
26-34	65.0	65.1	0.08
35-49	64.3	64.5	0.24
50+	65.7	65.8	0.05
Total	66.6	66.7	0.10

The adjusted weighted response rates are better than the unadjusted. That is, the response rates would have been higher in 1999 if the Newton had not been used for screening. The CAI rate would have been only slightly higher (68.68 vs. 68.55 percent). The PAPI rate would have been higher by about one-tenth of a percent (66.65 vs. 66.55 percent). Although the impact of the convenient-selection factor was very large in some cases, the impact on response rates was fairly small for the full survey. For example, the selection rate for females aged 35 to 49 years old was significantly higher in 1998 (***Table 3.11***) than would be expected by the sample design, but the change in overall response rate for that age group was only one-third of a percentage point (0.33 percent). This is because the opportunity for convenient selection occurs only in the households with more than one person in a selected age group. This represents a small share of the overall sample and therefore has little impact on the overall response rates.

3.4 Evaluation of Difference in Response Rate Patterns Between CAI and PAPI Interview Modes

Another major change to the data collection design in 1999 was the transition from a PAPI survey instrument to a CAI. The majority of the sample was worked using the CAI instrument, but a portion was collected with the PAPI design. There were 66,706 interviews collected with the

CAI and 13,809 with the PAPI. The primary purpose of the split-sample design was to permit comparison of estimates between 1998 and 1999. The split sample also provided data for an analysis of the response rates by mode.

3.4.1 Mode-Effect Hypothesis

It is possible that some subgroups in the population would be more responsive to a computerized instrument than to a paper instrument. For example, it is reasonable to expect younger or better educated persons to be more comfortable using a computer than older or less-well-educated persons. These characteristics aside, the isolated impact of the use of CAI may encourage respondents to cooperate due to an increased appearance of confidentiality and professionalism as a result of the computerized instrument (Lessler & O'Reilly, 1997). This section evaluates the difference between the two modes while controlling for the environmental, respondent, and interviewer characteristics in ***Table 3.4***. The mode-effect hypothesis evaluated in this section states

Response rates are higher for the CAI than the PAPI, when controlling for the set of environmental, respondent, and field interviewer characteristics from ***Table 3.4.***

3.4.2 Regression Analysis

Thirteen weighted linear regression models were fitted to evaluate the mode-effects hypothesis.[4] These regressions are presented in ***Table 3.15***.[5] The analysis weights include all dwelling-unit level weights and person-level design weight components. These models were run against the full set of selected CAI and PAPI cases for 1999, coded as either a completed interview or not (0,1). The first model contained only a constant and an

---

[4]The models were run using weighted linear regression and weighted logistic regression. Although the logistic models are more appropriate for the dichotomous dependent variables, the linear models produce estimates that can be interpreted as response rates. The linear and the logistic models yield the same results for the hypothesis tests. The linear models are reported here for ease of interpretation.

[5]The SUDAAN® (Survey Data Analysis) software package, developed by RTI, was used to fit the models. The sample structure was recognized using standard NHSDA analysis NEST statements for variance strata and variance replicates. SUDAAN is a registered trademark of RTI (© 2002, Research Triangle Institute).

Table 3.15 Mode Effects for Interview Response Rate under Alternative Models—1999 CAI vs. 1999 PAPI—Weighted Linear Regression

Model		Interview Response Rate		Difference between PAPI and CAI		
		CAI	PAPI	Estimate	SE	p-Value
1	Unadjusted mean	68.6	66.6	-2.0	1.1	.062
	Unadjusted mean by age					
2.1	12-17	78.1	76.3	-1.8	1.3	.162
2.2	18-25	71.2	67.8	-3.4	1.3	.009[a]
2.3	26-34	69.4	65.0	-4.5	1.5	.004[a]
2.4	35+	66.1	65.0	-1.0	1.4	.447
	Unadjusted mean by gender					
3.1	Male	67.1	64.6	-2.5	1.4	.069
3.2	Female	69.8	68.3	-1.5	1.1	.187
	Unadjusted mean by experience					
4.1	No experience	68.3	65.2	-3.1	1.1	.005[a]
4.2	Previous NHSDA experience	69.1	71.0	-1.9	2.1	.361
5	Adjusted for demographics only			-2.0	1.0	.045[a]
6	Adjusted for prior experience and demographics			-1.8	1.0	.056
7	Adjusted for prior experience, screener order, and demographics			-1.8	1.0	.056
8	Adjusted for prior experience, screener order, interviewer demographics, and respondent demographics			-1.9	0.9	.045[a]

[a]Difference is statistically significant at the .05 level.

indicator for the PAPI cases (PAPI =1). The parameter estimate for the constant (68.6) is equal to the response rates for the 1999 CAI; the parameter estimate for the indicator of PAPI is equal to the difference between the PAPI and CAI response rates (-2.0), and the standard error of the estimate (1.1) can be used to demonstrate the significance of the difference ($p < .062$).

Models 2.1 through 4.2 repeat this process of producing unadjusted means for the listed subsets on the independent variables. For example, the response rate for 12-17 year olds in CAI was 78.1 percent; the rate for PAPI for the same age group was 1.8 percentage points lower. Models 5 through 8 are estimated against the full set of selected and eligible cases for CAI and PAPI while controlling for interviewer experience and some of the known correlates of nonresponse. These models are adjusted for a number of variables, so only the estimate of the difference between the two years given the independent variables is available.

The results of the weighted linear regression models in ***Table 3.15*** can be used to evaluate the mode-effect hypothesis. The difference estimate in the base model indicates that the response rate for the CAI was higher by about 2 percentage points, but that the difference was not significant at $p < .05$ although it was very close. This pattern holds for most of the sub-group analyses. The fully specified model (Model 8) shows a marginally significant and negative impact of PAPI on response rates. ***Table 3.15*** supports the expectations of the mode-effect hypothesis; in general, the CAI yielded a higher response rate when controlling for the set of environmental, respondent, and field interviewer characteristics from ***Table 3.4***.

Table 3.16 contains the parameter estimates used to generate the difference in Model 8 of ***Table 3.15***. This table provides some interesting context for the role that each of the independent variables plays in explaining the response rates for the two modes. As expected, interviewers with prior experience on the NHSDA had an expected response rate that was 4.7 percentage points higher than those with no prior experience. On-the-job experience was not significant in this model. This is likely a result of including in the model the quarter in which the survey was conducted because the bulk of new interviewers began the project in the first quarter.[6]

3.5 Evaluation of Factors Contributing to Decline in Response Rate from 1998 to 1999

The models presented in ***Sections 3.3*** and ***3.4*** evaluate the impact of changes in the study design in 1999 on the response rates. ***Section 3.3*** demonstrated that the transition to the Newton improved the accuracy of the screening process and, as a result, lowered the response rates. However, the magnitude of the impact on response rates was very small and does not explain the large decline in 1999. ***Section 3.4*** demonstrated that PAPI response rates were about 2 points lower than CAI in 1999 when controlling for a set of environmental, respondent, and interviewer characteristics. Therefore, it is unlikely that the transition from PAPI to CAI led to the large decline in the 1999 response rates. This section compares the response rates between the two years for PAPI only while controlling for the set of environmental, respondent, and interviewer characteristics used in the previous models.

[6] On-the-job experience was significant when Model 8 was estimated without the quarter variable.

Table 3.16 1999 CAI vs. 1999 PAPI OLS—Full Model Coefficients

Factor	Level	Estimate	SE	p-Value
	Intercept	0.707	0.028	.000
Survey Design				
Mode	PAPI	-0.019	0.009	.045[a]
Environment				
Region (vs. West)	Northeast	-0.025	0.015	.106
	North Central	0.042	0.014	.003[a]
	South	0.010	0.015	.486
Urbanicity (vs. NonMSA Rural)	MSA ≥=1 mil	-0.090	0.015	.000[a]
	MSA 250K-999K	-0.054	0.015	.000[a]
	MSA < 250K	-0.015	0.017	.371
	NonMSA nonRural	0.030	0.016	.059
Quarter conducted (vs. Q4)	Q1	-0.069	0.016	.000[a]
	Q2	0.002	0.012	.881
	Q3	0.011	0.011	.341
Respondent Characteristics				
Age (vs. 35+)	12-17	0.098	0.009	.000[a]
	18-25	0.036	0.008	.000[a]
	26-34	0.016	0.010	.121
Gender	Male	-0.036	0.001	.000[a]
Age × gender	Male 12-17	0.023	0.012	.056
	Male 18-25	-0.008	0.012	.500
	Male 26-34	-0.005	0.015	.715
Race/Ethnicity (vs. else)	Hispanic	0.077	0.012	.000[a]
	Black	0.037	0.011	.000[a]
Interviewer Characteristics				
NHSDA experience	Yes	0.047	0.010	.000[a]
On-the-job experience (# of completed screenings) (vs. 300+)	0-59	0.010	0.015	.507
	60-119	-0.020	0.017	.218
	120-179	-0.001	0.015	.944
	180-299	-0.004	0.014	.746
Race (vs. white)	Am. Indian	-0.107	0.044	.014[a]
	Asian	0.036	0.084	.674
	Black	0.006	0.024	.796
	Other	-0.002	0.020	.924
Age (vs. 60+)	Missing	0.024	0.021	.260
	0- 40	0.011	0.015	.478
	41-50	0.005	0.013	.686
	51-60	0.033	0.013	.014[a]
Gender	Male	-0.027	0.010	.008[a]

[a]Indicates significance at $p < .05$.

One of the primary explanations for the decline in the response rates between 1998 and 1999 is derived from the large increase in the sample size. The increase resulted in a large number of inexperienced interviewers working the 1999 survey. This section evaluates the change in response rates between 1998 and 1999 while controlling for the level of interviewer experience and a set of known correlates of nonresponse (Groves & Couper, 1998; RTI, 2000). It is expected that, if as many as possible of the known changes between 1998 and 1999 are controlled, the difference in response rates between the two years will be small and insignificant.

3.5.1 Regression Analysis

As was done for the mode-effect hypothesis, 13 weighted linear regression models were fitted to evaluate the trend (see Footnote 4). These regressions are presented in ***Table 3.17*** (see Footnote 5). The analysis weights include all dwelling-unit-level weights and person-level design weight components. The models were run against the full set of selected PAPI cases for 1998 and 1999, coded as either a completed interview or not (0,1). The first model contained only a constant and an indicator for the cases in the 1999 survey. The estimate for the constant (76.0) is equal to the response rate for the 1998 PAPI. The parameter estimate for the indicator of 1999 is equal to the difference between 1998 and 1999 PAPI response rates (-9.5). The standard error of the estimate (1.2) can be used to demonstrate the significance of the difference ($p < .05$).

Models 2.1 through 4.2 repeat this process of producing unadjusted means for the listed subsets on the independent variables. For example, the response rate for youths aged 12 to 17 in 1998 was 81.3 percent; the rate in 1999 for the same age group was 5.1 percentage points lower. Models 5 through 8 are estimated against the full set of selected and eligible PAPI cases for 1998 and 1999 while controlling for interviewer experience and some of the known correlates of nonresponse. Because these models are adjusted for a number of variables, only the estimate of the difference between the two years given the independent variables is available.

The results of the weighted linear regression models in ***Table 3.17*** can be used to evaluate the degree to which the regression models explain the decline in response rates from 1998 to 1999. If the set of known correlates fully captured the relevant changes between 1998 and 1999, then the difference estimate would not be significant. The negative and significant difference between 1998 and 1999 response rates exists for all of the subcategories on the independent variables. Furthermore, the difference remains negative and significant in the multivariate models, including the most fully specified Model 8. Model 8 shows that the expected response rate for 1999 is five points (-5.0) lower than 1998, and the difference is significant beyond the .000 level.

Table 3.17 Trend Effects for Interview Response Rate under Alternative Models—1998 PAPI vs. 1999 PAPI—Weighted Linear Regression

Model		1998	1999	Difference between 1998 and 1999 Estimate	SE	p-Value
1	Unadjusted mean	76.0	66.6	-9.5	1.2	.000[a]
	Unadjusted mean by age					
2.1	12-17	81.3	76.3	-5.1	1.5	.001[a]
2.2	18-25	76.9	67.8	-9.1	1.5	.000[a]
2.3	26-34	75.1	65.0	-10.1	1.7	.000[a]
2.4	35+	75.2	65.0	-10.1	1.5	.000[a]
	Unadjusted mean by gender					
3.1	Male	73.5	64.6	-8.9	1.5	.000[a]
3.2	Female	78.1	68.3	-9.8	1.2	.000[a]
	Unadjusted mean by experience					
4.1	Previous NHSDA exp.	78.0	71.0	-6.9	2.1	.001[a]
4.2	No previous NHSDA exp.	70.3	65.2	-5.1	1.8	.004[a]
5	Adjusted for demographics only			-9.5	1.1	.000[a]
6	Adjusted for prior experience and demographics			-5.7	1.3	.000[a]
7	Adjusted for prior experience, screener order, and demographics			-5.0	1.3	.000[a]
8	Adjusted for prior experience, screener order, interviewer demographics, and respondent demographics			-5.0	1.3	.000[a]

[a]Difference is statistically significant at the .05 level.

This suggests that the decline in response rates in 1999 cannot be attributed entirely to the decrease in the level of experience in the interviewing staff. Furthermore, this suggests that there are important differences between the two years that affect response rates that are not captured by the measures of interviewer experience, interviewer demographics, urbanicity, and respondent demographics.

Table 3.18 contains the parameter estimates used to generate the difference in Model 8 of ***Table 3.17***. This table provides some interesting context for the role that each of the independent variables plays in explaining the response rates for the two years.

As expected, the level of FI experience had a positive and significant relationship with response rates. This held for the measure of past NHSDA experience and on-the-job experience. Interviewers with prior experience on the NHSDA had expected response rates that were 6.3 percentage points higher than those with no prior experience. Interviewers with little on-the-job experience had a lower expected response rate than those with a great deal of experience. The very new interviewers (0-59 completed screenings) had an expected rate 5.5 percentage points lower than the most experienced, and the fairly new (60-119 completed screenings) had an expected rate 4.1 percentage points lower.[7]

The controls performed as expected. Sample members in the Northeast were significantly less likely to complete the survey than were those in the West, and those from medium and large cities were also less likely to complete the survey than those in rural regions. This supports the prior findings about the role of urbanicity in response rates. Younger sample members were more likely to complete the interview, males were less likely (except for young males), and Hispanics and blacks were more likely than the baseline (whites and others).

There was little relationship between the age or gender of the interviewers and response rates. Interviewers with past NHSDA experience had a higher response rate than those who had never worked on the NHSDA. Interviewers with little on-the-job experience had significantly lower response rates than those with extensive experience.

The decrease in interviewer experience that resulted from the increase in sample size does not fully explain the decline in response rates between 1998 and 1999, even while controlling for other known correlates of nonresponse. The decrease in the aggregate experience levels did reduce the response rates in 1999; however, some differences between the two years remain unexplained. Many factors that explain response rates could not be included in the model due to measurement problems.

3.6 Summary and Conclusions

The design changes between 1998 and 1999 corresponded with a large decrease in response rates. A series of management efforts was taken to address the decrease, both in anticipation of the design changes and in reaction to unexpected results of the changes. In general, the efforts were successful.

---

[7] The new interviewers were more equally distributed across quarters in 1998 than in 1999. This resulted in a lower correlation between on-the-job experience and quarter of the interview than found in the CAI-PAPI comparison models.

Table 3.18 1998 PAPI vs. 1999 PAPI OLS—Full Model Coefficients

Factor	Level	Estimate	SE	p-Value
	Intercept	0.710	0.034	.000
Study Design				
Study year	1999	-0.050	0.013	.000[a]
Environment				
Region (vs. West)	Northeast	-0.031	0.017	.064
	North Central	0.055	0.017	.001[a]
	South	0.026	0.016	.102
Urbanicity (vs. NonMSA rural)	MSA ≥=1 mil	-0.067	0.018	.000[a]
	MSA 250K-999K	-0.036	0.018	.049[a]
	MSA < 250K	0.001	0.021	.954
	NonMSA nonRural	0.023	0.019	.224
Quarter conducted (vs. Q4)	Q1	0.006	0.018	.749
	Q2	0.037	0.013	.005[a]
	Q3	0.020	0.013	.137
Respondent Characteristics				
Age (vs. 35+)	12-17	0.063	0.010	.000[a]
	18-25	0.015	0.010	.163
	26-34	0.000	0.012	.987
Gender	Male	-0.046	0.010	.000[a]
Age × gender	Male 12-17	0.041	0.014	.005[a]
	Male 18-25	0.003	0.014	.854
	Male 26-34	-0.005	0.018	.779
Race/Ethnicity (vs. else)	Hispanic	0.069	0.012	.000[a]
	Black	0.042	0.011	.000[a]
Interviewer Characteristics				
NHSDA experience	Yes	0.063	0.012	.000[a]
On-the-job experience (# of completed screenings) (vs. 300+)	0-59	-0.055	0.017	.002[a]
	60-119	-0.041	0.017	.017[a]
	120-179	-0.021	0.016	.197
	180-299	-0.008	0.013	.549
Race (vs. white)	Am. Indian	-0.101	0.046	.030[a]
	Asian	0.031	0.109	.780
	Black	-0.005	0.030	.869
	Other	0.013	0.024	.576
Age (vs. 60+)	Missing	0.026	0.029	.375
	0- 40	0.016	0.019	.389
	41-50	0.007	0.014	.629
	51-60	0.029	0.015	.057
Gender	Male	-0.009	0.012	.448

[a] Indicates significance at $p < .05$.

The extended analysis presented in this chapter summarizes the current understanding of the decline in response rates in 1999. First, it appears that the previous understanding of the correlates of nonresponse was correct, but it does not completely explain the difference between 1998 and 1999. Second, management efforts taken during 1999 appear to have been successful in diminishing the decline in response rates, and this success has carried over to 2000. Third, the computerized screening instrument reduced sampling bias by removing interviewer effects from the screening routine, and this had a small and negative impact on response rates. Fourth, the transition from PAPI to CAI increased the response rates. Finally, much of the decline in 1999 can be attributed to changes in the composition of the field staff resulting from the large increase in sample size, most notably the reduced number of experienced field interviewers working on the project. However, this does not fully explain the decline in 1999.

3.7 References

Caspar, R. (1992). *Followup on nonrespondents in 1990* (DHHS Pub. No. ADM 92-1929). Rockville, MD: National Institutes of Health.

Couper, M.P., & Groves, R.M. (1996). Social environmental impacts on survey cooperation. *Quality and Quantity*, *30*, 173-188.

Gfroerer, J., Lessler, J.T., & Parsley, T. (1997). *Studies of nonresponse and measurement error in the National Household Survey on Drug Abuse*. NIDA Research Monograph 167. Rockville, MD: National Institutes of Health.

Groves, R.M., & Couper, M.P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley & Sons.

Huisman, M., Krol, B., & Van Sonderen, E. (1998). Handling missing data by re-approaching non-respondents. *Quality and Quantity, 32*, 77-91.

Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics, 15*, 269-282.

Lessler, J.T., & O'Reilly, J.M. (1997). *Mode of interview and reporting of sensitive issues: Design and implementation of audio computer-assisted self-interviewing*. NIDA Research Monograph 167, 366-383. Rockville, MD: National Institutes of Health.

RTI. (2000). *1999 NHSDA nonresponse analysis.* Rockville, MD: Substance Abuse and Mental Health Services Administration.

Snijkers, G., Hox, J., & De Leeuw, E. (1999). Interviewers' tactics for fighting survey nonresponse. *Journal of Official Statistics*, *15*, 185-198.

Substance Abuse and Mental Health Services Administration. (1996). *The development and implementation of a new data collection instrument for the 1994 NHSDA* (DHHS Pub. No. SMA 96-3084). Rockville, MD: Author+.

4

Assessment of the Computer-Assisted Instrument

Rachel Caspar and Michael Penne

The conversion of the National Household Survey on Drug Abuse (NHSDA) to computer-assisted interviewing (CAI) offered an opportunity to improve the quality of the data collected in the NHSDA in a number of ways. Some of these improvements were implemented easily and manifested themselves in more complete data—for example, the ability to eliminate situations where questions were inadvertently left blank by the respondent. However, other improvements could only be realized through careful development and implementation of new procedures. Thorough testing was needed to determine whether these new procedures did, in fact, result in higher quality data.

In this chapter, two significant revisions to how key NHSDA data items are collected and the effect of these revisions on the quality of the data obtained in the 1999 NHSDA are described. The first of these revisions was the addition of a methodology for resolving inconsistent or unusual answers provided by the respondent. This methodology was incorporated into the collection of a large number of the data items that are considered critical to the reporting needs of the NHSDA. The second revision dealt specifically with the way data on frequency of substance use over the past 12-month period was reported. This chapter also provides a review of several basic measures of data quality including rates of Don't Know and Refused responses, breakoff interviews, and the observational data provided by the interviewers at the conclusion of each interview. Where possible, these measures are compared between the CAI and paper-and-pencil interview (PAPI) NHSDA instruments as a means of assessing the effect of the move to CAI on data quality.

4.1 Respondent′s Ability to Verify Unusual Responses and Resolve Inconsistent Data

4.1.1 Background

One potential benefit of converting the NHSDA to a computer-assisted format was the opportunity to resolve inconsistent or unusual responses at the time of interview rather than editing the data after the fact. The self-administered answer sheet methodology used in the PAPI NHSDA had been shown to improve reporting of illegal substance use (Turner, Lessler, & Gfroerer, 1992). However, to preserve the respondent′s privacy, interviewers did not review the answer sheets for completeness prior to leaving the respondent′s home and mailing the PAPI questionnaire back to RTI. Thus, unless the respondent specifically asked for help, the interviewer was often unaware that the respondent was having difficulty. To fulfill the promise of full confidentiality to respondents, no additional contacts were made to resolve problems identified once the questionnaires arrived at RTI. This resulted in a need for significant post-data-collection editing to deal with response patterns that were inconsistent, incomplete, or unusual.

The conversion to audio computer-assisted self-interviewing (ACASI) was expected to significantly reduce the need for editing after the fact because the computer, rather than the respondent, controlled the routing through the questionnaire. This ensured that all applicable questions were presented to the respondent. In addition, because the program required some type of response to each applicable question, missing data were expected to be reduced. Yet the possibility of inconsistent and unusual responses remained, as did the potential for missing data (i.e., Don′t Know and Refused). Although the program could be easily adapted to identify such responses, it was less clear how to resolve the responses with the respondent. To maintain the privacy benefits of the ACASI component of the interview, any resolution methodology had to be simple enough for the respondent to complete without significant intervention by the interviewer. Because there was no existing research on the feasibility of respondents resolving inconsistencies in their data as part of a self-administered computer-assisted interview, work was undertaken to develop such a methodology as part of the NHSDA conversion work.

4.1.2 Development of Resolution Methodology

A resolution methodology was developed that combined three stages. At the first stage, the CAI instrument was programmed to identify inconsistent or unusual responses. At the second stage, respondents were asked to verify that the last response entered was correct. So, for example, when a respondent who had earlier reported her current age as 20 reported that she was 51 the first time she used marijuana (two clearly inconsistent answers), the computer was programmed to prompt the respondent to verify that the last answer entered was correct. This step was included to eliminate inconsistencies that could be due to keying errors. If the respondent indicated that the answer was incorrect, she was routed to a screen where she could answer the original question about age of first marijuana use again (perhaps this time entering the age as 15). In this case, the inconsistency was resolved. If, however, the respondent verified her age at first use of marijuana (51) as correct, the third stage of the process was implemented. At this stage, the respondent had the opportunity to resolve the inconsistent responses (in this example, the respondent's age, which had been recorded as 20, and the age of first marijuana use which had been recorded as 51). The computer routed the respondent to a screen where the two inconsistent answers were presented and the respondent had the opportunity to correct one or both entries. ***Figure 4.1*** describes the three stages of the resolution methodology for inconsistent responses.

In the example described above, the respondent provided two answers that were clearly inconsistent with one another. The methodology worked similarly when the respondent provided what could be considered an "unusual" response. Two checks were incorporated for such unusual responses. Based on input from staff responsible for editing the PAPI NHSDA data, it seemed possible that some respondents might be misreading the question that asked about their age at first use of a substance and instead reporting their current age. Thus, any response to an age-at-first-use question where the age reported was equal to the respondent's reported current age was verified. In addition, any age-at-first-use question where the age entered was less than 10 was also verified. This was done for two reasons:

- For many of the substances asked about in the NHSDA, first use at an age younger than 10 would be rather unusual.
- A single-digit response could indicate that the respondent had pressed the [Enter] key too quickly (resulting in a response of 2 instead of 25, for example).

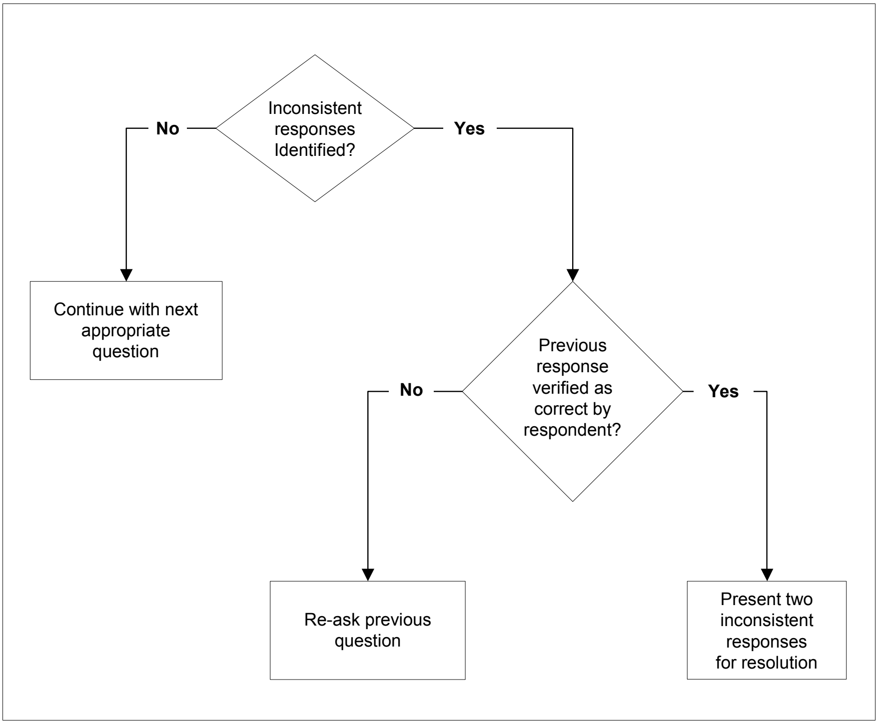

Figure 4.1 Resolution methodology flowchart for inconsistent responses.

Regardless of the type of unusual response recorded (either equal to current age or less than 10), the resolution methodology worked the same way. Once an unusual response was identified, the respondent was asked to verify the answer. If the respondent indicated the answer was correct, he/she was routed to the next applicable question. If the respondent indicated the answer was incorrect, he/she was routed to a screen where the original question was presented again.

In developing the actual text of the verification and resolution screens, several features were incorporated that were deemed to be important in aiding respondent comprehension:

- Original responses that were identified as inconsistent were displayed on screen for the respondent to enhance recall and comprehension.
- Resolution screens were worded so as not to explicitly place the responsibility for the inconsistency on the respondent (e.g., "your answers"), but rather to imply that the computer might have been in error (e.g., "the computer recorded").

- Respondents were asked to identify the incorrect response when two items were inconsistent to facilitate the flow of questioning.
- When the respondent indicated that both answers were incorrect, the respondent was re-asked the two items in the same order in which they were originally presented to maintain consistency.
- Inconsistencies were identified only between items that appeared close to each other in the interview to minimize respondent confusion.

The resolution methodology was first tested in the cognitive laboratory. Using three rounds of testing, screen text was developed that was understandable to a majority of the subjects (SAMHSA, 2001a). The primary changes made as a result of the laboratory testing were to scale back the amount of text on the inconsistency resolution screens significantly and to ask the respondent to identify the correct item when two items were found to be inconsistent.

The final format of the resolution methodology for inconsistent responses is shown below. In this example, the screen text is shown for a situation in which a respondent has reported drinking alcohol on more days in the past 30 days than in the past 12 months:

The answers for the last question and an earlier question disagree.
Which answer is correct?

1	I drank alcohol on [*x*] days in the past 12 months.
2	I drank alcohol on [*y*] days in the past 30 days.
3	Neither answer is correct.

Note that the values of *x* and *y* were generated by the computer program based on the respondent's answers to the earlier questions. Depending on the answer selected, the respondent was then routed to the appropriate follow-up question(s); in this case, either the number of days he/she drank alcohol in the past 12 months if response 2 was selected, the number of days he/she drank alcohol in the past 30 days if response 1 was selected, or to both questions if response 3 was selected.

The verification process was even more straightforward. In the example below, the verification text is shown for a situation where the respondent reports his/her age of first heroin use as less than 10 (again, the value of *x* was generated by the computer program based on the respondent's answer to the original question):

The computer recorded that you were [*x*] years old the first time you used heroin. Is this correct?

1 Yes
2 No

The respondent was either routed on to the next applicable question if the answer to the verification question was yes or re-asked the original question about his/her age the first time he/she used heroin if the answer was no.

The next step in the development of the resolution methodology was to test the procedure within the larger NHSDA questionnaire and in a realistic field setting. This was done as part of the 1997 CAI Field Experiment. The resolution methodology was included as one of three experimental treatments in this study with respondents randomly assigned to either receive the resolution items or not. Interviewing was completed during the fourth quarter of 1997. A total of 1,982 respondents were interviewed. Briefly, the results of the field test showed the following (SAMSHA, 2001a):

- Among the respondents assigned to the resolution treatment, 28 percent triggered at least one resolution item; of those respondents, just over half (55 percent) triggered a verify check as opposed to an inconsistency check.
- Youth respondents (those aged 12 to 17) accounted for approximately 60 percent of the inconsistency checks triggered.
- The average number of inconsistency checks triggered during an interview was slightly greater than 1 (1.2). The average number of verify checks triggered during an interview was 1.2 as well.

Of particular interest was whether the consistency of the data provided by respondents was improved when the resolution methodology was used. In analyzing the data from the field test, it was determined that only 17 percent of the data items (n = 57) remained inconsistent following the resolution task.

Two operational issues related to the use of the resolution methodology were also examined: breakoff rates and interview length. No differences were found in the breakoff rates for respondents assigned to the resolution treatment when compared with those who were not assigned to that treatment. Similarly, there were no differences found in interview length when the resolution methodology was added.

4.1.3 Results from 1999 NHSDA

Due to the limited sample size for the 1997 CAI Field Experiment (n = 1,982), demographic subgroup analyses were not possible. However, the 1999 NHSDA, with its significantly larger sample size afforded us the ability to look at the relative improvements in data consistency within subgroups as well as overall. The following inconsistency checks were included in the 1999 NHSDA:

- 30-day frequency of use greater than 12-month frequency of use (included for alcohol, marijuana, cocaine, "crack," heroin, hallucinogens, and inhalants)
- Zero days used in the past 30 days for respondents who reported the past 30 days as their time of most recent use (included for alcohol, marijuana, cocaine, "crack," heroin, hallucinogens, and inhalants)
- Age at first use greater than respondent's current age (included for cigarettes, chewing tobacco, snuff, cigars, alcohol, marijuana, cocaine, "crack," heroin, hallucinogens, inhalants, stimulants, sedatives, tranquilizers, and pain relievers)
- Age of first cigarette smoking greater than age of first smoking cigarettes daily
- Number of days consumed five or more drinks of alcohol on the same occasion during the past 30 days greater than the 30-day frequency of use for alcohol
- Last use of methamphetamine more recent than the respondent's last use of any stimulant
- Last use of LSD more recent than the respondent's last use of any hallucinogen
- Last use of PCP more recent than the respondent's last use of any hallucinogen.

In addition, two checks for unusual responses were included:

- Age at first use is less than 10 (included for cigarettes, chewing tobacco, snuff, cigars, alcohol, marijuana, cocaine, "crack," heroin, hallucinogens, inhalants, stimulants, sedatives, tranquilizers, and pain relievers)

- Age at first use is equal to the respondent's current age (included for cigarettes, chewing tobacco, snuff, cigars, alcohol, marijuana, cocaine, "crack," heroin, hallucinogens, inhalants, stimulants, sedatives, tranquilizers, and pain relievers).

Two features of the check items are important to note. First, as with the 1997 CAI Field Experiment, respondents were routed through an inconsistency check only once. If their responses remained inconsistent following the check, no additional attempt was made to resolve the inconsistency. However, it was possible for the answers provided during an inconsistency check to subsequently trigger an unusual data check (e.g., when the respondent provided an age less than 10 to an age-of-first-use question during the inconsistency check).

To begin, ***Table 4.1*** presents data for the respondents who triggered at least one inconsistency check. Results are grouped by the type of inconsistency triggered: (1) an inconsistency related to the respondent's answer to a 30-day frequency-of-use question (15 possible check items), (2) an inconsistency related to the respondent's recency of use for LSD, PCP, or methamphetamine (three possible check items), and (3) an inconsistency related to the respondent's answer to an age-of-first-use question (16 possible check items). The data are presented for the total sample and then split out by age, race/ethnicity, gender, and education. It is important to note that triggering a check item is necessarily correlated with substance use since a respondent who has never used a particular substance will not encounter any of the check items associated with the questions regarding that substance. Thus, the rates of triggering check items could be higher for subgroups where substance use is more prevalent.

The results in ***Table 4.1*** show that inconsistency checks were triggered by only 3.4 percent of the NHSDA respondents. Although no direct comparisons to the PAPI data are available, this finding suggests that the quality of data provided by respondents via the computer-assisted NHSDA questionnaire is quite good. The total number of checks triggered (2,374) is only slightly greater than the number of respondents who triggered at least one inconsistency (2,281), indicating that the vast majority of respondents provided only one inconsistent response. One specific check item, the comparison of the number of days in the past 30 days that a respondent "binged" on alcohol (had five or more drinks on the same occasion) as compared with the number of days he/she drank alcohol at all

Table 4-1. Demographic Characteristics of 1999 NHSDA Respondents Who Triggered at Least One Inconsistency Check†

				30-Day Frequency of Use								Age of First Use			
	Total 1999 Sample	Total Over All Checks		30-Day Use, 0 Days		30-Day Use > Past Year Use		30-Day Binge > 30-Day Alc††		Recency Logic		AFU > Current Age		Age Smoked Daily < AFU††	
Domains	**N**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Total respondents	66,706	2,281	3.42	728	1.09	179	0.27	1,009	1.51	172	0.26	98	0.15	152	0.23
Age group‡															
12-17	25,357	675	2.66	269	1.06	61	0.24	241	0.95	57	0.22	42	0.17	25	0.10
18-25	21,933	1,019	4.65	277	1.26	69	0.31	527	2.40	97	0.44	33	0.15	45	0.21
26-34	7,878	251	3.19	71	0.90	25	0.32	127	1.61	3	0.04	9	0.11	17	0.22
35-49	6,246	204	3.27	64	1.02	12	0.19	87	1.39	13	0.21	10	0.16	22	0.35
50+	5,292	132	2.49	47	0.89	12	0.23	27	0.51	2	0.04	4	0.08	43	0.81
Race / ethnicity‡															
Hispanic	8,481	353	4.16	105	1.24	33	0.39	167	1.97	19	0.22	13	0.15	27	0.32
Non-Hisp black	8,171	288	3.52	101	1.24	35	0.43	115	1.41	6	0.07	17	0.21	22	0.27
Non-Hisp non-black	50,054	1,640	3.28	522	1.04	111	0.22	727	1.45	147	0.29	68	0.14	103	0.21
Gender‡															
Male	32,092	1,367	4.26	406	1.27	96	0.30	638	1.99	104	0.32	70	0.22	90	0.28
Female	34,614	914	2.64	322	0.93	83	0.24	371	1.07	68	0.20	28	0.08	62	0.18
Education‡,§															
Less than high school	7,458	456	6.11	102	1.37	35	0.47	232	3.11	41	0.55	17	0.23	49	0.66
High school	14,845	618	4.16	182	1.23	47	0.32	285	1.92	44	0.30	16	0.11	55	0.37
Greater than high school	19,046	532	2.79	175	0.92	36	0.19	251	1.32	30	0.16	23	0.12	23	0.12
Total Triggered Checks among All Respondents#		2,374		754		180		1,009		178		101		152	

† Percentages are based on the total 1999 Respondent Sample.
†† Alc = alcohol; AFU = age of first use.
‡ Cochran-Mantel-Haenszel χ^2 Statistics indicate a significance at the 0.001 level across domain categories.
§ Education includes only respondents 18 years and older.
# Sample sizes of total number of triggered checks greater than the total number of respondents indicate that some respondents triggered multiple checks.

accounted for almost half of the inconsistencies. To trigger the item, a respondent would have to enter a larger number of "binge" days than drinking days. Given that this item was triggered by 1,009 respondents (44 percent of the respondents who triggered an inconsistency check) and accounted for 42 percent of the total number of checks triggered, it seems likely that these two questions, and the implied relationship between them, were not adequately understood by respondents. The ability to identify the inconsistent data and bring it to the respondent′s attention for resolution before the interview is completed is the strength of this methodology.

Another type of 30-day frequency of use check also accounted for a significant proportion of the inconsistency checks triggered in the 1999 NHSDA. For seven of the substance types (alcohol, marijuana, cocaine, "crack," heroin, hallucinogens, and inhalants), an inconsistency check was included to identify situations where a respondent reported using a substance during the past 30 days but then recorded a response of zero to the question that asked for the frequency of use during that time period. ***Table 4.1*** shows that this check was triggered by 728 respondents (32 percent of the respondents who triggered an inconsistency check) and accounted for 32 percent of the total number of checks triggered. The remaining inconsistency checks were triggered by relatively few respondents.

Population groups most likely to trigger inconsistency checks included young adults (aged 18 to 25) and males. As these demographic groups are more likely to use substances (SAMHSA, 2001), their chance of triggering a check item is increased. Hispanic respondents were somewhat more likely than other racial/ethnic groups to trigger these checks, although rates of substance use are generally not higher for Hispanics. Finally, rates of inconsistent reporting decrease with increasing levels of education.

Table 4.2 provides similar information on the verification items that were triggered. The most obvious finding from these data is that many more respondents triggered verification items than inconsistency items. Nearly a fifth of the respondents (17 percent) triggered at least one verification item. For the total sample, the two types of verification items were triggered with similar frequency. Respondents were more likely to trigger multiple verification items during the course of their interview; the average number of verification checks triggered was 1.4.

Not surprisingly, youth respondents were more likely to trigger a verification check. ***Table 4.2*** shows that youth respondents were more likely to trigger the check associated with their age of first use being the

Table 4.2 Demographic Characteristics of 1999 NHSDA Respondents Who Triggered at Least One Verification Check†

	Total 1999 Sample	Total Over All Verifications		Age of First Use = Current Age of Respondent		Age of First Use <10 Years of Age	
Domains	**N**	**N**	**%**	**N**	**%**	**N**	**%**
Total respondents	66,706	11,147	16.71	6,366	9.54	5,433	8.14
Age group‡							
12-17	25,357	5,781	22.80	4,085	16.11	2,186	8.62
18-25	21,933	3,684	16.80	2,052	9.36	1,821	8.30
26-34	7,878	762	9.67	154	1.95	623	7.91
35-49	6,246	530	8.49	60	0.96	476	7.62
50+	5,292	390	7.37	15	0.28	377	7.12
Race / ethnicity‡							
Hispanic	8,481	1,323	15.60	828	9.76	565	6.66
Non-Hisp black	8,171	1,170	14.32	627	7.67	602	7.37
Non-Hisp non-black	50,054	8,654	17.29	4,911	9.81	4,266	8.52
Gender‡							
Male	32,092	6,128	19.10	3,084	9.61	3,430	10.69
Female	34,614	5,019	14.50	3,282	9.48	2,003	5.79
Education‡,§							
Less than high school	7,458	1,192	15.98	455	6.10	796	10.67
High school	14,845	1,962	13.22	890	6.00	1,149	7.74
Greater than high school	19,046	2,212	11.61	936	4.91	1,352	7.10
Total triggered verifications#		15,886		8,631		7,255	

†Percentages are based on the total sample triggering at least one check as displayed in Table 4.1.
‡Cochran-Mantel-Haenszel χ^2 Statistics indicate a significance at the 0.001 level across domain categories.
§Education includes only respondents 18 years and older.
#Sample sizes of total number of triggered verifications greater than the total number of respondents indicates that some respondents triggered multiple verifications.

same as their current age. When checking for differences among multilevel domains such as age groups, the χ^2 statistic only indicates the presence of a significant difference among the various age groups but does not specify exactly where the difference occurs. Thus, additional significance tests were completed to determine this. The results of these tests indicate that the youth respondents were significantly more likely than the adults to trigger this check. No doubt this is because many youth are trying substances for the first time and thus their current age and their age at first use are more likely to be the same. Males were more likely than females to trigger the check for an age of first use less than 10. Again, this is consistent with patterns of substance use (SAMHSA, 2001b).

Perhaps more important than who triggered the inconsistency and verification check items is the issue of whether the data quality of the

NHSDA was improved as a result of including the checks. A review of the data showed that there were 381 respondents with at least one inconsistency remaining in their data — 17 percent of the respondents who had inconsistent data originally. This means that data quality was improved for over 80 percent of the respondents who originally provided inconsistent responses.

Among respondents who triggered inconsistency checks based on their responses for binge alcohol frequency and any alcohol frequency, 27 percent were unable to resolve this inconsistency. In fact, 70 percent of the total 381 remaining inconsistencies were attributable to this check alone. This result suggests that there continued to be respondent confusion with these two items. Additional work is needed to understand the nature of this confusion and how best to clarify the relationship between these two items, as presenting the respondents with their inconsistent responses did not provide sufficient information to assist them in reconciling their data.

The data items associated with the respondent's age of first use remained inconsistent for approximately 14 percent of the respondents who originally had inconsistent data for these checks. A slightly higher percentage of respondents (23 percent) were unable to resolve inconsistencies in their recency data. A similar percentage (22 percent) were unable to resolve inconsistencies associated with their 30-day use being greater than their use during the past year.

Understanding the resolution of the cases where a respondent reported use in the past 30 days but then reported zero days of use requires an additional explanation of this check item. When respondents provided this type of inconsistent response, the check item asked them to verify their zero response. Respondents who indicated the zero response was incorrect were provided with an opportunity to answer the 30-day frequency question again. Respondents who indicated that the zero response was correct were routed to the next appropriate question with no additional follow-up. Thus, no complete resolution of the inconsistency occurred (e.g., by confronting the respondent with the two inconsistent answers and asking him/her to revise one or both of them). For this reason, respondents who triggered this check item and verified their zero response were not defined as consistent respondents. During editing, these cases were assigned a 12-month recency, and their 30-day frequency was set to "bad data." A final recency and

30-day frequency was then assigned during imputation.[1] A review of the data from this check item revealed the following:

- As noted earlier, 728 respondents triggered this type of check item. Among these respondents, 632 (87 percent) either verified their zero response was correct or indicated the zero response was incorrect, but, when given the opportunity to revise their answer, again entered zero.
- Among the remaining respondents, 95 (13 percent) indicated that their zero response was incorrect and revised their 30-day frequency to a numeric value other than zero. One respondent entered a Don't Know/Refused in response to the check item.

The fact that so many respondents verified their zero response (either directly or by reentering the zero response) suggests the need for further research to understand the cognitive processes that lead a respondent to initially report 30-day use but then to indicate a frequency of zero for that time period. One possibility may be that respondents are initially willing to report the recent use but then feel uncomfortable reporting a frequency and thus enter zero as a way to avoid revealing sensitive information. Another possibility is that respondents may believe they have used a particular drug during the past 30 days but when asked to report a frequency realize that their use was actually longer ago, leading them to record a zero response to the frequency question. From an instrumentation perspective, it may be necessary to revise this check item so that a complete resolution of the inconsistency is conducted and there is no ambiguity about the respondent's final recency status.

Finally, although the verification checks did not identify inconsistent data, there was the possibility that in providing the opportunity for a respondent to revise an unusual response he/she might enter inconsistent data. A review of the responses that were revised during verification checks showed that none of the respondents who triggered the age-at-first-use-equal-to-current-age check revised their responses inconsistently. Nine respondents who triggered the age-at-first-use-less-than-10 check revised their responses in such a way as to generate inconsistent data. Thus, the verification process increased our confidence in the quality of data collected from over 11,000 respondents while introducing new data inconsistencies for only nine.

---

[1]For a complete discussion of these editing and imputation rules, see Chapters 5 and 6.

Taken in total, the results of the use of a resolution methodology for inconsistent or unusual responses improved the consistency of the 1999 NHSDA; thus, one could argue the overall quality of the 1999 NHSDA data. Many responses that would otherwise have had to be revised through the use of editing programs after the fact were able to be clarified by respondents during the course of the interview. This is the preferred approach since the respondent is the most knowledgeable source for information about his/her substance use history. Plans call for the continued use of this resolution methodology for the NHSDA in the future and the identification of additional data items in the questionnaire that could benefit from the inclusion of check items.

4.2 Collecting 12-Month Frequency of Use Data in CAI

4.2.1 Background

Information on the frequency of substance use during the past 12 months is a key analysis item in the NHSDA. When the NHSDA used a paper questionnaire, the frequency of use question included nine numeric categories as well as two additional categories necessary because skip patterns were not used. The full set of response categories used for each substance covered in the NHSDA questionnaire was:

- More than 300 days (every day or almost every day)
- At least 201 but not more than 300 days (5 to 6 days a week)
- At least 101 but not more than 200 days (3 to 4 days a week)
- At least 51 but not more than 100 days (1 to 2 days a week)
- At least 25 but not more than 50 days (3 to 4 days a month)
- At least 12 but not more than 24 days (1 to 2 days a month)
- At least 6 but not more than 11 days (less than 1 day a month)
- At least 3 but not more than 5 days in the past 12 months
- At least 1 but not more than 2 days in the past 12 months
- I have used [DRUG NAME], but not during the past 12 months
- I have never used [DRUG NAME] in my life.

Over the years that this item has been included in the NHSDA questionnaire, two problems have been noted. First, the question can be difficult for many respondents to answer because it requires the respondent to recall information over a long period of time. The text in parentheses was thought to reduce this reporting burden by showing the respondent how the number of days would calculate out across the 12 months. However, this additional text may have caused confusion in and of itself because it suggested a regular periodicity of substance use and seemed to exclude episodic use. For respondents with a more episodic pattern to their substance use, it may have been difficult to choose a response category because the text shown in parentheses could contradict the number of days the respondent would otherwise have chosen. As an example, a respondent who only smoked marijuana during the 2 weeks when he was on vacation would have difficulty choosing a category because the number of days he used (14) falls into a category that is also identified as "1 to 2 days a month."

In preparing the NHSDA for CAI administration, a third potential problem was identified with the 12-month frequency of substance use item — the large number of response categories. Even excluding the last two response options (which could be deleted due to the use of skip routing in CAI), the list still included nine fairly lengthy response options. Particularly when listening to the audio, it seemed that the number of categories was too great and that respondents might lose track of which response category matched up with which response code. For this reason, it was decided that some change was needed to reduce the number and complexity of the response categories. Once the decision was made to revise the items, it was also decided to develop a response set that could be collected in a way that would minimize respondent recall burden and work equally well for both episodic and regular users.

4.2.2 Development of Methodology

As with the work on the resolution methodology described above, the initial development work began in the cognitive laboratory. To study how respondents understood the response categories, the two parts of the response categories were first split apart. Subjects were first asked to report the number of days they drank alcohol during the past 12 months using a showcard that displayed only the text shown in parentheses from the list above. Based on the category they selected, the interviewers provided them with a numerical estimate of their 12-month frequency of use. In each case, the estimate given was the numeric range originally associated with the parenthesized text. Subjects were asked whether the numeric estimate seemed right for them and, if not, whether the actual frequency was higher or lower than the estimate.

Not surprisingly, respondents had difficulty recalling their alcohol use over the past 12 months. The use of the categories did ease the recall task, although a number of subjects reported formulating their responses based on their use over just the past few months. Subjects with more sporadic use patterns had more difficulty formulating an answer than did subjects with regular use patterns. As had been hypothesized, this difficulty was generally due to the regularity implied by the response options.

In an effort to ease the reporting task for infrequent or sporadic users, work was undertaken to restructure the 12-month frequency-of-use item to remove the implied regularity from the response options. At the same time, work was done to reduce the total amount of text that would appear for the response options. The revised item actually collects the 12-month frequency data in two steps. First, respondents are asked to select a preferred unit for reporting their use. Respondents can choose to report the number of days per week, the number of days per month, or the total number of days they used a certain substance during the past 12 months. Once the preferred unit is selected, a follow-up question asks the respondent to report his/her 12-month use in those units. The actual number of days the substance was used during the past 12 months can then be calculated by multiplying a respondent's monthly answer by 12 or a weekly answer by 52. No additional computation is necessary for respondents who choose to report their 12-month use as a total number of days.

In a final round of cognitive testing, the two-stage approach was administered to see whether it would ease the recall task. In nearly all cases, the two-stage approach was received favorably by subjects. Respondents reported selecting the units based on the regularity of their use; respondents who drank quite regularly chose to report the number of days per week, those who drank slightly less often chose to report the number of days per month, and those who drank rarely or only on special occasions chose to report the total number of days during the past 12 months (SAMHSA, 2001a).

4.2.3 Results from 1999 NHSDA

Two types of analyses were prepared to provide information on the quality of data collected under the new 12-month frequency reporting methodology. First, the correspondence between the reporting units originally selected by respondents and the actual units in which they reported was examined. To understand this analysis, an additional explanation of the instrument's programmed routing is necessary. As noted earlier, each respondent who reported use of a substance during the past 12 months was first asked to choose a reporting unit (days per week, days per month, or total number of days) and then routed to a tailored follow-up question that asked for his/her 12-month frequency of use in those units. Respondents who did not choose a reporting unit (e.g., gave a Don't Know or

Refused response) were not immediately routed out of the question sequence, however. Instead, they were routed to the follow-up question and asked to record their 12-month frequency using the total number of days as the reporting unit. If a second Don't Know or Refused response was given, the respondents were then asked to report an answer using days per month as the reporting unit. If a third Don't Know or Refused response was entered, respondents were given a final opportunity to report their 12-month frequency using days per week as the reporting unit. Thus, respondents who did not originally select a reporting unit received up to three opportunities to provide the 12-month frequency data. Similarly, respondents who chose a reporting unit but then provided a Don't Know or Refused response to the tailored follow-up question were given additional opportunities to report their 12-month frequency using the remaining reporting units. The order of the follow-up questions was always the same: total number of days, days per month, and days per week. So, for example, a respondent who chose days per month as his/her preferred reporting unit but then provided a Don't Know response to that version of the 12-month frequency follow-up question would have one additional opportunity to provide the 12-month frequency data by responding to the days-per-week version of the 12-month frequency follow-up question. In all cases, once a substantive response was given to the frequency question, the respondent was routed out of the remaining follow-up questions.

An analysis of the 12-month frequency data shows that the overwhelming majority of respondents (over 98 percent in each case) who initially selected a preferred unit for reporting went on to report their 12-month frequency using those units. Only a small percentage of respondents who chose a reporting unit went on to give a Don't Know or Refused response for the follow-up item to which they were routed (less than 2 percent in each case). This suggests the two-step reporting task was not overly confusing for respondents and may in fact have had the intended effect of simplifying the reporting task for this item. In addition, only a very small percentage of respondents reported their 12-month frequency in a unit other than the one initially selected (less than 1 percent in each case). What is perhaps more interesting, however, is that the majority of respondents who did not initially select a reporting unit (but instead entered a Don't Know or Refused response) went on to provide a substantive answer to the 12-month frequency of use item as a result of being routed through the subsequent follow-up items. In both cases (initial Don't Knows and initial Refuseds), the majority of respondents — 77 percent of initial Don't Knows and 70 percent of initial Refuseds — ultimately provided usable data for the 12-month frequency-of-use item. As a result of this routing pattern, the overall rate of missing data for the 12-month frequency-of-use items was much lower than it would have been if respondents who initially provided a Don't Know or Refused response had been routed out of the follow-up questions.

Results were also reviewed by demographic subgroups. The data were consistent with the results reported above for the total sample.

A second analysis is presented in ***Table 4.3.*** These data provide a comparison of the distribution of respondents across the response categories for 12-month frequency of drinking alcohol. The categories are those that were used in previous years of the NHSDA and for the PAPI cases in 1999. The continuous data from the CAI cases were collapsed to create a comparable categorical variable. The data are shown both with the inclusion of the respondents who did not drink alcohol during the past 12 months and with these respondents excluded.

A difference of means test was conducted and ***Table 4.3*** shows that the weighted CAI mean is significantly higher than that for PAPI. Means were calculated using the midpoint of the frequency range from each response option. This result is consistent with what was found in the 1997 CAI Field Experiment data (see SAMHSA, 2001) where a higher frequency of drinking alcohol was reported under the CAI methodology than the PAPI. It is possible this difference was the result of the revised question wording. Because respondents were generating their own responses without being able to see other response options (as was the case on the answer sheets used in PAPI), their responses may have been less influenced by social desirability. The ability to influence frequency reporting by altering the endpoints of the scale has been well-documented in the literature (see, for example, Schwarz & Hippler, 1991). A similar phenomenon may have occurred here, in that respondents' answers were more accurate because they were less influenced by outside factors. However, it is also possible that the difference is attributable to the use of the CAI mode of interviewing, which was designed to further enhance the respondent's privacy. Having improved the privacy of the interview, respondents may now be reporting more honestly. Statistically significant differences were found for nearly all demographic subgroups (data not presented here), and, in all cases, the CAI mean was higher than that for PAPI. Subgroups for which the results were not statistically significant include respondents 50 years old or older, blacks, and those with less than a high school education.

Table 4.3 1999 NHSDA 12-Month Substance Use Frequency Comparison for Alcohol

Contrast Mean for Difference Testing:† 16.09 (86.64 – 70.55) p-value = .000048										
	1999 CAI					**1999 PAPI**				
Weighted Mean for Difference Test‡	86.64					NA				
Full Sample Means	Unweighted: 72.38 Weighted: 86.17					Unweighted: 56.81 Weighted: 70.55				
	All Categories			**12-Month Users Only**		**All Categories**			**12-Month Users Only**	
		Percent		**Percent**			**Percent**		**Percent**	
Frequency Categories	**Sample**	**Unwgtd**	**Wgtd**	**Unwgtd**	**Wgtd**	**Sample**	**Unwgtd**	**Wgtd**	**Unwgtd**	**Wgtd**
More than 300 days (almost every day)	1,066	1.60	3.55	2.78	5.67	453	3.28	5.51	5.40	8.51
201-300 days (5-6 days a week)	2,377	3.56	5.24	6.20	8.38	285	2.06	2.90	3.40	4.48
101-200 days (3-4 days a week)	6,941	10.41	11.02	18.10	17.62	594	4.30	4.76	7.09	7.36
51-100 days (1-2 days a week)	6,097	9.14	9.78	15.90	15.63	1,199	8.68	9.26	14.30	14.29
25-50 days (3-4 days a month)	4,681	7.02	7.19	12.21	11.49	1,081	7.83	9.14	12.90	14.11
12-24 days (1-2 days a month)	6,188	9.28	10.03	16.14	16.03	1,160	8.40	8.83	13.84	13.62
6-11 days (< 1 day a month)	2,874	4.31	4.58	7.49	7.32	1,010	7.31	7.89	12.05	12.18
3-5 days	4,532	6.79	6.89	11.82	11.01	1,240	8.98	8.70	14.79	13.43
1-2 days	3,591	5.38	4.29	9.36	6.86	1,360	9.85	7.78	16.23	12.02
Never used or did not use in past year	28,359	42.51	37.43	NA§	NA§	5,427	39.30	35.22	NA§	NA§

†Contrast mean for the Test of Differences is defined as the difference between the CAI and PAPI means (H_0: Contrast Mean = 0). The PAPI mean for testing is the same Full Sample Weighted Mean (no subsetting necessary for PAPI sample).
‡Weighted Means for the Difference Test only includes CAI and PAPI sample in the overlap FI Regions.
§NA = not applicable.

With no external source available to validate the NHSDA data, it is impossible to determine which frequency distribution (that produced under CAI or PAPI) is more accurate. However, research on the reporting of sensitive behaviors routinely assumes that higher reporting reflects greater accuracy (see, for example, Tourangeau & Smith, 1996; O'Reilly, Hubbard, Lessler, Biemer, & Turner, 1994; Turner et al., 1992). This assumption is based on the belief that respondents have reason to underreport sensitive behaviors (e.g., to avoid embarrassment or because they fear legal sanctions) and do not have reason to overreport such behaviors. Thus, it is assumed that the higher frequency reports found in CAI are an indication of improved data quality.

4.3 General Measures of Data Quality

Basic measures of data quality that are routinely examined in survey research are rates of missing data. High rates of missing data can compromise data quality. In addition, for self-administered items, high rates may indicate problems with respondent comprehension or recall burden. When comparing such rates for the CAI and PAPI NHSDA questionnaires, it is important to keep several facts in mind. First, a serious threat to data quality in the PAPI NHSDA were the data items left blank by the respondent. Because interviewers could not review respondents′ answer sheets before mailing completed cases to RTI for processing, they were often unaware of problems the respondent may have had completing the questions. For example, questions may have been left blank because respondents did not follow skip patterns correctly, failed to notice additional questions on the back of an answer sheet, or turned too many pages at one time within an answer sheet booklet (thus missing some of the questions). Items the respondent could not or did not wish to answer may also have been left blank because respondents did not receive explicit instructions on how to mark a Refused or Don′t Know response unless they specifically asked the interviewer for this information. Thus, the true meaning of a blank data item in the PAPI questionnaire is somewhat unclear.

The move to CAI data collection provided an opportunity to eliminate blank data items, an important improvement to overall data quality. With CAI, careful programming can ensure that each respondent is routed only to those questions that are appropriate based on his/her previous answers and can require that an answer be input before the respondent can move to the next question. Although this eliminates blanks, explicit instructions must be provided for entering Refused or Don't Know responses. In the case of the CAI NHSDA, two function keys were labeled for this purpose. Respondents were instructed to press F3 to record a Don't Know response and F4 to record a Refused response.

In considering whether data quality was improved in the CAI NHSDA by reducing the rate of missing data, it is necessary to consider all three types of missing data (Don't Know, Refused, and blank) even though the rates of the individual types will differ by mode. In ***Table 4.4***, these rates of missing data are presented for a set of critical items in the NHSDA. Two items are reviewed — lifetime use and recency of use. These items are included across each of the core modules of the NHSDA and have been central to the reporting needs of SAMHSA and its constituency since the inception of the NHSDA. The obvious finding from ***Table 4.4*** is that missing data were, for the most part, a small problem across both modes of data collection in the 1999 NHSDA.

Interestingly, the rate of missing data for the PAPI mode increased more for the rarer (and perhaps more sensitive) substance types, which are asked about later in the interview, than the more commonly used (and less sensitive) substances. This is not the case for the CAI data where rates remained more constant across substance types. It is possible that the missing data rates increased in PAPI because the respondent became more concerned about reporting such sensitive behaviors using the answer sheet methodology. Perhaps the answer sheet methodology was viewed as sufficiently private for reporting less sensitive types of substance use but not for the rarer substance. It is also possible that respondents who had never used substances grew weary of marking the specific Never Used category for each question on an answer sheet and instead simply began to leave many of the questions blank. It may also be possible that the missing data rate may have increased for the later substances because the format of the answer sheets became more complicated and respondents may have had a more difficult time understanding how to record their answers.

Another aspect that might contribute to the differences between CAI and PAPI, especially for the rarer substances, is the usable case rule discussed in ***Chapter 5***. Given that the CAI requires the cigarette lifetime plus nine other substance lifetime questions to be answered with either a "yes" or "no," there are minimal opportunities for respondents to respond with a Don't Know or Refused and still be considered a usable case. The PAPI, on the other hand, requires respondents to answer only the alcohol, marijuana, and cocaine lifetime questions. This implies that many of the

Table 4.4 Distribution of Missing Data by Substance of Interest for Lifetime Use and Recency of Use: Total Sample

			1999 CAI Sample			1999 PAPI Sample†		
				Percent			Percent	
Substance Type	**Question**	**Type of Missing Data**	**N**	**Unwgtd**	**Wghtd**	**N**	**Unwgtd**	**Wghtd**
Cigarettes	Lifetime use	Don't know	0	0.000	0.000	0	0.000	0.000
		Refused	0	0.000	0.000	2	0.014	0.001
		Blank	NA	NA	NA	53	0.384	0.301
	Recency period	Don't know	242	0.363	0.386	3	0.022	0.003
		Refused	67	0.100	0.053	1	0.007	0.000
		Blank	NA	NA	NA	164	1.188	1.073
Alcohol	Lifetime use	Don't know	23	0.034	0.016	0	0.000	0.000
		Refused	54	0.081	0.027	1	0.007	0.012
		Blank	NA	NA	NA	32	0.232	0.061
	Recency period	Don't know	298	0.447	0.370	2	0.014	0.006
		Refused	98	0.147	0.071	1	0.007	0.012
		Blank	NA	NA	NA	36	0.261	0.189
Marijuana	Lifetime use	Don't know	25	0.037	0.056	0	0.000	0.000
		Refused	200	0.300	0.297	0	0.000	0.000
		Blank	NA	NA	NA	59	0.427	0.367
	Recency period	Don't know	130	0.195	0.122	1	0.007	0.002
		Refused	118	0.177	0.158	0	0.000	0.000
		Blank	NA	NA	NA	64	0.463	0.624
Cocaine	Lifetime use	Don't know	20	0.030	0.040	0	0.000	0.000
		Refused	94	0.141	0.197	0	0.000	0.000
		Blank	NA	NA	NA	95	0.688	0.547
	Recency period	Don't know	24	0.036	0.036	0	0.000	0.000
		Refused	37	0.055	0.066	0	0.000	0.000
		Blank	NA	NA	NA	77	0.558	0.770
Crack	Lifetime use	Don't know	7	0.010	0.009	0	0.000	0.000
		Refused	11	0.016	0.008	0	0.000	0.000
		Blank	NA	NA	NA	111	0.804	0.650
	Recency period	Don't know	7	0.010	0.005	0	0.000	0.000
		Refused	11	0.016	0.014	1	0.007	0.000
		Blank	NA	NA	NA	99	0.717	0.840
Heroin	Lifetime use	Don't know	18	0.027	0.011	0	0.000	0.000
		Refused	57	0.085	0.105	0	0.000	0.000
		Blank	NA	NA	NA	176	1.275	1.360
	Recency period	Don't know	6	0.009	0.011	0	0.000	0.000
		Refused	9	0.013	0.006	0	0.000	0.000
		Blank	NA	NA	NA	101	0.731	0.882
Hallucinogens‡	Lifetime use	Don't know	415	0.622	0.368	4	0.029	0.010
		Refused	132	0.198	0.196	0	0.000	0.000
		Blank	NA	NA	NA	608	4.403	4.135
	Recency period	Don't know	85	0.127	0.067	1	0.007	0.001
		Refused	51	0.076	0.048	0	0.000	0.000
		Blank	NA	NA	NA	126	0.912	1.095
Inhalants‡	Lifetime use	Don't know	181	0.271	0.113	0	0.000	0.000
		Refused	110	0.165	0.112	1	0.007	0.042
		Blank	NA	NA	NA	456	3.302	2.663
	Recency period	Don't know	162	0.243	0.114	0	0.000	0.000
		Refused	99	0.148	0.095	2	0.014	0.064
		Blank	NA	NA	NA	311	2.252	2.692

† PAPI data are raw and have not been subjected to edit routines that impute missing data.
‡ Hallucinogens and Inhalants ask several questions pertaining to lifetime use. These statistics are based on respondents answering at least one of the questions with a Don't Know or Refused response or leaving the question blank. In this regard, percentages are not additive.

blank responses encountered within the PAPI sample may be due to cases that met the minimal criteria for the usable case rule but subsequently did not answer any of the remaining questions about other substances. Given this aspect, the extent of nondeterminant data may be more of an issue with editing/usable case specifications then with actual respondent refusal or miscomprehension.

Similar data showing the missing data rates for youth and adult respondents are presented in ***Tables 4.5*** and ***4.6***, respectively. These data show a similar pattern to that for the overall sample.

Taken together, it appears that rates of missing data decreased under the CAI mode. In an effort to further improve the quality of data collected in the NHSDA, follow-up questions for respondents who provided Refused or Don't Know responses to critical items were added for the 2000 NHSDA to encourage those respondents to provide substantive answers. So, in the future, missing data rates may be even lower for the CAI NHSDA.

Another general measure of data quality that is often reviewed is the rate of partial interviews, also referred to as "breakoffs." A high rate of breakoffs may indicate that a questionnaire is too long or too labor-intensive. Data quality will obviously suffer when breakoff rates increase because more respondents will have incomplete data records, which can hamper analysis of the data. The rate of breakoff interviews for both CAI and PAPI were reviewed to determine whether rates were comparable between the modes. Only 11 cases that were finalized as breakoffs for PAPI (0.08 percent of all completed PAPI interviews) and 91 breakoff cases for CAI (0.14 percent of all completed CAI interviews) were found.[2] Although the CAI rate is double that of PAPI, both rates are extremely small and, thus, data quality is unlikely to be jeopardized. It is encouraging to note that the move to CAI data collection did not significantly deter respondents from completing the full NHSDA interview. This had been one area of concern during the conversion process since, at that time, no study had attempted to conduct such a lengthy ACASI interview.

---

[2]Of the 11 PAPI breakoffs, 5 (45 percent) were ultimately determined to be unusable cases based on the rules described in ***Chapter 5***. Of the 91 CAI breakoffs, 41 (45 percent) were determined to be unusable.

Table 4.5 Distribution of Missing Data by Substance of Interest for Lifetime Use and Recency of Use: Aged 12 to 17

			1999 CAI Sample			1999 PAPI Sample†		
				Percent			Percent	
Substance Type	Question	Type of Missing Data	N	Unwgtd	Wghtd	N	Unwgtd	Wghtd
Cigarettes	Lifetime use	Don't know	0	0.000	0.000	0	0.000	0.000
		Refused	0	0.000	0.000	0	0.000	0.000
		Blank	NA	NA	NA	8	0.058	0.013
	Recency period	Don't know	117	0.175	0.044	1	0.007	0.002
		Refused	38	0.057	0.010	0	0.000	0.000
		Blank	NA	NA	NA	31	0.224	0.058
Alcohol	Lifetime use	Don't know	17	0.025	0.008	0	0.000	0.000
		Refused	29	0.043	0.011	0	0.000	0.000
		Blank	NA	NA	NA	11	0.080	0.014
	Recency period	Don't know	180	0.270	0.076	2	0.014	0.006
		Refused	71	0.106	0.025	0	0.000	0.000
		Blank	NA	NA	NA	10	0.072	0.028
Marijuana	Lifetime use	Don't know	12	0.018	0.005	0	0.000	0.000
		Refused	62	0.093	0.027	0	0.000	0.000
		Blank	NA	NA	NA	10	0.072	0.022
	Recency period	Don't know	61	0.091	0.021	1	0.007	0.002
		Refused	40	0.060	0.016	0	0.000	0.000
		Blank	NA	NA	NA	11	0.080	0.041
Cocaine	Lifetime use	Don't know	9	0.013	0.004	0	0.000	0.000
		Refused	15	0.022	0.006	0	0.000	0.000
		Blank	NA	NA	NA	19	0.138	0.029
	Recency period	Don't know	9	0.013	0.005	0	0.000	0.000
		Refused	9	0.013	0.004	0	0.000	0.000
		Blank	NA	NA	NA	13	0.094	0.044
Crack	Lifetime use	Don't know	3	0.004	0.002	0	0.000	0.000
		Refused	2	0.003	0.001	0	0.000	0.000
		Blank	NA	NA	NA	20	0.145	0.046
	Recency period	Don't know	0	0.000	0.000	0	0.000	0.000
		Refused	2	0.003	0.001	0	0.000	0.000
		Blank	NA	NA	NA	18	0.130	0.060
Heroin	Lifetime use	Don't know	15	0.022	0.006	0	0.000	0.000
		Refused	13	0.019	0.003	0	0.000	0.000
		Blank	NA	NA	NA	28	0.203	0.081
	Recency period	Don't know	3	0.004	0.001	0	0.000	0.000
		Refused	2	0.003	0.001	0	0.000	0.000
		Blank	NA	NA	NA	16	0.116	0.062
Hallucinogens‡	Lifetime use	Don't know	271	0.406	0.111	1	0.007	0.000
		Refused	58	0.087	0.022	0	0.000	0.000
		Blank	NA	NA	NA	114	0.826	0.349
	Recency period	Don't know	43	0.064	0.019	0	0.000	0.000
		Refused	27	0.040	0.010	0	0.000	0.000
		Blank	NA	NA	NA	25	0.181	0.076
Inhalants‡	Lifetime use	Don't know	132	0.198	0.051	0	0.000	0.000
		Refused	70	0.105	0.028	0	0.000	0.000
		Blank	NA	NA	NA	105	0.760	0.318
	Recency period	Don't know	116	0.174	0.042	0	0.000	0.000
		Refused	58	0.087	0.020	0	0.000	0.000
		Blank	NA	NA	NA	60	0.434	0.186

† PAPI data are raw and have not been subjected to edit routines that impute missing data.
‡ Hallucinogens and Inhalants ask several questions pertaining to lifetime use. These statistics are based on respondents answering at least one of the questions with a Don't Know or Refused response leaving the question blank. In this regard, percentages are not additive.

Table 4.6 Distribution of Missing Data by Substance of Interest for Lifetime Use and Recency of Use: Aged 18 or Older

			1999 CAI Sample			1999 PAPI Sample†		
				Percent			Percent	
Substance Type	Question	Type of Missing Data	N	Unwgtd	Wghtd	N	Unwgtd	Wghtd
Cigarettes	Lifetime use	Don't know	0	0.000	0.000	0	0.000	0.000
		Refused	0	0.000	0.000	2	0.014	0.001
		Blank	NA	NA	NA	45	0.326	0.289
	Recency period	Don't know	125	0.187	0.342	2	0.014	0.001
		Refused	29	0.043	0.043	1	0.007	0.000
		Blank	NA	NA	NA	133	0.963	1.014
Alcohol	Lifetime use	Don't know	6	0.009	0.008	0	0.000	0.000
		Refused	25	0.037	0.016	1	0.007	0.012
		Blank	NA	NA	NA	21	0.152	0.046
	Recency period	Don't know	118	0.177	0.294	0	0.000	0.000
		Refused	27	0.040	0.046	1	0.007	0.012
		Blank	NA	NA	NA	26	0.188	0.161
Marijuana	Lifetime use	Don't know	13	0.019	0.051	0	0.000	0.000
		Refused	138	0.207	0.270	0	0.000	0.000
		Blank	NA	NA	NA	49	0.355	0.345
	Recency period	Don't know	69	0.103	0.101	0	0.000	0.000
		Refused	78	0.117	0.143	0	0.000	0.000
		Blank	NA	NA	NA	53	0.384	0.583
Cocaine	Lifetime use	Don't know	11	0.016	0.036	0	0.000	0.000
		Refused	79	0.118	0.191	0	0.000	0.000
		Blank	NA	NA	NA	76	0.550	0.518
	Recency period	Don't know	15	0.022	0.032	0	0.000	0.000
		Refused	28	0.042	0.062	0	0.000	0.000
		Blank	NA	NA	NA	64	0.463	0.726
Crack	Lifetime use	Don't know	4	0.006	0.007	0	0.000	0.000
		Refused	9	0.013	0.006	0	0.000	0.000
		Blank	NA	NA	NA	91	0.659	0.604
	Recency period	Don't know	7	0.010	0.005	0	0.000	0.000
		Refused	9	0.013	0.013	1	0.007	0.000
		Blank	NA	NA	NA	81	0.587	0.781
Heroin	Lifetime use	Don't know	3	0.004	0.005	0	0.000	0.000
		Refused	44	0.066	0.102	0	0.000	0.000
		Blank	NA	NA	NA	148	1.072	1.279
	Recency period	Don't know	3	0.004	0.010	0	0.000	0.000
		Refused	7	0.010	0.005	0	0.000	0.000
		Blank	NA	NA	NA	85	0.616	0.821
Hallucinogens‡	Lifetime use	Don't know	144	0.216	0.257	3	0.022	0.010
		Refused	74	0.111	0.174	0	0.000	0.000
		Blank	NA	NA	NA	494	3.577	3.787
	Recency period	Don't know	42	0.063	0.048	1	0.007	0.001
		Refused	24	0.036	0.038	0	0.000	0.000
		Blank	NA	NA	NA	101	0.731	1.019
Inhalants‡	Lifetime use	Don't know	49	0.073	0.063	0	0.000	0.000
		Refused	40	0.060	0.084	1	0.007	0.042
		Blank	NA	NA	NA	351	2.542	2.345
	Recency period	Don't know	46	0.069	0.073	0	0.000	0.000
		Refused	41	0.061	0.075	2	0.014	0.064
		Blank	NA	NA	NA	251	1.818	2.506

† PAPI data are raw and have not been subjected to edit routines that impute missing data.

‡ Hallucinogens and Inhalants ask several questions pertaining to lifetime use. These statistics are based on respondents answering at least one of the questions with a Don't Know or Refused response or leaving the question blank. In this regard, percentages are not additive.

4.4 Results From Interview Debriefing Items

At the conclusion of each interview, the interviewer completed a series of debriefing items designed to provide additional information about the environment in which the interview was completed and any information that might be useful in interpreting the respondent's data. The interviewer completed these items without input from the respondent, so the data collected may not accurately reflect a respondent's actual opinions, but only the interviewer's interpretation of these opinions based on his or her own assessment of the respondent. As a result of this aspect and missing data, only unweighted percentages are displayed. A subset of these debriefing items were the same for both the CAI and PAPI interviews that were completed as part of the 1999 NHSDA. A comparison of these items is provided as one final check on the data quality of the CAI NHSDA.

In both the CAI and PAPI NHSDA, respondents were responsible for completing a large portion of the interview on their own. For this reason, it was very important that the interview mode be easy for the respondents to complete. In ***Table 4.7***, data on the interviewers' assessments of respondent difficulty with the interview are presented.

For both the CAI and PAPI cases, the majority of respondents were rated as having no difficulty with the interview. In all but one of the subgroups (respondents aged 50 or older), a larger percentage of CAI respondents than PAPI respondents were rated as having no difficulty with the interview. Perhaps more interesting is the percentage of respondents who were rated as having a lot of difficulty with the interview. Fortunately, these rates are low overall (0.5 percent for CAI and 1.5 percent for PAPI) and for most subgroups. Two subgroups are worth reviewing, however. The rates of respondents having a lot of difficulty with the interview were highest for respondents aged 50 years or older and respondents who had less than a high school education. These subgroups were of special concern during the development of the CAI instrument because they were expected to have less experience working with computers; thus, it is surprising to note that the percentage of less educated respondents having difficulty was higher for PAPI than for CAI. Respondents 50 and older also exhibited greater difficulty with PAPI than with CAI, although the results are not statistically significant. Based on these data, it seems that the answer sheets presented more of a challenge to these respondents than did the CAI instrument. The reason may be that respondents did not have to follow routing instructions in CAI or that, because of the use of skip patterns in the

Table 4.7 Respondent Difficulty with Interview Task

Domains	Interviewer Report of Difficulty	1999 CAI		1999 PAPI	
		N	%	N	%
Total †	No difficulty	57,991	87.2	11,223	83.4
	Some difficulty	8,189	12.3	2,032	15.1
	A lot of difficulty	338	0.5	204	1.5
Age group					
12-17†	No difficulty	21,823	85.1	2,752	81.5
	Some difficulty	3,371	13.1	592	17.5
	A lot of difficulty	438	1.8	31	1.0
18-25†	No difficulty	20,048	91.6	3,094	86.8
	Some difficulty	1,778	8.1	433	12.1
	A lot of difficulty	62	0.3	36	1.1
26-34†	No difficulty	7,027	89.3	2,489	86.5
	Some difficulty	787	10.0	356	12.4
	A lot of difficulty	50	0.7	34	1.1
35-49†	No difficulty	5,476	88.0	1,914	84.8
	Some difficulty	715	11.5	312	13.8
	A lot of difficulty	33	0.5	30	1.4
50+	No difficulty	3,617	68.6	974	70.3
	Some difficulty	1,538	29.2	339	24.5
	A lot of difficulty	117	2.2	73	5.2
Race/Ethnicity					
Hispanic†	No difficulty	6,633	78.5	2,227	75.1
	Some difficulty	1,723	20.4	667	22.5
	A lot of difficulty	91	1.1	70	2.4
Non-Hisp black†	No difficulty	6,675	81.9	2,520	78.5
	Some difficulty	1,400	17.2	626	19.5
	A lot of difficulty	71	0.9	63	2.0
Non-Hisp non-black	No difficulty	44,683	89.5	6,476	88.9
	Some difficulty	5,066	10.1	739	10.1
	A lot of difficulty	176	0.4	71	1.0
Gender					
Male†	No difficulty	27,613	86.3	5,024	81.0
	Some difficulty	4,188	13.1	1,059	17.1
	A lot of difficulty	193	0.6	117	1.9
Female†	No difficulty	30,378	88.0	6,199	85.4
	Some difficulty	4,001	11.6	973	13.4
	A lot of difficulty	145	0.4	87	1.2
Education					
Less than high school†	No difficulty	5,281	71.1	1,455	63.7
	Some difficulty	1,972	26.6	703	30.8
	A lot of difficulty	171	2.3	125	5.5
High school†	No difficulty	12,909	87.1	2,882	84.5
	Some difficulty	1,839	12.4	492	14.3
	A lot of difficulty	66	0.5	36	1.2
More than high school	No difficulty	17,978	94.6	4,134	94.1
	Some difficulty	1,007	5.3	245	5.6
	A lot of difficulty	25	0.1	12	0.3

† Cochran-Mantel-Haenszel χ^2 Statistics indicate a significant difference in means across response categories at the .001 level between CAI and PAPI samples.

core of the CAI instrument, respondents no longer had to answer questions that seemed irrelevant to them (e.g., answering questions about a substance they had never used by marking a box each time to indicate this). However, this is only speculation. Overall, although the differences are not large, it is encouraging to see that the computer technology did decrease interview difficulties for respondents. It is assumed that this can only result in improvements in overall data quality for the NHSDA.

A second comparison can be made between CAI and PAPI for an item that asked the interviewer to assess the privacy of the interview setting. Due to the sensitive nature of the questions in the NHSDA, study protocol required interviewers to seek out a private setting for the interview in order to encourage honest reporting by the respondent (and thus improve data quality). ***Table 4.8*** shows the data for this item. For the total sample, there is a three-point difference in the percentage of cases where the interview setting was completely private, with the CAI rate being higher than PAPI. For youth respondents, 67 percent of the CAI interviews took place in a completely private environment versus 61 percent for the PAPI interviews, a striking difference. This is notable since historically it has been more difficult for NHSDA interviewers to achieve a private interview setting with youth respondents because adult members of the household often want to be present. If the CAI interview can improve the privacy of the interview by reducing the degree to which others are able to view the process, it seems likely that it will have a positive impact on data quality.

One debriefing item was included only for the CAI interviews. Interviewers were asked to rate the influence of the computer on the respondents′ decision to participate. Although these data were collected only for cases that resulted in a completed interview, it seems possible that data collected from respondents who viewed the computer negatively might be adversely affected. Overall, the computer was reported to be a positive influence in two-thirds of the cases (see ***Table 4.9***). Not surprisingly, this varied by demographic subgroup. The computer was rated as a positive influence for only 49 percent of the interviews conducted with the oldest respondents, and it was rated as a negative influence in 10 percent of these cases. This is consistent with concerns raised by project staff regarding the conversion of the instrument to CAI as well as the informal feedback received from interviewers who reported that it took more persuasion to convince older people to participate in the CAI NHSDA. Whether data quality has been sacrificed for these older respondents is not entirely clear. However, based on the other analyses presented in this chapter, this does not appear to be the case.

Table 4.8 Privacy of Interview Setting

Domains	Interviewer Report of Privacy	1999 CAI N	1999 CAI %	1999 PAPI N	1999 PAPI %
Total†	Completely private	48,534	73.0	9,514	70.0
	Some lack of privacy	15,895	23.9	3,543	26.0
	Constant presence of others	2,100	3.1	550	4.0
Age Group					
12-17†	Completely private	16,945	67.0	2,065	60.8
	Some lack of privacy	7,377	29.2	1,160	34.1
	Constant presence of others	954	3.8	172	5.1
18-25†	Completely private	16,707	76.3	2,619	72.8
	Some lack of privacy	4,545	20.8	858	23.8
	Constant presence of others	639	2.9	123	3.4
26-34†	Completely private	5,873	74.7	2,097	71.7
	Some lack of privacy	1,777	22.6	718	24.6
	Constant presence of others	213	2.7	108	4.7
35-49†	Completely private	4,918	79.0	1,688	74.0
	Some lack of privacy	1,195	19.2	525	23.0
	Constant presence of others	112	1.8	69	3.0
50+†	Completely private	4,091	77.6	1,045	74.4
	Some lack of privacy	1,001	19.0	282	20.1
	Constant presence of others	182	3.4	78	5.5
Race/Ethnicity					
Hispanic	Completely private	5,798	68.6	2,026	67.6
	Some lack of privacy	2,320	27.4	841	28.1
	Constant presence of others	331	4.0	130	4.3
Non-Hisp black	Completely private	5,668	69.6	2,205	68.1
	Some lack of privacy	2,206	27.1	900	27.8
	Constant presence of others	275	3.6	133	4.1
Non-Hisp non-black†	Completely private	37,068	74.2	5,283	71.7
	Some lack of privacy	11,369	22.8	1,802	24.4
	Constant presence of others	1,494	3.0	287	3.9
Gender					
Male†	Completely private	23,529	73.5	4,376	69.8
	Some lack of privacy	7,480	23.4	1,627	26.0
	Constant presence of others	993	3.1	263	4.2
Female†	Completely private	25,005	72.4	5,138	70.0
	Some lack of privacy	8,415	24.4	1,916	26.1
	Constant presence of others	1,107	3.2	287	3.9
Education					
Less than high school	Completely private	4,888	65.8	1,504	64.9
	Some lack of privacy	2,186	29.4	687	29.7
	Constant presence of others	356	4.8	125	5.4
High school†	Completely private	11,120	75.1	2,494	72.3
	Some lack of privacy	3,271	22.1	813	23.6
	Constant presence of others	424	2.8	141	4.1
More than high school†	Completely private	15,581	82.0	3,451	77.6
	Some lack of privacy	3,061	16.1	883	19.9
	Constant presence of others	366	1.9	112	2.5

† Cochran-Mantel-Haenszel χ^2 Statistics indicate a significant difference in means across response categories at the .001 level between CAI and PAPI samples.

Table 4.9 Influence of CAI on Respondents

Domains	Interviewer Report of Influence	1999 CAI	
		N	%
Total	Influenced in a positive way	44,348	66.6
	No influence	21,212	31.9
	Influenced in a negative way	980	1.5
Age Group			
12-17	Influenced in a positive way	18,452	73.0
	No influence	36,733	26.6
	Influenced in a negative way	96	0.4
18-25	Influenced in a positive way	14,473	66.1
	No influence	7,263	33.2
	Influenced in a negative way	155	0.7
26-34	Influenced in a positive way	5,049	64.2
	No influence	2,724	34.6
	Influenced in a negative way	91	1.2
35-49	Influenced in a positive way	3,809	61.1
	No influence	2,304	37.0
	Influenced in a negative way	120	1.9
50+	Influenced in a positive way	2,565	48.7
	No influence	2,188	41.5
	Influenced in a negative way	518	9.8
Race/Ethnicity			
Hispanic	Influenced in a positive way	6,041	71.5
	No influence	2,306	27.3
	Influenced in a negative way	104	1.2
Non-Hisp black	Influenced in a positive way	5,761	70.7
	No influence	2,265	27.8
	Influenced in a negative way	118	1.5
Non-Hisp non-black	Influenced in a positive way	32,546	65.2
	No influence	16,641	33.3
	Influenced in a negative way	758	1.5
Gender			
Male	Influenced in a positive way	21,406	66.9
	No influence	10,176	31.8
	Influenced in a negative way	428	1.3
Female	Influenced in a positive way	22,942	66.4
	No influence	11,036	32.0
	Influenced in a negative way	552	1.6
Education			
Less than high school	Influenced in a positive way	4,642	62.4
	No influence	2,440	32.8
	Influenced in a negative way	352	4.8
High school	Influenced in a positive way	9,344	63.1
	No influence	5,154	34.8
	Influenced in a negative way	317	2.1
More than high school	Influenced in a positive way	11,910	62.3
	No influence	6,885	36.2
	Influenced in a negative way	215	1.5

4.5 Conclusions

In this chapter, a number of different analyses have been presented that provide information on the overall quality of the data collected in the 1999 NHSDA. The results suggest that the move to CAI data collection has improved data quality, although in some cases the increase is fairly small because data quality for the PAPI NHSDA data was already quite high. Perhaps the most significant improvement to data quality comes as a result of the inclusion of the inconsistent and unusual data checks described in ***Section 4.1***. This enhancement is a radical departure from the PAPI NHSDA and one that was possible only under the CAI mode of interview. The results presented here show that, although the CAI data did not suffer from a large amount of inconsistent or unusual data, the methodology was able to resolve a large number of these cases in a way that is both cost-effective and likely to enhance overall data quality for the items involved.

Results of the change in the 12-month frequency of use item are somewhat difficult to interpret. The revised method for collecting these data does result in higher reported frequencies, but whether this is due to the revision or simply the move from PAPI to CAI is impossible to determine. The distribution of responses and the fact that the mean frequency is higher under CAI than PAPI do provide anecdotal support for the revised method of collecting these data, however.

Finally, results from basic measures of data quality and interviewer debriefing items suggest that the CAI methodology reduces interview difficulties among respondents, helps to further enhance the degree of privacy, and appears to contribute positively to item-level response rates.

4.6 References

O'Reilly, J.M, Hubbard, M.L., Lessler, J.T., Biemer, P.P., & Turner, C.F. (1994). Audio and video computer assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics*, *10(2)*, 197-214.

Schwarz, N., & Hippler, H. (1991). Response alternatives: The impact of their choice and presentation order. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman, Eds., *Measurement errors in surveys,* 45-56. New York: John Wiley & Sons.

Substance Abuse and Mental Health Services Administration. (2001). *Development of computer-assisted interviewing procedures for the NHSDA*. Office of Applied Studies, Methodological Series M-3 (DHHS Pub. No. SMA 01-3514). Rockville, MD: Author.

Substance Abuse and Mental Health Services Administration. (2001b). *Summary of findings from the 2000 NHSDA*. Office of Applied Studies, NHSDA Series H-13 (DHHS Pub. No. SMA 01-3549). Rockville, MD: Author.

Tourangeau, R., & Smith, T.W. (1996). Asking sensitive questions. *Public Opinion Quarterly*, *60* (2), 275-304.

Turner, C.F., Lessler, J.T., & Gfroerer, J.C., Eds. (1992). *Survey measurement of drug use: Methodological studies* (DHHS Pub. No. ADM 92-1929). Rockville, MD: National Institutes of Health.

5

Development of Editing Rules for CAI Substance Use Data

Larry Kroutil and Lawrence Myers

As noted in ***Chapter 1***, a major change to the study protocol for the 1999 National Household Survey on Drug Abuse (NHSDA) was the shift from paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI). Although many of the substance use questions are similar in the two instruments, there are some differences. In addition, whereas the PAPI questionnaire required respondents to answer all questions in most sections, the CAI instrument makes extensive use of skip instructions. These significant differences in the nature of the data obtained in the new and old instruments necessitated the development of entirely new editing rules.

This chapter discusses the development of the new editing rules for the NHSDA CAI data and presents the results of an investigation of alternative editing methods for CAI data. The analysis was based primarily on data from the first 6 months of data collection in 1999.[1]

Section 5.1 discusses data quality issues that were affected by conversion to CAI. ***Section 5.2*** presents the general methodological approach used to define and test alternative rules for defining a usable case. ***Section 5.3***

[1]In late 1999, a fully weighted 6-month dataset was created to facilitate a number of preliminary analyses, including a review of substance use prevalence estimates, and to resolve key editing and imputation issues prior to the completion of the full year of data collection. Because the NHSDA employs a quarterly sampling methodology, these data were based on representative State and national samples of about half the sample size planned for the full-year study.

discusses alternative editing rules, along with the final rule that was implemented. ***Section 5.4*** presents selected substance use measures to compare the impact of the new CAI editing procedures with the former PAPI procedures.

5.1 Data Quality Issues Affected by Conversion to CAI

The CAI technology was designed to produce more internally consistent data while still allowing respondents to answer in privacy by using audio computer-assisted self-interviewing (ACASI) for the more sensitive parts of the interview. The ACASI approach allowed respondents to enter answers to sensitive questions directly into a hand-held computer, away from the view of the field interviewer or any other household members who were at home when the interview was conducted. The questions were displayed on the screen for the respondents to read. All respondents also were encouraged to listen on headphones to an audio recording of the questions. This feature of ACASI was especially useful for respondents with limited reading ability.

A major objective of introducing CAI technology was to improve the quality of the data by avoiding the inconsistencies and other data quality problems that arise naturally in the PAPI approach. Specifically, because of the paper-and-pencil administration, the substance use data collected through the PAPI instrument were subject to the following potential data quality problems:

- illegible responses, multiple response categories marked, or out-of-range values
- items left blank that respondents should have answered, with no apparent reason for the nonresponse
- inconsistencies among related variables.

Resolution of these data quality issues in PAPI was handled primarily through logical editing of the data, because respondents were not recontacted to supply missing information or resolve inconsistencies in their answers. These logical editing procedures were implemented through computer programs. Hence, the term that was commonly used for these procedures was "machine editing."

In addition, under the prior PAPI approach, the sensitive sections of the interview were completed in private by the respondent using self-administered answer sheets. The answer sheets pertaining to use of specific substances contained few or no skip patterns that allowed respondents to skip over questions that did not apply. Instead, respondents were instructed to respond to each question and were allowed to check an appropriate "not applicable" response if the question did not apply to them (e.g., never used the substance or used the substance but not in the period of interest). In principle, then, respondents who were not users would take approximately the same amount of time to complete an answer sheet as those who reported use. This procedure was designed to protect respondent confidentiality by concealing to the interviewers whether a respondent used substances. Further, skip patterns could provide an incentive for respondents to under-report substance use to avoid having to answer additional questions about use of that substance (i.e., a burden reduction issue).

However, requiring PAPI respondents to answer every item on a substance's answer sheet provided opportunities for them to answer inconsistently, such as reporting that they last used a substance more than 30 days ago, but then answering questions pertaining to the past 30 days as though they had used during that period. On the other hand, requiring respondents to answer every item on an answer sheet increased the amount of information available in the editing process for use in replacing missing data with nonmissing values.

In contrast, conversion of the NHSDA interview to a CAI format eliminated the problem of multiple responses because, for most questions, the computer program permits entry of only one response at a time per item.[2] Similarly, the instrument was programmed not to allow out-of-range values for certain items, such as frequency-of-use items, thereby reducing the amount and types of out-of-range values that would otherwise need to be addressed through machine editing.

In addition, the routing logic in the CAI instrument was designed to reduce the amount of missing data by skipping respondents past questions that did not apply to them instead of requiring respondents to repeatedly mark a "does not apply" response category. The routing logic also was designed to reduce the occurrence of inconsistent data by not giving

---

[2]In situations where respondents could enter more than one response that applied (e.g., specific psychotherapeutic drugs shown on a list), these responses were captured as separate variables.

respondents the opportunity to provide inconsistent answers. For example, respondents who reported last using a substance more than 12 months ago were skipped out of questions pertaining to their frequency of use of these substances during the past 30 days or past 12 months. The occurrence of inconsistent data was further reduced through the use of consistency checks built into the CAI program that prompt respondents to resolve inconsistencies that occur between related items.

Despite the potential for improvements in data quality through a CAI instrument, it was recognized that conversion to CAI would not completely eliminate data problems. For example, missing data would not be completely eliminated because CAI respondents still had the option of entering a response of Don't Know or Refused when answering a given item. Unlike the situation with PAPI (where respondents could leave items blank), however, analysts would know why a CAI respondent did not answer an item. In addition, items that were unanswered because of a breakoff would still have a missing value. Similarly, even though skip patterns and consistency checks were designed to reduce inconsistent reporting, opportunities for inconsistent data remained. For example, if a respondent did not resolve an inconsistent report when given the opportunity, the interview simply proceeded with the inconsistency in place.

5.2 Definition of a Usable Case

This step in the editing procedures establishes the minimum-item response required for cases to be retained for weighting and further analysis (i.e., "usable" cases). These procedures are designed to eliminate cases with unacceptable levels of item nonresponse (i.e., missing data), thereby retaining cases with lower levels of missing data and reducing the amount of statistical imputation that would be needed for any given record. If a respondent provided sufficient data to meet the usable case requirements, the case would be classified as a unit response, and any missing data would be accepted as item nonresponse and potentially subjected to further editing or imputation.

In the PAPI editing procedures, a case was considered "edit-complete" (i.e., usable) if the respondent had answered enough questions on the alcohol, marijuana, cocaine, and crack cocaine answer sheets to make a determination of whether the respondent was a lifetime user or nonuser of alcohol, marijuana, and cocaine (in any form). Because respondents were instructed to answer every question on these three answer sheets as well as on the crack cocaine answer sheet, identification of "edit-complete" cases in PAPI required machine edits to be run for the recency-of-use variables for alcohol, marijuana, and cocaine (including crack). For example, a respondent in PAPI could, in principle, leave the entire cocaine answer sheet blank but report use of crack cocaine on that answer sheet, in which case the PAPI edits would infer that the respondent was a cocaine user (and, therefore, that the case was potentially "edit-complete").

In CAI, however, respondents were asked more detailed questions about different substances only if they reported lifetime use of that substance (or lifetime use of one or more substances within a broader category, such as hallucinogens).[3] Consequently, whether a CAI respondent was a lifetime user or nonuser of substances of interest could be more readily determined from review of respondents′ answers to the raw lifetime question(s) for that substance (or category of substances). Further, CAI cases could be identified as usable (or not) at an earlier stage of the machine-editing process compared with the stage at which such cases were identified in the PAPI editing procedures.

5.2.1 Discussion of Alternative Rules

As noted above, an extensive series of analyses was undertaken late in 1999 on the first 6 months of data. For these preliminary analyses involving the first 6 months of data, a case was considered usable if the respondent provided data on lifetime use of cigarettes, alcohol, and marijuana. This is similar to the PAPI "edit-complete" requirements that lifetime use or nonuse be fully defined for three "criterion" substances, with the following exceptions: (1) one of the criterion substances for defining a usable case was cigarettes, not cocaine, and (2) identification of such cases could be accomplished earlier in the editing process, as noted above.

For this sensitivity study, three usable case rules were defined. The three variations on the usable case rules used in these analyses are shown in ***Table 5.1***. The more relaxed and more stringent rules were defined

---

[3]In all modules except the "Hallucinogens," "Inhalants," "Pain Relievers," "Tranquilizers," "Stimulants," and "Sedatives" modules, the logic for asking more detailed questions about use of that substance was based on the answer to a single yes/no question (e.g., "Have you ever, even once, used marijuana or hashish?"). In the "Hallucinogens" through "Sedatives" modules, the logic for asking more detailed questions about use of that category of substances was based on respondents' answers to multiple yes/no questions about the lifetime use or nonuse of specific substances within that category (e.g., lifetime use or nonuse of the specific hallucinogens LSD, PCP, peyote, mescaline, psilocybin/mushrooms, "Ecstasy," or "any other" hallucinogen).

relative to the rule that was used for the preliminary 6-month analysis (rule 2).

Table 5.1 Usable Case Rules, Quarters 1 and 2†

Rule No.	Description	Usable Cases
1	**More relaxed rule**: Respondent (R) must have provided basic demographics and must have responded to at least one lifetime use question.	26,528
2	**6-month rule**: R must have provided basic demographics and must have responded to the lifetime use questions for cigarettes, alcohol, and marijuana with a "yes" or "no" answer.	26,352
3	**More stringent rule**: R must have provided basic demographics; complete recency on cigarettes, alcohol, and marijuana; and lifetime use data on cocaine and heroin.	26,009

† Counts of the total number of usable cases resulting from the different rules are based on the first 6 months of data.

Reports for 10 substances were examined (cigarettes, chewing tobacco, snuff, alcohol, marijuana, cocaine, heroin, hallucinogens, prescription pain relievers, and prescription sedatives). For the more relaxed rule (rule 1), respondents needed to provide only some basic demographic information (i.e., age, sex) and provide information about lifetime use or nonuse for at least one of these substance categories. For hallucinogens, pain relievers, and sedatives, where respondents were asked multiple "yes/no" questions about their lifetime use of specific substances within the category (e.g., LSD, PCP, peyote, mescaline, psilocybin, "Ecstasy," or any other hallucinogen), the requirements were considered to be satisfied if at least lifetime use or nonuse was reported for at least one specific substance within the category. Consequently, cases that would be rejected under rule 2 because lifetime use or nonuse of cigarettes, alcohol, and cocaine were not fully defined might be retained as usable under the more relaxed rule. This rule would retain more cases for subsequent analysis, but would require more subsequent imputation for item nonresponse.

The more stringent usable case rule (rule 3) required not only that lifetime use or nonuse of cigarettes, alcohol, and marijuana be defined, but that the period when they last used each of these three drugs also be defined. In addition, lifetime use or nonuse of cocaine and heroin had to be defined. Consequently, a case where the respondents answered the lifetime questions for cigarettes, alcohol, and marijuana could still be rejected by this more stringent rule if a respondent answered with a Don′t Know or

Refused response when asked when he/she last used any of these three substances. In addition, cases could be rejected if they provided complete information on when they last used cigarettes, alcohol, and marijuana, but answered with a Don't Know or Refused response when asked whether they ever used cocaine or heroin.

5.2.2 Results for Alternative Rules

To examine the general quality of data included or excluded by each usable case rule, distributions were examined according to the number of substances for which the lifetime use question was answered. ***Table 5.2*** shows these results.

Table 5.2 Lifetime Use Response Patterns for 10 Selected Substances, Quarters 1 and 2

No. of Lifetime Use Questions Answered	Cases Admitted by Rule 3		Additional Cases Admitted by Rule 2		Additional Cases Admitted by Rule 1	
	N	%	N	%	N	%
10	25,648	98.61	280	81.63	0	0.00
9	301	1.16	38	11.08	77	43.75
8	47	0.18	15	4.37	31	17.61
7	10	0.04	4	1.17	15	8.52
6	3	0.01	4	1.17	19	10.80
5	0	0.00	2	0.58	9	5.11
4	0	0.00	0	0.00	10	5.68
3	0	0.00	0	0.00	3	1.70
2	0	0.00	0	0.00	2	1.14
1	0	0.00	0	0.00	10	5.68
Total	26,009	100.00	343	100.00	176	100.00

A total of 26,009 cases from the first 6 months of data collection would have been admitted as usable cases under the most stringent rule (rule 3). An additional 343 cases were admitted by rule 2 that were not admitted by the most stringent rule, and an additional 176 cases were admitted by the more relaxed rule (rule 1) that were not admitted by rule 2. A total of 85 cases did not meet even the most relaxed criteria for rule 1 (data not shown).

Table 5.2 also shows percentages for the number of lifetime use questions answered for the cases admitted by the different usable case rules. Under rule 3, almost 99 percent of the cases had a valid response ("yes" or "no") to the lifetime usage question for all 10 of the selected substances, and none had fewer than 6 of the 10 substances reported. The additional cases admitted by rule 2 had approximately 82 percent with valid responses to the lifetime usage question for all 10 of the selected substances; all cases had at least 5 of the 10 selected substances reported. Some of the cases admitted by rule 1 had a high proportion of missing data, but more than half had 8 or more of the 10 selected substances with responses on the lifetime usage question.

These options for defining cases as usable also were reviewed by NHSDA expert consultants and discussed in February 2000. One recommendation was to adopt a two-tiered approach in which cases that answered the lead (i.e., gate) questions for cigarettes, alcohol, and marijuana as yes or no would be considered usable (i.e., rule 2). Cases that did not make this first cut could still be considered usable if they answered a sufficient number of other gate questions as yes or no.

5.2.3 Development of Final Rule

However, subsequent analyses that were run on three quarters of data indicated that a much simpler set of rules could be applied that would retain numbers of cases similar to those of the more complicated two-tiered approach described above. These analyses on the first three quarters of data were based on examination of 14 sets of gate questions. All of the previous 10 were included, plus cigars, inhalants, prescription tranquilizers, and prescription stimulants. Data for lifetime use of crack cocaine were not checked because the logic for asking about crack cocaine was dependent upon the respondent having answered the lifetime cocaine question as yes. Although the CAI instrument also asked about pipe tobacco, this was not included because there was only one other question about pipe tobacco in addition to the gate question.

Results of this analysis from the first three quarters of 1999 NHSDA CAI data are shown in ***Table 5.3***. This analysis focused on the usable case rule that was used for the 6-month analysis (rule 2) and the more relaxed rule (rule 1).

The usable case rule for the 6-month analysis that required use or nonuse of cigarettes, alcohol, and marijuana to be fully defined would have retained a total of 47,399 cases from the first three quarters. Of these 47,399 usable cases, 99.5 percent answered all 14 of the above gate questions. Only 12 respondents who answered the cigarette, alcohol, and marijuana gate questions as yes or no provided data to fewer than 10 gates overall, and only 6 respondents provided data to fewer than 9 gates.

Table 5.3 Results for Usable Case Rules 1 and 2, Quarters 1 through 3

No. of Gates Answered	Qtr 1 - 3 Cases Admitted by Rule 2				Additional Qtr 1 - 3 Cases Admitted by Rule 1			
	N	%	Cumul N	Cumul %	N	%	Cumul N	Cumul %
14	47,157	99.49	47,157	99.49	0	0.00	0	0.00
13	139	0.29	47,296	99.78	159	48.48	159	48.48
12	45	0.09	47,341	99.88	42	12.80	201	61.28
11	26	0.05	47,367	99.93	21	6.40	222	67.68
10	20	0.04	47,387	99.97	24	7.32	246	75.00
9	6	0.01	47,393	99.99	11	3.35	257	78.35
8	2	0.00	47,395	99.99	9	2.74	266	81.10
7	2	0.00	47,397	100.00	5	1.52	271	82.62
6	2	0.00	47,399	100.00	4	1.22	275	83.84
5	0	0.00			21	6.40	296	90.24
4	0	0.00			9	2.74	305	92.99
3	0	0.00			2	0.61	307	93.60
2	0	0.00			4	1.22	311	94.82
1	0	0.00			17	5.18	328	100.00
0	0	0.00			0	0.00		
Total	47,399				328			

An additional 328 cases from the first three quarters answered at least 1 of the 14 gate questions but did not answer all 3 of the cigarette, alcohol, and marijuana gate questions. Of these 328 cases, 257 respondents provided information for 9 or more of the 14 gates. Thus, a rule that set a cutpoint based on the total number of gate questions with usable data, regardless of whether lifetime use (or nonuse) was fully defined for cigarettes, alcohol, and marijuana, would (1) be simpler to explain and implement, (2) result in a minimal loss of cases that fully answered all of these gate questions, and (3) include additional cases that met or exceeded the cutpoint but did not fully answer these gate questions.

However, one advantage of usable case rule 2 was that lifetime use or nonuse was fully defined for three substances. An advantage of having lifetime use or nonuse fully defined for at least one substance was that data for that substance could be used in subsequent statistical imputations for other substances where lifetime use/nonuse was undefined. In contrast, lifetime use or nonuse of at least one substance would not necessarily be fully defined for all cases admitted by a rule based solely on the number of gate questions answered. Therefore, a decision was made to include the requirement in the final usable case rule that lifetime use or nonuse of cigarettes had to be defined in order for a respondent's data to be considered usable.

The following requirements for the final usable case rule were decided upon:

- The lifetime cigarette gate question CG01 had to have been answered as yes or no, as noted above.
- At least nine of the following additional gates had to have answers of yes or no: chewing tobacco, snuff, cigars, alcohol, marijuana, cocaine (in any form), heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives.

As in rule 1, the multiple gate substances, hallucinogens through sedatives, were considered to have usable data if at least one lead lifetime question in the series was answered yes or no (e.g., if at least one question in the series LS01a through LS01h was answered yes or no for hallucinogens).

Implementation of this usable case rule on data from the first three quarters of 1999 resulted in 422 cases (0.9 percent) being classified as not usable. When this rule was run on the fourth quarter of data, an almost identical percentage did not meet the usable case criteria (196 cases, or 1.0 percent). Although the absolute numbers of cases not meeting the usable case criteria might appear high, more than half of the cases from the first three quarters and nearly half of the Quarter 4 cases that did not meet the usable case criteria had not even gotten to the very first substance use question pertaining to lifetime use of cigarettes (data not shown). This indicated that respondents broke off the interview in the initial questions on demographics or the interviews were terminated because the respondents were ineligible. Specifically, interviews were terminated prior to respondents being asked any substance questions if respondents were ineligible because they were too young (i.e., under age 12) or because they were on active military duty. Thus, ineligible cases by definition would not meet the usable case criteria.

5.3 Editing of Substance Recency Variables

Editing in the NHSDA has historically involved using data provided by respondents themselves to accomplish the following main objectives:

- to replace missing values with nonmissing values, including situations where questions were legitimately skipped, based on respondents′ answers to previous questions
- to make data in related items consistent.

Edits of the variables that establish when respondents last used a substance of interest (i.e., recency variables) are probably the most critical. These recency variables are the precursors for the final measures that establish the prevalence of use in the past 30 days, past 12 months, and lifetime.

Under the edits for the old PAPI format, as a general rule, if a respondent indicated in one question on a substance's answer sheet that he/she had never used a substance and indicated in another question that he/she had used a substance, logical editing coded the person as a user of that substance. If a respondent reported two (or more) different answers on the same answer sheet with respect to how recently he/she had used a substance, the editing procedures typically assigned the category indicating the more recent use. Relatively little statistical imputation was done to the PAPI recency variables following the editing step.

Conversion of the NHSDA instrument to a CAI format provided an excellent opportunity to reexamine the procedures and underlying assumptions for editing the recency variables. Further, the logic in the CAI instrument, in which respondents were skipped past questions that did not apply to them, precluded these same kinds of edits for the CAI recency variables as were done for PAPI.

5.3.1 Discussion of Alternative Rules

Four possible ways were examined for editing the CAI recency variables. The four rules are summarized in ***Table 5.4***. They vary in the extent to which they utilize other data in the questionnaire to resolve questions about recency of use or use statistical imputation. ***Table 5.5*** compares how these four edit rules handled different types of conflicting information that can appear within a substance's section. Edit rule 2 was used for the preliminary 6-month analysis.

Table 5.6 summarizes the strengths and limitations of these four different edit rules. For example, a potential strength of edit rule 1 is that it makes fewer revisions to a respondent's original answers to the recency questions. Thus, if a respondent originally reported last using a substance more than 12 months ago, edit rule 1 would retain this answer in most situations. Consequently, there would be less missing data for questions pertaining to use of the substance in the past 12 months or past 30 days because one could infer that a respondent did not use during these periods. However, a potential drawback of edit rule 1 is that it typically would ignore other information provided by a respondent in his/her interview record. With this approach, analysts may conclude that the data are

Table 5.4 Edit Rule Descriptions

Rule No.	Description
1	**"Lower-bound" edit rule**: Generally disregards data that would suggest more recent use than what the respondent (R) originally indicated. For alcohol through inhalants, where Rs can originally report use in the past 30 days but then report use on 0 days in that period, this rule also sets Rs to being past year users if they confirmed use on 0 days in the past month. Exceptions are that more recent crack use will infer more recent cocaine use, reports of more recent use in response to consistency checks still assign the R to the category indicating more recent use, and Rs who do not resolve inconsistencies between related recency variables (e.g., LSD and any hallucinogen) are set to the period indicating more recent use.
2	**6-month analysis rule**: Generally uses data from that drug's module to infer more recent use than what the R originally indicated. However, confirmations of use on 0 days in the past month set an R to being a past year user, unless the original answer of past month use is supported by other data suggesting past month use.
3	**"Upper-bound" edit rule:** As is the case with rule 2, generally uses any information within a substance's module to infer more recent use. In addition, disregards confirmations of use on 0 days in the past month, such that an R's original answer of most recent use in the past 30 days is retained. However, responses to consistency checks that change an answer to less recent use can still set the recency to indicate less recent use than what the R originally indicated.
4	**"Flag and impute" rule**: This rule flags inconsistencies between a recency variable and related variables but does not make a decision about the final recency category. Rather, this rule leaves these inconsistent recency-of-use data to be statistically imputed. For example, if an R originally reported last using a drug in the past 30 days but then confirmed use on 0 days in that period, edit rule 4 would infer that the R is *at least* a past year user. In the imputation procedures, this case would be eligible to be imputed to use in the past 30 days or use more than 30 days ago but within the past 12 months.

"cleaner" than they actually are. In addition, respondents may deliberately report less recent use in order to avoid being asked more detailed questions about their use in the past 12 months or past 30 days. Therefore, this approach could underestimate the true prevalence of use, particularly for less commonly used substances.

Edit rules 2 and 3 make more use of other data in a respondent's record to arrive at a decision about when the respondent last used a substance. Although use of the CAI instrument limits situations where the respondent potentially indicated more recent use than he/she indicated in the recency question, these rules would yield less conservative estimates than those obtained under rule 1. However, these rules could be criticized as "making up data," and some estimates might be disputed because of how the data were edited.

Unlike the first three rules, edit rule 4 takes the editing task away from second-guessing where the truth lies, which is a major strength. It leaves all inconsistent data to be statistically imputed. A second advantage is the ease with which it can be explained to people who are not intimately familiar with the NHSDA instrument or data. However, implementation of this approach could still be seen as "making up data," albeit through statistical techniques. Furthermore, the more data that must be imputed relative to the number of respondents who reported use in a given period (i.e., and whose recency was not imputed), the less defensible a given estimate will be. In addition, some editing may still be needed for inconsistencies not directly related to the recency-of-use variables (e.g., unresolved inconsistencies between the number of days in the past 30 days when respondents had five or more drinks on a single occasion and the total number of days that they drank in that period).

5.3.2 Results for Alternative Rules

Data from the first two quarters of the 1999 NHSDA were used to assess the anticipated effects of the choice of an edit rule on substance use prevalence estimates. Because of the potential for interaction between usable case and edit rules, edit rule results were compared within each usable case rule. (Original analyses were performed using data that had been imputed after having been edited. However, the resulting data obscured the effects on prevalence that could be attributed directly to editing.)

The following measures of substance use were created for this chapter using the edited variables from the first two quarters of data: (1) any lifetime use (which could include use in the past 12 months or past 30 days); (2) any past year use (which could include use in the past 30 days); and (3) past month use (i.e., use in the past 30 days). In creating these substance use measures, data from respondents who reported using a substance but whose most recent use of that substance was not fully defined were handled as follows:

- If a respondent was unclear about past month use, but unambiguously indicated past year use, then the respondent was classified as a lifetime and past year user, with missing data for past month use.

Table 5.5 How the Four Different Edit Rules Handle Inconsistencies Involving CAI Recency Variables

Edit Rule 1	Edit Rule 2	Edit Rule 3	Edit Rule 4
Type of Inconsistency: Recency originally indicates use in past 30 days, but use on 0 days in past 30 days is confirmed.			
Infers that the respondent (R) is a past-year (but not past-month) user and retains the answer of use on 0 days.	Same as edit rule 1.	Ignores the report of use on 0 days; retains the original answer of most recent use in the past 30 days.	Infers that the R is at least a past year user, and potentially a past month user.
Type of Inconsistency: Recency originally indicates use more than 30 days ago but within past 12 months, but the 12-month frequency indicates use on more than 335 days in that period.			
Ignores the 12-month frequency data; retains original answer from the recency question.	Infers that the R is a past month user.	Same as edit rule 2.	Infers that the R is at least a past year user and potentially a past month user.
Type of Inconsistency: Recency does not indicate use in past 30 days, but the R reports first using the drug in the same month as interview took place.			
Ignores the month- and year-of-first-use data; retains original answer from the recency question.	Infers that the R is a past month user.	Same as edit rule 2.	If the recency originally indicated use in the past 12 months, infers that the R is at least a past year user and potentially a past month user. If the recency originally indicated lifetime use (or was missing), infers that the R is at least a lifetime user (and potentially a past year or past month user).
Type of Inconsistency: Recency does not indicate use in past 12 months, but the age-at-first-use equals the R's current age.			
Ignores the age-at-first-use data; retains original answer from the recency question.	Infers that the R is a past year user.	Same as edit rule 2.	Infers that the R is at least a lifetime user (and potentially a past year or past month user).
Type of Inconsistency: Recency does not indicate use in past 12 months, but the R reported first using in a month and year that falls within 12 months of interview date.			
Ignores the month- and year-of-first-use data; retains original answer from the recency question.	Infers that the R is a past year user.	Same as edit rule 2.	Infers that the R is at least a lifetime user (and potentially a past year or past month user).

Table 5.5 (Continued)

Edit Rule 1	Edit Rule 2	Edit Rule 3	Edit Rule 4
Type of Inconsistency: Crack use is more recent than use of any cocaine (no consistency check available).			
Infers that the R last used cocaine in the period indicated by the crack recency.	Same as edit rule 1.	Same as edit rule 1.	If the cocaine recency originally indicated use in past 12 months, infers that the R is at least a past year user of both cocaine and crack. If the cocaine recency originally indicated lifetime use, infers that the R is at least a lifetime user of cocaine and crack.
Type of Inconsistency: R does not resolve inconsistencies between most recent LSD/PCP use and use of any hallucinogen despite being prompted to do so; R does not resolve inconsistencies between most recent methamphetamine use and use of any stimulant despite being prompted to do so.			
Assigns the most recent period of use indicated, based on the consistency check information.	Same as edit rule 1.	Same as edit rule 1.	If the "parent" (e.g., hallucinogen) recency originally indicated use in past 12 months, infers that the R is at least a past year user. If the "parent" recency originally indicated lifetime use, infers that the R is at least a lifetime user (and potentially a past year or past month user).

Table 5.6 Comparative Strengths and Limitations of the Four Different Edit Rules for CAI Recency Variables

Edit Rule	Strengths	Drawbacks
1	• Preserves more of the respondent's (R's) original answers; requires less editing than other options • Less vulnerable to criticism of "making up data" • Still provides a means of addressing inconsistencies that Rs did not resolve themselves • Produces fewer missing data for 30-day and 12-month variables because they can be inferred to have been legitimately skipped • Produces fewer missing data for "non-core" sections that are administered only if the R is a past year user	• Ignores other information provided directly by the R • May underestimate use of some drugs, if Rs learn to report less recent use or to refuse in order to skip out of 30-day and 12-month sections • May underestimate use of some drugs due to "time boundary" problems with recency questions (e.g., if an R is on borderline between use in past 12 months and use more than 12 months ago) • Allows Rs who originally reported past 30-day use to reverse themselves and choose a more socially desirable answer
2	• Can compensate for potential underreporting, although such opportunities are limited • Allows data to be used to replace responses of "don't know/refused" with nonmissing values • Attempts to take into account a broader set of the R's answers, instead of ignoring other data • Unlike PAPI edit rules, edit decisions are not always in the direction of inferring more recent use	• Open to criticism of "making up" data • Some edit decisions may not be as defensible as others • Logically inferred past month and past year users have missing data for the 12-month and 30-day items in a section because they were skipped out based on their original answer to the recency question
3	• Same potential strengths as edit rule 2, except that edit decisions are even more likely to be in the direction of inferring more recent use • Does not allow Rs who originally reported past month use to reverse themselves to choose a more socially desirable/less difficult answer	• Same potential drawbacks as edit rule 2
4	• Resolution of inconsistencies through statistical methods may be more defensible to users of NHSDA data • Still allows Rs to be assigned to a more recent category of use, but based on statistical methods • Does not involve second-guessing what the "truth" is if an R answered inconsistently • Simple rule to explain: If the recency is inconsistent with other data, impute it	• Still may be viewed as a form of "making up" data • Edit rules for related recency variables (e.g., LSD and any hallucinogen) can still be fairly complicated; resulting imputations for related recency variables also can get complicated • Requires more imputing than other approaches; strength of the approach is diminished if much imputation is required • Could delay subsequent data processing steps or yield missing/inconsistent data if processing of other variables proceeds before imputations are finished • Some editing decisions may still be required for inconsistencies that do not involve the recency variables (e.g., frequency of consumption of five or more drinks per occasion in the past 30 days); otherwise, this would further increase the amount of imputation needed.

- Similarly, if it could be determined that a respondent's last use of a substance occurred at some point in his/her lifetime (but could have been in the past 12 months or past 30 days), the respondent was classified as a lifetime user but was considered to have missing data for past year and past month use.

Consequently, within a given usable case rule, sample sizes varied depending on whether the drug measure of interest was assigned a missing value for a given edit rule. Within usable case rule 3 (most stringent rule), for example, cell sizes ranged from 25,646 for past month alcohol use under edit rule 4 to 26,009 for lifetime measures of cigarettes, alcohol, marijuana, cocaine, and heroin (same sample size for all four edit rules). For the analyses described in this section, however, no adjustments were made to the 6-month analysis weights to correct for varying amounts of missing data.

Table 5.7 shows odds ratios based directly on the weighted data for the total population aged 12 years and older; odds ratios also were computed for the age groups 12 to 17, 18 to 25, and 26 and older. Aside from calculating odds ratios for specific age groups, no adjustments were made for other potential covariates such as gender or race/ethnicity. The SAS statistical software package was used to compute these unadjusted odds ratios. The unadjusted odds ratios were calculated as $[P_1/(1-P_1)]/[P_2/(1-P_2)]$, where P_1 was the weighted prevalence estimate for a given substance under edit rule 1, 3, or 4 and P_2 was the weighted prevalence estimate under edit rule 2. For example, if a given substance's past month prevalence was 1.0 percent under edit rule 2 and was 0.95 percent under edit rule 1 (not actual data), the resulting unadjusted odds ratio would be 0.9495, calculated as follows:

> $[.0095 / (1 - .0095)] / [.0100 / (1 - .0100)]$, where $P_1 = 0.0095$ (prevalence from edit rule 1) and $P_2 = 0.0100$ (prevalence from edit rule 2).

The above odds ratio indicates that respondents would have approximately 95 percent of the odds of being classified as a past month substance user under edit rule 1 compared with edit rule 2. However, when prevalence estimates are low, as often is the case for illegal substance use, the actual difference in prevalence can be small, as shown by this example.

As shown in ***Table 5.5***, many of the potential inconsistencies between the reported most recent use of a substance and other data in a module applied to recent initiates of a substance, such as those who reported first using a substance at their current age. Stated differently, recent initiates

Table 5.7 Unadjusted Odds Ratios Based on Edited Data for Persons Aged 12 or Older

	Usable Case/Edit Rule								
	Usable Case Rule 1			Usable Case Rule 2			Usable Case Rule 3		
	Edit Rule			Edit Rule			Edit Rule		
Period of Use/Drug	**1**	**3**	**4**	**1**	**3**	**4**	**1**	**3**	**4**
Lifetime Use									
Cigarettes	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Smokeless tobacco	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Alcohol	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Marijuana	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cocaine	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Heroin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hallucinogens	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Pain relievers	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Sedatives	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Past Year Use									
Cigarettes	0.9920	1.000	0.9943	0.9920	1.000	0.9943	0.9923	1.000	0.9946
Smokeless tobacco	0.9807	1.000	0.9817	0.9808	1.000	0.9817	0.9810	1.000	0.9819
Alcohol	0.9917	1.000	0.9967	0.9917	1.000	0.9967	0.9920	1.000	0.9970
Marijuana	0.9854	1.000	0.9865	0.9856	1.000	0.9867	0.9869	1.000	0.9880
Cocaine	0.9657	1.000	0.9319	0.9657	1.000	0.9318	0.9670	1.000	0.9329
Heroin	0.9357	1.000	0.9358	0.9357	1.000	0.9358	0.9357	1.000	0.9358
Hallucinogens	0.9316	1.000	0.9287	0.9314	1.000	0.9285	0.9319	1.000	0.9298
Pain relievers	0.9778	1.000	0.9784	0.9777	1.000	0.9783	0.9791	1.000	0.9797
Sedatives	0.9900	1.000	0.9901	0.9900	1.000	0.9901	0.9900	1.000	0.9901
Past Month Use									
Cigarettes	0.9977	1.000	0.9999	0.9977	1.000	0.9999	0.9977	1.000	0.9999
Smokeless tobacco	0.9978	1.000	0.9988	0.9978	1.000	0.9988	0.9978	1.000	0.9988
Alcohol	0.9909	1.017	0.9971	0.9909	1.017	0.9971	0.9909	1.017	0.9971
Marijuana	0.9905	1.021	0.9924	0.9905	1.021	0.9924	0.9904	1.021	0.9924
Cocaine	0.9035	1.012	0.8924	0.9035	1.012	0.8924	0.9025	1.012	0.8912
Heroin	0.6984	1.000	0.6986	0.6984	1.000	0.6986	0.6984	1.000	0.6986
Hallucinogens	0.9845	1.040	0.9621	0.9845	1.040	0.9620	0.9851	1.040	0.9622
Pain relievers	0.9938	1.000	0.9945	0.9938	1.000	0.9944	0.9938	1.000	0.9944
Sedatives	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

could have more opportunities to provide data that were inconsistent with their responses on when they last used a substance compared with respondents who had initiated use less recently. Therefore, ***Table 5.8*** presents unadjusted odds ratios specifically for the 12- to 17-year-old age group.

Test models for measures of lifetime, past year, and past month drug use also were run on data from usable case rule 2 using SUDAAN PROC MULTILOG. This procedure allowed observations on the different edit rules to be treated as repeated measures from the same respondent. The

Table 5.8 Unadjusted Odds Ratios Based on Edited Data for Youth Aged 12 to 17

	Usable Case/Edit Rule								
	Usable Case Rule 1			Usable Case Rule 2			Usable Case Rule 3		
	Edit Rule			Edit Rule			Edit Rule		
Period of Use/Drug	**1**	**3**	**4**	**1**	**3**	**4**	**1**	**3**	**4**
Lifetime Use									
Cigarettes	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Smokeless tobacco	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Alcohol	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Marijuana	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Cocaine	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Heroin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hallucinogens	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Pain relievers	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Sedatives	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Past Year Use									
Cigarettes	0.9417	1.000	0.9542	0.9419	1.000	0.9543	0.9451	1.000	0.9576
Smokeless tobacco	0.9203	1.000	0.9238	0.9214	1.000	0.9248	0.9218	1.000	0.9253
Alcohol	0.9402	1.000	0.9587	0.9400	1.000	0.9585	0.9450	1.000	0.9637
Marijuana	0.9386	1.000	0.9461	0.9395	1.000	0.9470	0.9465	1.000	0.9540
Cocaine	0.8811	1.000	0.8504	0.8811	1.000	0.8504	0.8890	1.000	0.8574
Heroin	0.9286	1.000	0.9289	0.9286	1.000	0.9289	0.9286	1.000	0.9289
Hallucinogens	0.8738	1.000	0.8692	0.8738	1.000	0.8691	0.8756	1.000	0.8709
Pain relievers	0.9604	1.000	0.9621	0.9603	1.000	0.9619	0.9630	1.000	0.9646
Sedatives	0.9417	1.000	0.9420	0.9417	1.000	0.9420	0.9417	1.000	0.9420
Past Month Use									
Cigarettes	0.9863	1.000	0.9984	0.9864	1.000	0.9984	0.9863	1.000	0.9983
Smokeless tobacco	0.9893	1.000	0.9933	0.9893	1.000	0.9932	0.9892	1.000	0.9931
Alcohol	0.9773	1.044	0.9938	0.9772	1.044	0.9938	0.9774	1.045	0.9940
Marijuana	0.9747	1.021	0.9855	0.9747	1.021	0.9855	0.9746	1.021	0.9854
Cocaine	0.9624	1.047	0.9073	0.9624	1.047	0.9073	0.9608	1.049	0.9030
Heroin	0.6495	1.000	0.6501	0.6495	1.000	0.6501	0.6495	1.000	0.6501
Hallucinogens	0.9445	1.042	0.9253	0.9445	1.042	0.9253	0.9468	1.043	0.9273
Pain relievers	0.9899	1.000	0.9917	0.9898	1.000	0.9916	0.9898	1.000	0.9916
Sedatives	0.9997	1.000	1.000	0.9997	1.000	1.000	0.9997	1.000	1.000

models included main effects for the edit rule, age group, and race/ethnicity, as well as interactions between each pair of demographic characteristics. Variances were evaluated using the robust variance estimation procedures (Binder's method) within SUDAAN. This approach also used the standard design assumptions normally used for producing tables from the NHSDA.

However, the model-based adjusted odds ratios from these exploratory analyses varied little from the corresponding unadjusted odds ratios. Furthermore, the large sample sizes yielded differences in the adjusted odds ratios that were of statistical significance, even though many of the odds ratios were close to unity. Stated differently, situations in which odds ratios were close to one another suggest that the differences were of statistical, not practical, significance. Therefore, for the sake of simplicity, the remaining discussion in this section focuses on the unadjusted odds ratios.

Within a given usable case rule, there was no variation among the edit rules in terms of classifying respondents as lifetime users or nonusers of a particular substance. This pattern held for each age group (data not shown for age groups 18 to 25 and 26 and older). The different edit rules had no effect on estimates of lifetime use, because lifetime use or nonuse was established from the "gate" questions. Once lifetime use had been established from respondents' answers to a substance's gate question(s), subsequent questioning in CAI about use of a given substance was predicated on the respondent being a user. Consequently, the edit rules developed for the CAI method of administration would affect the period of use to which a respondent was classified, once the respondent had already been identified as a user.

For estimates of past year use, there was no variation between edit rules 2 and 3. As shown in ***Table 5.5***, edit rules 2 and 3 handled inconsistencies in a similar manner with respect to use in the past year, but differed in how they handled inconsistencies that might have inferred past month use. For all past year estimates shown in ***Table 5.7***, rules 1 and 4 yielded lower estimates of past year use after editing had been done compared with estimates produced by edit rules 2 and 3. For the overall population, however, the odds ratios for edit rules 1 and 4 were still close to unity. Furthermore, the properties of edit rule 4 could allow respondents to be assigned to a category indicating more recent use; however, any such assignment would be made through statistical procedures.

Similar patterns held for past year use among youths aged 12 to 17. However, odds ratios for past year cocaine use and past year hallucinogen use were less than 0.90 for edit rules 1 and 4, relative to edit rule 2. These notably lower odds ratios for cocaine and hallucinogens among youths were probably due to the low prevalence of use of these substances, such that even modest differences between the edit rules had greater impacts on prevalence estimates. Specifically, 6-month edited estimates for past year cocaine use in this age group ranged from 1.35 to 1.40 percent for edit rules 1 and 4, compared with estimates of 1.57 to 1.59 percent for edit rules 2 and 3 (data not shown). Similarly, 6-month edited estimates for past year hallucinogen use among youths ranged from 3.48 to 3.50 percent for edit rules 1 and 4, compared with 3.97 to 3.98 percent for edit rules 2 and 3 (data not shown).

For past month use, odds ratios for edit rule 3 relative to rule 2 were consistently greater than or equal to 1. That finding is consistent with the properties of rule 3, which edited cases to be past month users in situations in which rule 2 did not (***Table 5.5***). As was the case with past year estimates, edit rules 1 and 4 yielded odds ratios less than 1, relative to rule 2. As was the case with past year use, differences between the edit rules were most pronounced for the less commonly used substances. In particular, respondents had only about 70 percent of the odds of being classified as past month heroin users under edit rules 1 and 4, compared with their odds of being classified as past month users under edit rule 2. Within each usable case rule, however, there were only 24 respondents (unweighted) who were assigned to past month heroin use status under edit rules 1 and 4 and only 28 respondents who were assigned to this status under rules 2 and 3 (data not shown).

Overall, then, these data from the first 6 months of the 1999 NHSDA would suggest that the choice of edit rules would have little practical impact on substance use estimates. Nevertheless, the impact of edit rules 2 and 3 was more noticeable for the less commonly used substances, based on a different analysis of the proportion of users assigned to a given status through editing. In particular, about 7 or 8 percent of past year hallucinogen users, 11 or 12 percent of past year cocaine users, and 17 percent of past year heroin users underwent editing based on rules 2 and 3 (data not shown).[4] Similarly, about 11 percent of past month cocaine users were assigned to that status through edit rules 2 and 3. The impact of these two rules was even more striking for past month heroin use when the data were weighted. Specifically, 30 percent of past month heroin users were assigned to that status through edit rules 2 and 3. In comparison, only about 1 to 2 percent of past month cocaine users and less than 8 percent of past month heroin users were assigned to that status through edit rule 4 and subsequent imputation.

5.3.3 Summary of Final Rules

As discussed earlier, the logic in the computer-assisted interviewing instrument limited the amount of data that can be used in editing a

---

[4]These percentages of users assigned to a given status through editing are based on 6-month weighted data.

respondent's answers. This was borne out by the odds ratio data presented in the preceding section, which showed minimal differences between edit rules.

Therefore, the final edit rule that was adopted was primarily the "flag and impute" rule (edit rule 4). The beauty of this edit rule lies in its simplicity: If a respondent gives an answer within a substance's section that conflicts with the original answer to the recency-of-use question, then statistically impute the recency.

A second attractive feature of this rule is that, if the respondent provides conflicting information, it is not necessary to try to deduce from the data when the respondent last used the substance. Moreover, this rule does not automatically discount indications of more recent use than the respondent originally reported in a recency question (as edit rule 1 often does), nor does it automatically infer that the respondent last used a substance more recently than he/she originally reported (as rules 2 and 3 often do). In particular, decisions in edit rules 2 and 3 to infer more recent use can have an appreciable impact on estimates of current (i.e., past month) use for less commonly used substances, such as cocaine and heroin. Although edit rule 4 requires more statistical imputation than the other three edit rules, that may be a relatively small price to pay, especially for the less commonly used substances.

However, one modification was made to this basic "flag and impute" rule for recency-of-use variables that were related (i.e., any cocaine and crack cocaine, any hallucinogen and LSD or PCP, any stimulant and methamphetamine). If the more specific recency variable (e.g., crack cocaine) indicated more recent use than did the more general recency variable (e.g., any cocaine), the edit procedures inferred more recent use for the general recency. The assumption behind this edit was that respondents may not necessarily have made the connection between the two related substances (e.g., making the connection that crack is a type of cocaine).

5.4 Comparisons with Former Edit Procedures

Figures 5.1 and ***5.2*** show the effects of editing and imputation for past month marijuana use and cigarette use in the 1999 CAI and 1998 data (Bowman, Penne, Westlake, & Singh, 1999). Weighted percentages are shown in these tables for raw data prior to any editing, following editing, and following imputation.

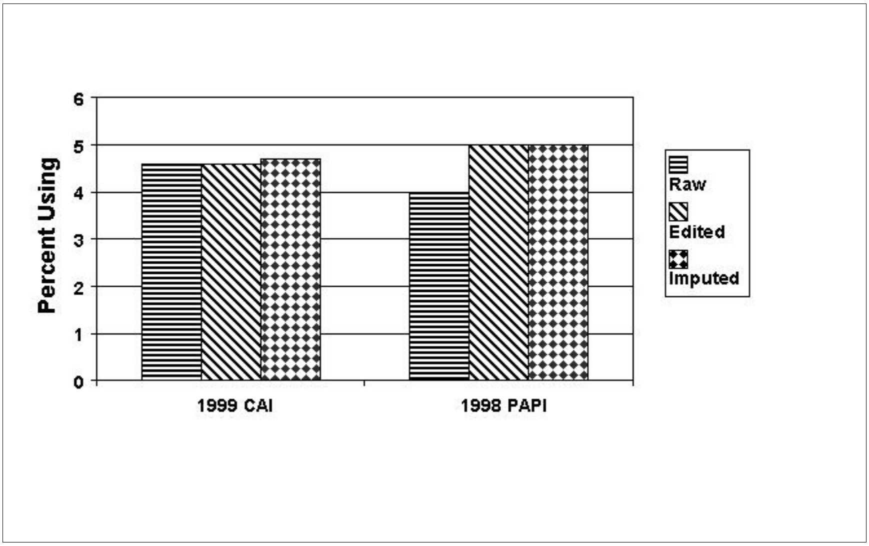

Figure 5.1 Past month marijuana results, 1999 CAI and 1998 PAPI.

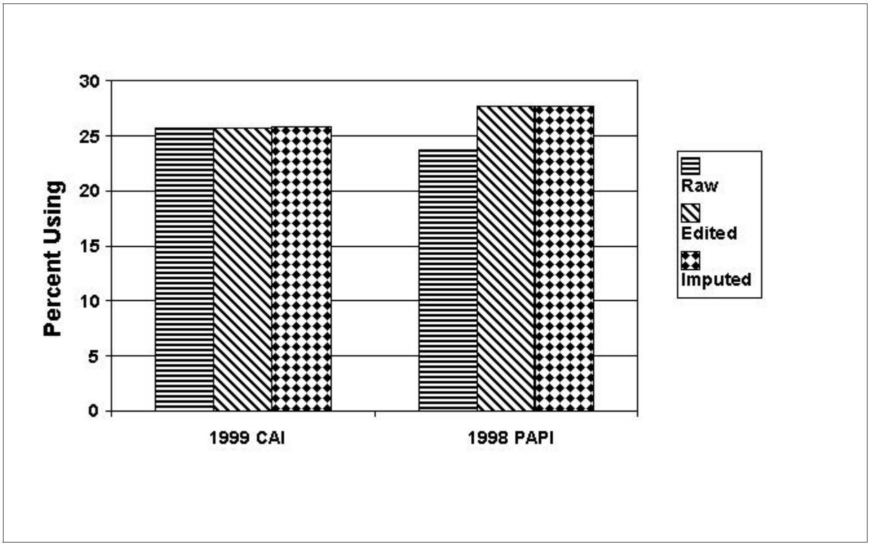

Figure 5.2 Past month cigarette results, 1999 CAI and 1998 PAPI.

For both of the substances shown in ***Figure 5.1***, there was little change in going from the raw CAI data provided by the respondents to the final imputed estimates. For example, the estimate of marijuana use in the past month changed from 4.6 percent in the raw data to 4.7 percent after editing and imputation. In comparison, editing made a greater contribution to estimates of past month marijuana use and cigarette use in the 1998 data. For past month marijuana use, the raw estimate in 1998 was 4.0 percent (weighted), the estimate after editing was 5.0 percent, and the final imputed estimate was 5.0 percent. Thus, the editing procedures that had been used in the NHSDA since 1994 increased the 1998 estimate of past month marijuana use by about 25 percent relative to the raw data; the additional impact of imputation on the final estimate of past month marijuana use was virtually nil.

Differences in the impacts of the new 1999 CAI editing and imputation procedures and those used in prior years were even more pronounced for less commonly used substances. In terms of unweighted data,[5] 66 respondents in the 1999 CAI sample reported last using heroin in the past 30 days, which decreased to 64 after editing.[6] There were 36 CAI respondents following editing who were lifetime or past year heroin users who were potentially eligible to be imputed to be past month users, and there were another 75 respondents following editing who were eligible to be imputed to be either users or nonusers because their lifetime use or nonuse of heroin was not known. Following imputation, a total of 70 CAI respondents were classified as past month heroin users, or a net increase of only four cases relative to the raw and a net increase of six relative to the edited. In comparison, the editing procedures in 1998 nearly doubled the number of respondents classified as being past month heroin users (17 respondents in the raw data and an additional 11 cases who were assigned to this category through editing). No additional cases in 1998 were statistically imputed to be past month heroin users (Bowman, Penne, Westlake, & Singh, 1999). Thus, the new editing and imputation procedures for the 1999 CAI sample had less of an appreciable impact overall on the past month heroin data. Further, the CAI editing procedures had the potential to *decrease* the number of respondents classified as past month users. Thus, the net increase in respondents classified as past month heroin users occurred

[5]Weighted estimates comparing the effects of editing and imputation in the 1999 CAI sample had been run only for the priority substance measures discussed previously.

[6]Two respondents who originally reported last using heroin in the past month subsequently reported that they used it on "0 days" in that period.

through imputation, not in the editing. In all, these changes in the 1999 CAI editing procedures represent an improvement over the way missing or inconsistent data had been handled in that these issues are resolved primarily through statistical methods.

5.5 References

Bowman, K.R., Penne, M.A., Westlake, M.J., & Singh, A.C. (1999). *1998 National Household Survey on Drug Abuse. Editing, imputation, and outlier weight evaluation report*. Report prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-97-9007. Research Triangle Park, NC: Research Triangle Institute.

6

Predictive Mean Neighborhood Imputation for NHSDA Substance Use Data

Avinash Singh, Eric Grau, and Ralph Folsom, Jr.

In 1999, the instrument used to administer the National Household Survey on Drug Abuse (NHSDA) was changed from a paper-and-pencil format (PAPI) to a computer-assisted format (CAI).

In previous years, imputation of missing values in recency of use and frequency of use in the past 12 months was accomplished with an unweighted sequential hot deck procedure. In the spirit of improving the quality of estimates from the redesigned NHSDA and as a result of fundamental differences between PAPI and CAI, there was a need to change the way missing data were edited and imputed. The implementation of the "flag and impute" editing rule, described in Chapter 5, and the desire to impute more variables required a new method that was rigorous, flexible, and preferably multivariate. This chapter presents a new imputation method with these characteristics, termed Predictive Mean Neighborhoods (PMN), that was used to impute missing values in the NHSDA substance use variables.

Following a discussion of background in ***Section 6.1***, this chapter outlines the previously used hot deck method, along with its limitations, in ***Section 6.2***. The new method is described in general in ***Section 6.3***, followed by details of the method in ***Section 6.4***. ***Section 6.5*** compares the method with other available methods and provides details concerning the motivation for employing a new method. In the concluding section

(***Section 6.6***), the impact of imputation on substance use estimates is compared between PAPI and CAI.

6.1 Background

As is explained in Chapter 5, changes in the questionnaire format required changes in the editing rules. In particular, respondents had many opportunities in the PAPI questionnaire to enter inconsistent data, and these data were used in the editing process. In CAI, a respondent could be automatically skipped out of questions that did not apply to his/her usage status, and this greatly reduced the opportunities for the respondent to enter inconsistent data. As a result, it was no longer possible to do the extensive editing with the CAI data that was possible with PAPI. Even with the skip logic, however, inconsistent data still existed. With this reduced amount of inconsistent data, the "flag-and-impute" rule became the preferred approach.

These changes in turn required changes in the way missing data were imputed. With the editing rules used in previous years, inconsistent responses were usually edited toward more recent use. Hence, in most cases, imputation was required only if the response was completely missing. This resulted in editing having a larger role than imputation in the final estimates. However, with the flag-and-impute rule, nearly all inconsistent cases are left unresolved, to be imputed to a final value. This changed the relative contribution of edited and imputed values in the final estimates, so that a greater burden was placed on statistical imputation in 1999 relative to editing, as opposed to previous years. In addition, imputation was required for more variables in 1999. While recency of use and frequency of use in the past 12 months were the only substance use variables with missing values imputed in previous years, missing values in the 30-day frequency of use, age at first use, 30-day binge drinking frequency, and age at first daily cigarette use variables were also imputed in the CAI data.

The patterns of missingness among these variables can be complex, with missing values in some variables necessitating missingness in other related variables. For example, a respondent may indicate that he/she is a lifetime user of heroin. But if the respondent does not indicate whether he/she is a past year user, not only is the recency of use variable missing, he/she is also unable to answer questions about frequency of use in the past 12 months and the past 30 days.

These points illustrate the following: (1) it is imperative that an effective and flexible method of statistical imputation be employed, and (2) since the substance use measures are so closely related to each other (sometimes in a hierarchical manner), it would be advantageous to have a multivariate imputation procedure, which would maintain the relationship between the variables with missing data. However, the main problems in developing a multivariate imputation method are the presence of both discrete and continuous variables in the data set and the inability to specify a suitable covariance structure in the multivariate model. It will be shown later how the problems of multivariate modeling can be overcome with PMN by working with a series of univariate models. These models are defined conditionally and fitted sequentially according to a predetermined hierarchy of response variables.

6.2 Description and Limitations of Method used in Previous Years and the 1999 PAPI Sample

In previous NHSDAs, an (unweighted) sequential hot deck was the tool of choice for almost all of the imputation-revised variables, which included both demographic and substance use measures. Hot deck imputation involves the replacement of a missing value with a valid code taken from another respondent who is "similar" and has complete data. The data set is first partitioned into imputation classes, which are defined according to variables that are very closely related to the variable with missing values to be imputed. Responding and nonresponding units are then sorted together within these imputation classes by a variable or collection of variables closely related to the variable of interest Y. When applied to the NHSDA, an additional random number was added as the final sort variable under a priority rule for sorting variables. For sequential hot deck imputation (Little & Rubin, 1987, p. 62), a missing value of Y is replaced by the nearest responding value preceding it in the sequence. This method is random only to the extent that the final random number is used to break the ties on the other sorting variables.

Although the (unweighted) sequential hot deck was simple and quick to implement, it had a number of disadvantages, which are outlined below:

- *The ability to use covariates to determine donors is limited*. The first few sorting covariates (often only the first and second covariate) almost entirely determine what donor will be used for a particular respondent with missing data, regardless of how many sorting covariates are included. This is due to the sparseness of donors with matching variable levels for variables further down the sort order. For many of these substance use measures, age was the most important covariate that could predict level of substance use. However, many other variables would also be useful as indicators of substance use, including use of other substances. The ability of the hot deck procedure to account for this is limited.

- *The relative importance of covariates is determined a priori***.** Before implementing the hot deck, the order of the covariates must be decided. This order determines how closely donors and recipients will match on a given covariate. There is no mechanism based on the data to weight the covariates based on their relationship to the response variable.
- *Weights are not used to determine the most appropriate donor for a respondent with missing data*. This would not introduce bias in the estimation of population means and totals if the unweighted mean of donors is the same as the weighted mean of donors. This is unlikely in practice, however. The donor class is usually very small because it consists of respondents with ties on the other sorting variables. A way out would be to use the weighted hot deck whenever there are at least two respondents in the donor class.
- *The correlations across multiple outcome variables imputed to the same record are not accounted for when* finding a donor because the imputation is completely univariate. Note that multivariate imputation may not be feasible with the sequential hot deck because each outcome variable may need a different set of covariates.
- *The choice of donor, after the sort has been completed, may be deterministic*, which may introduce bias in estimating means and totals and hence make it rather difficult to determine the mean squared error of the estimator that accounts for imputation.

In addition, because of the increased number of variables requiring imputation, a large number of restrictions had to be placed on the donor set to ensure that imputed values were consistent with preexisting nonmissing values. A complicated hot deck program had been developed that disregarded ineligible donors based on logical restrictions, but expanding this program to account for more variables to be imputed and to incorporate weights in the program would have been quite difficult. Moreover, the number of restrictions that could be placed on the donor set would be limited because of the sparseness of available donors that would result. With more restrictions, more weight would be placed on the first or second covariate in the sort order.

6.3 The New Imputation Method

6.3.1 Introduction of Method and Its Development from Predictive Mean Matching and Nearest Neighbor Hot Deck

To address the deficiencies associated with the unweighted sequential hot deck, a new method of imputation called Predictive Mean Neighborhoods was developed for the NHSDA. PMN is a combination of two commonly used imputation methods: a non-model-based hot deck (nearest neighbor), and a modification of the model-assisted predictive mean matching method of Rubin (1986). PMN enhances the predictive mean matching (PMM) method in that it can be applied to both discrete and continuous variables either individually or jointly. PMN also enhances the nearest neighbor hot deck (NNHD) method in that the distance function used to find neighbors is no longer ad hoc.

The random NNHD, like the sequential hot deck, is a commonly used imputation method (Little & Rubin, 1987, p. 65). With this method, donors and recipients are distinguished by the completeness of their records with regard to the variable(s) of interest (the donor has complete data, the recipient does not). A donor set deemed close to the recipient with respect to a number of covariates is used to select a donor at random. For the NHSDA, the set of covariates typically would include demographic variables as well as some other nonmissing substance use variables. To further ensure that a donor matches the recipient as closely as possible, discrete variables (or discrete categories of continuous variables) strongly correlated with substance use, such as age categories, can be used to restrict the set of donors. Furthermore, other restrictions involving outcome variables can be imposed on the neighborhood. Note that in NNHD, unlike the sequential hot deck, a distance function is used to define closeness between the recipient and a donor. Accordingly, there is less of a problem with sparseness of the donor class, but the distance function involving categorical or nominal variables is typically ad hoc and often hard to justify.

The PMM method is suitable for continuous outcome variables. With this method, a distance function is used to determine distances between the predictive mean for the recipient, obtained under a model, and the response variable outcomes for candidate donors. The respondent with the smallest distance is chosen as the donor. Unlike the NNHD, the donor is not randomly selected from a neighborhood. The advantages of PMM include the following:

- Model bias in the predictive mean can be minimized by using suitable covariates.
- The PMM method is not a pure model-based method because the predictive mean is only used to assist in finding a donor. Hence, like NNHD, it has the flexibility of imposing certain constraints on the set of donors.

However, the choice of donor is nonrandom. This nonrandomness leads to bias in the estimators of means and totals. It also tends to make the distribution of outcome values skewed to the center. Furthermore, as mentioned earlier, the PMM method is not suitable for discrete variables because the distance function between the recipient's predictive mean (which takes continuous values) and the donor's outcome value (which takes discrete values) is not very sensitive in discriminating between donors with common outcome values.

6.3.2 Brief Overview of Predictive Mean Neighborhoods Method

In the univariate version of PMN (denoted as UPMN), as in the case of PMM, the prediction model is fit to the data from complete respondents. The innovative aspect of UPMN is that predictive means for both the recipient and the donor are computed to obtain the distance between the two predictive means. With this innovation, the distance is also well defined for discrete variables, and a delta neighborhood (delta chosen to be a small positive number signifying that donors in the neighborhood are within delta distance from the recipient, i.e., their predictive means are almost equal to that of the recipient) is defined to pick one donor at random similar to the unweighted hot deck. In the multivariate version of PMN (denoted as MPMN), to avoid the problem of multivariate modeling requiring complex covariance structure in general, another innovation was invoked in which several univariate predictive means conditional on the outcome variables assigned earlier on in a hierarchy of outcome variables are obtained, and then a neighborhood for picking a donor is defined using a vector delta, the Mahalanobis distance (see ***Section 6.4*** for definition), or both. For example, with the lifetime substance use variable, the logistic regression model can be used to find the predictive mean for the univariate PMN. With the substance recency of use among lifetime users, the recency of substance use vector (the first element indicating lifetime use but not past year use, the second element indicating past year use but not past month, and the third element indicating past month use) can be imputed by the MPMN. To compute the predictive mean vector, three univariate logistic models corresponding to the three elements of the recency vector are fitted hierarchically so that they are conditioned on the recency outcomes positioned higher up in the vector. In the process, UPMN is used to obtain provisional imputed values that are used in computing predictive means for recipients (see ***Section 6.4*** for more details).

6.3.3 Univariate and Multivariate Applications of Predictive Mean Neighborhoods

PMN is easily applicable to problems of both univariate and multivariate imputations. The need for univariate imputation arises when values of a single continuous variable, such as age at first use of marijuana, or a dichotomous discrete variable, such as lifetime use of marijuana, are missing for a respondent, and need to be imputed independently of other variables. The need for multivariate imputation arises when values of two or more variables are missing for a single respondent. The case of a single polytomous variable such as marijuana recency with missing values can be viewed as a multivariate imputation problem.

The standard approach to multivariate modeling with a given set of outcome variables (including both discrete and continuous) is likely to be tedious in practice because of the (1) computational problems that are due to the sheer number of model parameters and (2) the difficulty in specifying a suitable covariance structure. Following Little and Rubin's (1987) proposal of a joint model for discrete and continuous variables, and its implementation by Schafer (1997), it is possible to fit a pure multivariate model for multivariate imputation, but it would require making distributional assumptions. Moreover, none of the existing solutions take the survey design into account because of the obvious problem of specifying the probability distribution underlying survey data. However, in applying MPMN to the 1999 NHSDA, a multivariate model was fitted by a series of univariate parametric models (including the polytomous case) so that variables modeled earlier on in the hierarchy have a chance to be included in the covariate set for subsequent models in the hierarchy. In the multivariate modeling with MPMN, the innovative idea is to express the likelihood in the superpopulation model as a product of marginal and conditional likelihoods, which then allows for use of univariate techniques for fitting multivariate (but conditional) predictive means.

If it turns out that a donor set for MPMN is sparse, the UPMN can be used as an alternative. Assuming that the donor set (i.e., the set of complete records in a small neighborhood of the recipient with respect to all the elements of the predictive mean) is not sparse, having a single record to fill in all the missing values in an incomplete record is desirable because doing so preserves the relationships among the variables of interest. Moreover, if the predictive mean vector includes both missing and nonmissing variables (this could easily happen when models are fitted in a univariate manner under a hierarchy), one can also ensure that the predictive mean vector for the donor record is close to the recipient not only with respect to missing variables but also with respect to the nonmissing ones. Although the nonmissing values would not be replaced by the "corresponding values from the donor, some degree of correlation between missing and nonmissing variables is expected to be preserved because of the closeness between the donor and the recipient. This is because the predictive mean vector consists of conditional means (the substance use covariates in the conditioning set appear earlier on in the hierarchy); therefore, being close to the conditional means should help preserve the correlation among outcome variables on the recipient record.

6.4 Detailed Description of PMN Method and Application to 1999 NHSDA

The procedure for implementing UPMN and MPMN can be summarized in the following six steps. Steps 2 through 5, and sometimes step 6, are cycled through each of the substances and substance use measures in the order determined by Step 1. Steps 4 and 5 (Steps 4 through 6 when applicable) could be considered a variant of a random nearest neighbor hot deck.

Step 1. *Hierarchy definition*. In the first step, the order in which variables are modeled is determined so that variables early in the hierarchy may be used for modeling the conditional predictive mean, i.e., they have the potential to be part of the set of covariates for variables later in the hierarchy. Note that not all variables in the hierarchy may be missing for a particular incomplete record. Nevertheless, models are developed for all the variables in a univariate fashion for reasons mentioned in ***Section 6.3.2***. For example, in the substance modules in the CAI sample of the 1999 NHSDA, there are different substances to be modeled and there are different measures of substance use for each substance. It is therefore necessary to determine the order in which the combination of substances and substance use measures are to be handled. Using the sequence of variables determined by this step, cycle through Steps 2 through 5 and sometimes Step 6. In the application of the PMN to the NHSDA, the order of imputation for substances was determined by considering the factors such as the level of stigma associated with the substances, the level of missingness in the data, and the degree to which one set of substances could be used as predictors for other substances. The order of substances was given by cigarettes, smokeless tobacco, cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine, crack, and heroin. The order of substance use measures imputed was determined based on the natural hierarchy of the variables: lifetime usage, recency of use, frequency of use in the past 12 months, frequency of use in the past 30 days, and age of first use.

The following steps are conducted for each variable:

Step 2. *Setup for model building and hot deck assignment*. For each model that is fitted, two groups must be created: complete and incomplete data respondents (item respondents and item nonrespondents). Complete data respondents have complete data across the variables of interest, and incomplete data respondents encompass the remainder of respondents. If the final assignment is multivariate, then complete data respondents must have complete data across all the variables in the multivariate response vector. Models are constructed using complete data respondents only.

Step 3. *Sequential hierarchical modeling*. Build the model using the complete data respondents only, with weights adjusted for item nonresponse. For the CAI substance modules, lifetime usage indicators are modeled first, since all other substance use indicators depend on an indication of lifetime use or nonuse. Once the hierarchy of substances for lifetime usage has been determined, lifetime usage indicators for individual substances can be modeled sequentially. The sequence used for the remaining combinations of substances and substance use measures depends on which covariates are desired in the models and which variables are considered part of a multivariate set. Since drug use is so closely related to age, separate models can be fit within predetermined age groups. In the 1999 NHSDA, drug use is modeled separately within the following age groups: 12 to 17 years old, 18 to 25 years old, and 26 years old or over. To take advantage of the fact that a 50-state design is used, covariates are included in the model that account for the level of drug use of the respondent's state of residence.

Step 4. *Computation of predictive means and delta neighborhoods*. Once the model has been fitted, the predictive means for item respondents and item nonrespondents are calculated using the model coefficients. For models with a multivariate predictive mean vector (such as with a polytomous logit model), a single element out of that vector must be chosen, so that each respondent has exactly one predictive mean value.[1] This predictive mean is the matching variable in a random NNHD. It can either come directly from the model, be adjusted to account for the conditioning on the time period, or (if it is the predicted value based on a model with a transformed response variable), be back-transformed to the original units.

For each item nonrespondent, a distance is calculated between the predictive mean of the item nonrespondent and the predictive means of every item respondent. Those item respondents whose predictive means are "close" to (within a predetermined value delta of) the item nonrespondent are considered part of the "delta neighborhood" for the item nonrespondent and are therefore potential donors. If the number of item respondents who qualify as donors is greater than some number, such as *k*, then only those item respondents with the smallest *k* distances are eligible to be donors.

The pool of donors is further restricted to satisfy constraints to make imputed values consistent with the preexisting nonmissing values of the item nonrespondent. An example of this type of constraint, called a logical constraint, is given by the age of first crack use, which must not be younger than the age of first cocaine use. Other constraints, called likeness constraints, are placed upon the pool of donors to make the attributes of the neighborhood as close to that of the recipient as possible. For example, for age at first use, the age of the donor and the age of the recipient are restricted to be the same whenever possible, and the donor and recipient must come from states with similar usage patterns. A small value of delta could also be thought of as a likeness constraint. Whenever insufficient donors are available to meet the likeness constraints, including the preset small value of delta, the constraints are loosened in priority order according to their perceived importance. As a last resort, if an insufficient number of donors are available to meet the logical constraints (given the loosest set of likeness constraints allowable), then a donor is found using a sequential hot

---

[1]Alternatively, one could perform a provisional MPMN simply using the predicted probabilities from the polytomous model. The final MPMN would be built based on probabilities from the polytomous model, as well as predictive means for the other variables in the multivariate set. See Step 6 for a description of the MPMN.

deck, in which matching is done on the predictive mean. (Even though weights would not be used to determine the donor in the sequential hot deck, "unweighted" is not an accurate characterization of the imputation process, since weighting has already been incorporated in the calculation of the predicted mean.)

If a large number of variables is imputed in a single multivariate imputation, one has the advantage of preserving, as much as possible, correlations between variables that exist in the data. However, the more that variables are included in a multivariate set, the less likely it is that a neighborhood can be used for the imputation within a given delta. What is gained by doing a multivariate imputation is lost, in many instances, by not being able to find a neighborhood within the specified delta.

Step 5. *Assignment of imputed values using a univariate predictive mean (UPMN)*. Using a simple random draw from the neighborhood developed in Step 4, a donor is chosen for each item nonrespondent. If only one response variable is to be imputed, then the assignment step is just a simple replacement of a missing value by the value of the donor. It is possible, however, that a quantity is donated that is a function of the final imputed value. For example, for the 12-month frequency of substance use, since donors and recipients could potentially have a different maximum possible number of days in the year that they could have used a substance, the observed proportion of total period is donated, rather than the observed 12-month frequency, where the "total period" could range up to a year. In the assignment step, the donor's proportion of total period is multiplied by the recipient's maximum possible number of days in the year that they could have used the substance.

The assignment step is multivariate if several response variables are associated with a single predictive mean, provided more than one of those response variables is missing. In that case, all of the missing values will be imputed using the same donor. If there is more than one response variable associated with a single predictive mean, but not all of them are missing, only the missing values are replaced by those of the donor. The resulting imputed values are provisional if a multivariate neighborhood (MPMN) step is called for; otherwise, these values are final.[2] (Go to step 6 if the variables for which steps 2-5 have been completed are part of a complete

[2]If the variable is part of a multivariate set to which MPMN is to be applied, and provisional values are not needed for subsequent models, part of Step 4 (creation of delta neighborhood) and Step 5 could be skipped.

multivariate set that MPMN is to be applied to. Go to step 2 if the variables for which steps 2-5 are completed are not part of a complete multivariate set, and other variables are still to be imputed. Otherwise, the process is finished.)

Step 6. *Determination of multivariate predictive mean neighborhood (MPMN) and assignment of imputed values*. With MPMN, the neighborhood definition is based on a vector of predictive means, rather than a single predictive mean as in the univariate case. This vector may encompass a subvector of predictive means from a single categorical model (as with a polytomous logit model) in addition to scalar predictive means from any number of models with continuous response variables. For each item nonrespondent, a distance is calculated between the elements of this vector of predictive means where the observed values are missing, and the corresponding elements of the vector for every item respondent. To make all elements of the vector conditional on the same usage status in the full predictive mean vector, predictive means that are calculated on the basis of past year and past month users are adjusted to account for the probability that a respondent is a past year user or a past month user. For example, in the CAI sample of the 1999 NHSDA, the full predictive mean vector for alcohol includes the following elements:

1. Recency, past month: P(past month alcohol user|lifetime alcohol user)
2. Recency, past year, not past month: P(past year but not past month alcohol user lifetime|alcohol user)
3. 12-month frequency: (Proportion of days respondent used alcohol in the past year|past year user of alcohol)*P(past year user alcohol|lifetime alcohol user)[3]
4. 30-day frequency: (Proportion of days respondent used alcohol in the past month|past month user of alcohol)*P(past month alcohol user|lifetime alcohol user)
5. 30-day binge frequency: (Proportion of days respondent was a binge drinker in the past month|past month user)*P(past month alcohol user|lifetime alcohol user).

[3]For the 12-month frequency, 30-day frequency, and 30-day binge frequency, the models are fit using logits. In 1999, these logits are not converted to probabilities when creating the predictive mean vector. This is an oversight that is corrected in 2000.

The subset of elements that are used to determine a neighborhood for a particular item nonrespondent depends on the missingness pattern of that item nonrespondent.[4] Moreover, if partial information is available on the recency of use, the predictive means will be adjusted to account for that knowledge. For example, if a particular item nonrespondent was known to be a past year alcohol user and his 12-month frequency was known, then the elements above for which differences would be calculated would be element No. 1 conditioned on past year use, and Nos. 4 and 5, that is, P(past month alcohol user|lifetime alcohol user)/P(past year alcohol user|lifetime alcohol user), P(R used alcohol on a given day in the past month|past month user of alcohol))*P(past month alcohol user|lifetime alcohol user)/P(past year alcohol user|lifetime alcohol user), and P(R was a binge drinker on a given day in the past month|past month user)*P(past month alcohol user|lifetime alcohol user)/P(past year alcohol user|lifetime alcohol user).[5]

A neighborhood that results from this vector of distances can be constrained by a multivariate preset delta, where the distances associated with each element of the predictive mean vector must each be less than preset delta associated with that element.[6] From the donors that remain, a single neighborhood can be created out of a vector of differences by converting that vector to a scalar, called the Mahalanobis distance, which is given by

$$(\mu_R - \mu_{NR})^T \Sigma^{-1} (\mu_R - \mu_{NR}) \ ,$$

where

μ_R = predictive mean (sub)vector for a given item respondent

μ_{NR} = predictive mean (sub)vector for a given item nonrespondent.

---

[4]Alternatively, one could use the entire predictive mean vector to determine the neighborhood, regardless of the missingness pattern. Because many respondents in the multivariate set are only missing one item in the set, imputation could be accomplished using UPMN, which is computationally much faster. That is why the entire predictive mean vector is not used to determine the neighborhood in the 1999 imputation process.

[5]The recency of use probability is adjusted based on partial knowledge of the item nonrespondent's recency of use. This knowledge is not used in the adjustment of the frequency of use variables. Even though it is known that the item nonrespondent had more recent use, the predicted means are still adjusted using the probability conditioned on lifetime use rather than more recent use. This is an oversight in the implementation of the 1999 procedures that is rectified in 2000.

[6]In the 1999 imputation procedures, no multivariate delta is set. Instead, the neighborhood is constrained by excluding potential donors who are not part of the univariate neighborhood associated with one of the elements of the multivariate vector.

The matrix Σ is the variance-covariance matrix of the predictive means, calculated using the subvector of predictive means associated with each missingness pattern, using complete data respondents within each age group. The Mahalanobis distance is calculated only for those respondents who meet the delta constraint. The neighborhood is determined by selecting the k smallest Mahalanobis distances within this subset of item respondents for a given item nonrespondent.

If some of the variables in the response vector are not missing, then only those that are missing are replaced. However, logical constraints must be placed on the multivariate neighborhood so that imputed values are consistent with preexisting nonmissing values. For example, if a respondent is missing a 30-day frequency, but his/her non-missing 12-month frequency is 350, then a donor cannot have a 30-day frequency less than 350-335, or 15. If the number of respondents in the univariate subset who meet the logical constraints imposed upon the multivariate neighborhood is fewer than k but greater than 0, then all the respondents in the resulting subset are selected for the neighborhood. Finally, if there are no respondents within the univariate subset who meet the logical constraints imposed by the multivariate neighborhood, then the k smallest Mahalanobis distances who meet the logical constraints among all candidate donors for a given item nonrespondent are selected for the neighborhood. In addition to the multivariate delta, likeness constraints are used to make the donors in the neighborhood as much like the recipient as possible. These can be loosened if insufficient donors are available. Finally, as with the univariate neighborhood, an unweighted sequential hot deck is used as a last resort if insufficient donors are available who meet the logical constraints and the loosest allowable set of likeness constraints.

As with the univariate assignments, a donor is randomly drawn from the neighborhood for each item nonrespondent. For most variables, the observed value of interest is donated directly to the recipient. As in the univariate case, however, it is possible for a quantity to be donated that is a function of the final imputed value rather than the imputed value itself. The 12-month frequency example given in Step 5 applies here as well.

Implementation. The PMN methodology has been widely used for the imputation of a variety of variables in the NHSDA, including both continuous and categorical variables with one or more levels. The models were fit using standard modeling procedures in SAS and SUDAAN®, while SAS macros were used to implement the hot deck step, including the restrictions on the neighborhoods. Although creating a different neighborhood for each item nonrespondent was computationally intensive, the method was implemented successfully.

6.5 Comparison of PMN with Other Available Methods

The PMN methodology addresses all of the shortcomings of the unweighted sequential hot deck method used for earlier NHSDAs.

- *The ability to use covariates to determine donors is far greater than that in the hot deck***.** As with other model-based techniques, using models allows us to incorporate more covariates, including measures of use of other substances, in a systematic fashion, where weights can be incorporated without difficulty. Moreover, those covariates not explicitly modeled can be used as restrictions in choosing the neighborhood. This is possible because the predictive mean is used only to find a neighborhood and not directly for imputation.
- *The relative importance of covariates is determined by standard estimating equation techniques***.** In other words, there are objective criteria based on methodology (such as regression) that quantify the relationship between a given covariate and the response or outcome variable in the presence of other covariates. Thus, the response variable itself is used indirectly to determine donors.
- *The problem of sparse neighborhoods is considerably reduced, which makes it easier to implement restrictions on the donor set.* Because the distance function was defined as a continuous function of the predictive mean, it was possible to find donors arbitrarily close to the recipient. Thus, it was much less likely to have the problem of sparse neighborhoods for hot decking. Moreover, having sufficient donors in the neighborhood allowed for the imposition of extra constraints on the donor set that would have been difficult to incorporate directly in the model.
- *The correlations across response variables is accounted for by making the imputation multivariate.*
- *The choice of donor can be made random by choosing delta large enough so that the neighborhood is of a size greater than 1*. We need delta neighborhoods and random selection of the donor because this allows one to mimic the case that all donors have approximately equal means and the error distribution has mean zero. This helps to avoid

bias in estimating means and totals, variances of which can be estimated as in two-phase sampling or by suitable resampling methods.

- *Sampling weights are easily incorporated in the models.* Note that sampling weights should be used in estimating parameters of the mean function of the imputation model to ensure their design-model consistency. However, for ensuring asymptotic unbiasedness and the consistency of estimates of means and totals, sampling weights are not necessarily needed while choosing a donor at random. It is sufficient to ensure that the selection probabilities are such that the error distribution has mean zero. The existing method of weighted hot deck imputation is an important example in which the error distribution has mean zero around the recipient's predictive mean because the predictive mean is given by the weighted mean of the imputation class. Thus, the weighted hot deck method can be viewed as a special case of PMN.

In comparison to other model-based methods, discrete and continuous variables can be handled jointly and relatively easily in MPMN by using the idea of univariate (conditional) modeling in a hierarchical manner. In MPMN, one can objectively assign differential weights to different elements of the predictive mean vector depending on the variability of predictive means in the data set via the Mahalanobis distance.

As noted earlier, the PMN method has some similarity with the predictive mean matching method of Rubin (1986) except that, for the donor records, Rubin uses the observed variable value (not the predictive mean) to compute the distance function. Also, the well-known method of nearest neighbor imputation is similar to PMN, except that the distance function is in terms of the original predictor variables and often requires somewhat arbitrary scaling of discrete variables. Moreover, for this method, it is generally hard to decide objectively about the relative weights for different predictor variables.

6.6 Impact of Imputation on Substance Use Prevalence Estimates

Tables 6.1 *through* ***6.3*** give the relative impact of imputation for 1998 and 1999 (CAI) estimates of past month use, past year use, and lifetime use of all substances in the core section of the questionnaire. As a way to explain the data presented in these tables, suppose the following:

E = Edited weighted total number of users,

I = Increase in the final weighted total due to imputation,

P = Final estimated, weighted total population, and

T = E + I = Final estimated, weighted total number of substance users for any particular substance and time period.

The table columns present the following:

Final % = 100*(T/P)

Relative % from Imputes = 100*(I/T).

Because of numerous changes between the 1998 sample and the 1999 CAI sample (discussed in ***Chapter 2)***, it would not be advisable to compare the final prevalence estimates (Final Percent) between the two samples. However, some comments can be made about the comparison of the "Relative Percent from imputes" between the two samples. In general, imputation has greater impact on the prevalence estimates in the CAI sample than on the estimates in the PAPI sample. With the implementation of the "flag-and-impute" editing rule in the CAI sample, where inconsistencies would be resolved by imputation, this result is not surprising. The exceptions to this rule are either due to differences in questionnaire format between PAPI and CAI or to attributes of the modules themselves. In particular,

- In the PAPI sample, no frequency of use questions were available for smokeless tobacco or cigars (age at first use was not used to edit to more recent use). Hence, respondents could only indicate past month or past year use directly in the recency of use question. There were no opportunities to edit to more recent use based on responses to other questions, which resulted in a potentially larger role for statistical imputation in the PAPI sample. Moreover, for cigars, the PAPI questions were asked in the noncore section, and no questions were asked about past year use.
- For estimates of the prevalence of nonmedical use of prescription-type substances (pain relievers, tranquilizers, stimulants, and sedatives) in the past month, the impact due to imputation is much greater in the PAPI sample than it is in the CAI sample. In both CAI and PAPI, these modules differ from the others in that they do not include a 30-day frequency question. The CAI and PAPI modules for these substances do include questions about frequency

Table 6.1 Impact of Imputation on Past Month Use of Selected Substances, by Age Group

Substance	1998 PAPI Final %	1998 PAPI Relative % from Imputes	1999 CAI Final %	1999 CAI Relative % from Imputes
Aged 12-17				
Cigarettes	18.25	0.00	14.90	0.54
Smokeless tobacco	1.21	0.00	2.29	0.75
Cigars	5.64	5.51	5.36	1.53
Alcohol	19.08	0.00	16.54	5.20
Binge alcohol	7.70	—†	10.10	11.51
Marijuana	8.26	0.00	7.22	6.49
Cocaine	0.82	0.00	0.53	8.34
Crack	0.25	0.00	0.08	5.07
Heroin	0.16	0.00	0.17	19.67
Hallucinogens	1.82	0.00	1.10	7.34
Inhalants	1.11	0.00	1.05	11.63
Pain relievers	1.32	20.98	2.14	12.86
Tranquilizers	0.27	34.80	0.51	11.23
Stimulants	0.57	39.20	0.65	17.79
Sedatives	0.20	46.05	0.19	14.74
Aged 18 to 25				
Cigarettes	41.60	0.00	39.67	0.22
Smokeless tobacco	5.42	0.00	5.74	0.37
Cigars	11.86	3.27	11.50	0.55
Alcohol	60.02	0.00	57.24	1.86
Binge alcohol	31.72	—†	37.90	5.29
Marijuana	13.78	0.00	14.22	2.71
Cocaine	1.96	0.00	1.67	3.37
Crack	0.32	0.00	0.26	6.04
Heroin	0.21	0.00	0.13	2.30
Hallucinogens	2.65	0.00	1.87	2.54
Inhalants	1.10	0.00	0.64	2.78
Pain relievers	1.77	24.37	2.61	6.85
Tranquilizers	1.01	20.72	1.06	7.47
Stimulants	0.59	24.06	1.05	7.99
Sedatives	0.21	11.44	0.20	3.38
Aged 26 or Older				
Cigarettes	26.60	0.00	24.91	0.32
Smokeless tobacco	2.94	5.64	3.19	0.49
Cigars	6.29	4.57	4.50	0.79
Alcohol	54.68	0.00	48.73	1.58
Binge alcohol	14.01	—†	18.62	6.83
Marijuana	3.15	0.00	2.79	2.47
Cocaine	0.61	0.00	0.56	12.61
Crack	0.17	2.70	0.19	20.05
Heroin	0.02	0.00	0.05	18.02
Hallucinogens	0.21	0.00	0.08	5.35
Inhalants	0.09	0.00	0.09	9.79
Pain relievers	0.54	37.62	0.82	4.19
Tranquilizers	0.19	37.53	0.40	0.05
Stimulants	0.20	34.71	0.30	6.47
Sedatives	0.06	69.91	0.08	0.00

(Continued)

Table 6.1 (Continued)

	Aged 12 or Older			
	1998 PAPI		1999 CAI	
Substance	**Final %**	**Relative % from Imputes**	**Final %**	**Relative % from Imputes**
Cigarettes	27.65	0.00	25.76	0.31
Smokeless tobacco	3.08	4.14	3.42	0.48
Cigars	6.93	4.36	5.49	0.80
Alcohol	51.66	0.00	46.45	1.76
Binge alcohol	15.63	—†	20.21	6.70
Marijuana	5.04	0.00	4.73	3.21
Cocaine	0.80	0.00	0.70	9.43
Crack	0.20	1.80	0.19	16.92
Heroin	0.06	0.00	0.07	14.78
Hallucinogens	0.69	0.00	0.42	4.27
Inhalants	0.33	0.00	0.26	8.35
Pain relievers	0.78	30.85	1.19	6.59
Tranquilizers	0.30	30.05	0.50	3.31
Stimulants	0.29	32.87	0.43	8.75
Sedatives	0.10	48.50	0.10	3.69

†—Although an edited binge drink variable was available in the 1998 PAPI, no imputation-revised binge drink variable was created. The range of variables in PAPI to which imputation was applied was considerably less than in 1999 CAI. The binge drink variable was one of the variables with missing values imputed for the first time in 1999.

of use in the past 12 months. In CAI, respondents who report that they last used a prescription-type drug more than 12 months ago are not asked about their frequency of use in the past 12 months.

With the PAPI questionnaire, however, respondents could indicate past year use on the 12-month frequency question even if they reported less recent use on the recency question. PAPI editing procedures classified these cases as past year users. Because the PAPI questionnaire did not contain 30-day frequency questions for the prescription-type substances, there was no information for these respondents as to whether they were past month users of these drugs. Imputation was therefore applied to classify them as either past month users or not. For other substances, PAPI respondents could indicate past month usage in the 30-day frequency question, even if they did not indicate it in the recency question. In these situations, PAPI respondents were logically inferred to be past month users, and no further imputation was required.

- For other substances and prevalence measures, a large number of cases required no imputation in the PAPI sample. Because inconsistencies were always resolved in PAPI in favor of more recent use, imputation would only be required where no information was available (e.g., the respondent did not answer the gate question, or the respondent started using at his/her current age but did not answer the 30-day frequency question).

Table 6.2 Impact of Imputation on Past Year Use of Selected Substances, by Age Group

Substance	1998 PAPI Final %	1998 PAPI Relative % from Imputes	1999 CAI Final %	1999 CAI Relative % from Imputes
Aged 12-17				
Cigarettes	23.76	0.12	23.39	2.77
Smokeless tobacco	3.72	0.93	4.60	4.43
Cigars	NA†	NA†	12.56	5.54
Alcohol	31.81	0.00	34.13	4.57
Binge alcohol	—‡	—‡	—‡	—‡
Marijuana	14.06	0.00	14.15	4.62
Cocaine	1.67	0.00	1.61	11.13
Crack	0.45	0.00	0.39	8.72
Heroin	0.25	0.00	0.28	4.55
Hallucinogens	3.82	0.00	3.83	10.30
Inhalants	2.92	0.00	3.91	10.18
Pain relievers	3.07	0.18	5.46	15.45
Tranquilizers	1.09	0.00	1.56	13.04
Stimulants	1.17	0.00	2.12	12.67
Sedatives	0.57	0.00	0.48	15.65
Aged 18 to 25				
Cigarettes	47.11	0.18	47.49	0.54
Smokeless tobacco	8.96	0.82	9.21	1.25
Cigars	NA†	NA†	24.96	2.10
Alcohol	74.16	0.00	74.79	1.02
Binge alcohol	—‡	—‡	—‡	—‡
Marijuana	24.10	0.00	24.46	1.46
Cocaine	4.67	0.00	5.24	2.59
Crack	0.79	0.00	0.95	3.48
Heroin	0.40	0.00	0.45	5.22
Hallucinogens	7.17	0.00	6.77	1.80
Inhalants	3.20	0.00	2.62	2.54
Pain relievers	4.41	0.00	7.59	5.86
Tranquilizers	2.67	0.21	3.14	4.64
Stimulants	1.85	0.30	2.95	3.91
Sedatives	0.49	0.00	0.61	2.08
Aged 26 or Older				
Cigarettes	28.71	0.23	28.14	0.49
Smokeless tobacco	3.71	5.23	3.90	0.40
Cigars	NA†	NA†	9.45	1.63
Alcohol	66.67	0.00	64.05	0.75
Binge alcohol	—‡	—‡	—‡	—‡
Marijuana	5.23	0.00	5.23	1.31
Cocaine	1.27	0.00	1.11	3.06
Crack	0.39	1.21	0.40	1.47
Heroin	0.05	0.00	0.09	0.00
Hallucinogens	0.41	0.00	0.22	4.42
Inhalants	0.27	0.00	0.17	3.77
Pain relievers	1.28	0.00	1.86	3.79
Tranquilizers	0.56	0.00	0.87	1.03
Stimulants	0.42	0.00	0.57	3.35
Sedatives	0.15	0.00	0.20	2.42

(Continued)

Table 6.2 (Continued)

	Aged 12 or Older			
	1998 PAPI		1999 CAI	
Substance	**Final %**	**Relative % from Imputes**	**Final %**	**Relative % from Imputes**
Cigarettes	30.55	0.21	30.14	0.69
Smokeless tobacco	4.39	3.69	4.65	1.03
Cigars	NA†	NA†	11.77	2.19
Alcohol	64.00	0.00	62.29	1.01
Binge alcohol	—‡	—‡	—‡	—‡
Marijuana	8.57	0.00	8.64	1.93
Cocaine	1.74	0.00	1.69	3.68
Crack	0.44	0.81	0.47	2.62
Heroin	0.12	0.00	0.16	2.74
Hallucinogens	1.63	0.00	1.44	4.48
Inhalants	0.92	0.00	0.88	6.30
Pain relievers	1.86	0.03	2.98	6.71
Tranquilizers	0.89	0.08	1.23	3.81
Stimulants	0.68	0.11	1.04	5.56
Sedatives	0.24	0.00	0.29	4.66

†NA = not available. Since the question about cigar recency of use in the 1998 PAPI only asked about use in the past 30 days, no information about past year cigar use was available.

‡—Although an edited binge drink variable was available in the 1998 PAPI, no imputation-revised binge drink variable was created. The range of variables in PAPI to which imputation was applied was considerably less than in 1999 CAI. The binge drink variable was one of the variables with missing values imputed for the first time in 1999.

Table 6.3 Impact of Imputation on Lifetime Use of Selected Substances, by Age Group

Substance	1998 PAPI Final %	1998 PAPI Relative % from Imputes	1999 CAI Final %	1999 CAI Relative % from Imputes
Aged 12-17				
Cigarettes	35.81	0.06	37.07	0.00
Smokeless tobacco	8.85	0.01	9.79	0.03
Cigars	18.64	0.84	19.57	0.08
Alcohol	37.34	0.00	42.87	0.20
Binge alcohol	—†	—†	—†	—†
Marijuana	16.95	0.00	18.73	0.29
Cocaine	2.19	0.00	2.38	0.00
Crack	0.75	0.00	0.62	0.00
Heroin	0.35	0.00	0.44	0.00
Hallucinogens	5.35	0.00	5.67	0.82
Inhalants	6.10	0.00	9.13	0.34
Pain relievers	4.61	0.12	8.16	1.75
Tranquilizers	1.73	0.00	2.47	0.97
Stimulants	1.72	0.00	3.95	1.01
Sedatives	0.98	0.00	0.82	0.71
Aged 18 to 25				
Cigarettes	68.82	0.00	68.91	0.00
Smokeless tobacco	24.07	0.01	25.81	0.03
Cigars	39.31	1.01	43.94	0.09
Alcohol	83.20	0.00	83.91	0.06
Binge alcohol	—†	—†	—†	—†
Marijuana	44.61	0.00	46.81	0.44
Cocaine	10.03	0.00	11.88	0.29
Crack	2.70	0.00	3.29	0.39
Heroin	1.08	0.00	1.75	0.49
Hallucinogens	17.37	0.00	19.29	0.33
Inhalants	10.81	0.16	14.10	0.22
Pain relievers	8.24	0.00	15.16	0.49
Tranquilizers	5.08	0.11	7.87	0.51
Stimulants	3.94	0.14	8.99	0.27
Sedatives	1.25	0.00	1.97	0.88
Aged 26 or Older				
Cigarettes	74.48	0.02	72.29	0.00
Smokeless tobacco	17.24	0.48	19.23	0.10
Cigars	36.00	0.76	36.33	0.06
Alcohol	86.89	0.00	86.13	0.03
Binge alcohol	—†	—†	—†	—†
Marijuana	33.23	0.00	34.67	0.40
Cocaine	11.80	0.00	12.67	0.62
Crack	2.12	0.22	2.85	1.88
Heroin	1.19	0.00	1.45	0.00
Hallucinogens	9.26	0.79	10.77	0.46
Inhalants	4.88	0.00	6.50	0.49
Pain relievers	4.91	0.00	8.07	0.25
Tranquilizers	3.52	0.40	6.52	0.91
Stimulants	4.84	0.13	7.35	0.52
Sedatives	2.42	0.00	4.13	0.77

(Continued)

Table 6.3 (Continued)

	Aged 12 or Older			
	1998 PAPI		1999 CAI	
Substance	Final %	Relative % from Imputes	Final %	Relative % from Imputes
Cigarettes	69.73	0.02	68.16	0.00
Smokeless tobacco	17.24	0.37	19.09	0.08
Cigars	34.62	0.80	35.55	0.06
Alcohol	81.26	0.00	81.31	0.04
Binge alcohol	—†	—†	—†	—†
Marijuana	32.99	0.00	34.56	0.40
Cocaine	10.57	0.00	11.49	0.56
Crack	2.05	0.18	2.67	1.60
Heroin	1.09	0.00	1.38	0.08
Hallucinogens	9.89	0.57	11.33	0.45
Inhalants	5.76	0.04	7.75	0.41
Pain relievers	5.31	0.01	8.99	0.44
Tranquilizers	3.54	0.33	6.27	0.85
Stimulants	4.40	0.13	7.20	0.50
Sedatives	2.12	0.00	3.50	0.78

†—Although an edited binge drink variable was available in the 1998 PAPI, no imputation-revised binge drink variable was created. The range of variables in PAPI to which imputation was applied was considerably less than in 1999 CAI. The binge drink variable was one of the variables with missing values imputed for the first time in 1999.

6.7 References

Little, R.J.A, & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Rubin, D.B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics*, *4*, 87-94.

Schafer, J. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall.

7

Mode Effects on Substance Use Measures: Comparison of 1999 CAI and PAPI Data

James Chromy, Teresa Davis, Lisa Packer, and Joseph Gfroerer

The shift from paper and pencil interviewing (PAPI) to computer assisted interviewing (CAI) in the 1999 National Household Survey on Drug Abuse (NHSDA) was anticipated to have significant effects on the reporting of substance use by survey respondents. This expectation was based on several studies in the literature as well as field testing done within the NHSDA project that showed respondents were more willing to report sensitive behaviors using audio-computer assisted interviewing than with self-administered paper questionnaires. Because there is great interest in analyzing trends in substance use prevalence, a critical component of the 1999 NHSDA redesign was the supplemental sample that employed the "old" (PAPI) NHSDA data collection methodology. The intent of this dual-sample design in 1999 was primarily to make it possible to continue to measure trends, using estimates from both before and after the redesign. However, the large dual sample design of the 1999 NHSDA also can be viewed as an important survey research experiment assessing mode effects on the reporting of sensitive behaviors.

This chapter presents a comparison of the substance use prevalence estimates derived from the 1999 PAPI and CAI samples. In addition to providing NHSDA data users with information that will help them interpret NHSDA trends in substance use prevalence, the analysis also will be of interest to survey researchers concerned with mode effects.

7.1 Background

In 1999, the National Household Survey on Drug Abuse (NHSDA) was implemented using a new 50-state design together with CAI methods for both screening households and interviewing selected respondents. Prior to 1999, PAPI was used both for screening and interviewing.

The introduction of CAI technology was designed to produce more internally consistent data while still allowing the respondent to answer in private by using the audio computer-assisted self interviewing (ACASI) for the more sensitive parts of the interview. The ACASI approach allowed the respondent to enter answers to sensitive questions directly into the computer away from the view of the field interviewer or any other household members. In addition, the questions were displayed on the screen for the respondent to read and a recorded voice reading of the questions was also provided to the respondent via ear phones.

While the main purpose of simultaneously administering the survey using the old PAPI methodology in a supplemental sample was to relate prior PAPI-based estimates to levels that would have been achieved using CAI for long term trend reporting in the future, it also provides an opportunity to estimate mode effects based on large nationally representative samples. If the 1999 NHSDA had been designed as strictly a "mode" experiment, then the number of questionnaire changes that were introduced with CAI might have been limited. The CAI questionnaire does allow the survey designer to ask questions differently, to easily refer back to prior answers, to condition probes on initial responses, to implement complex skip and routing patterns without confusing the respondent, and to allow the respondent to resolve at least some of the apparent inconsistencies in initial answers. The mode effects reported in this chapter incorporate not only the effects of presenting the same or similar questions by a computer instead of a paper questionnaire and responding on a keyboard instead of with a pencil, but they also reflect the total impact of the entire conversion process from PAPI to CAI including the enhancements in the interviewing process made possible by the CAI application.

It is important to note that this is not an evaluation of the effects of self-administered questionnaires (SAQs) vs. interviewer-administered questionnaires (IAQs) for sensitive questions. Use of a self-administered questionnaire for the more sensitive portions of the NHSDA questionnaire has been the standard approach since inception of the survey (Barker et al., 1998). Turner et al. (1992) confirmed the importance of using a self-administered approach based on a 1990 NHSDA field experiment. Several other studies report on the effects of self vs. interviewer administration across modes defined by paper and pencil, telephone, and face-to-face interviewing (e.g., for summaries see Tourangeau and Smith, 1998; Turner et al., 1998b).

Early research on computer-assisted self-interviewing was reported by O′Reilly et al. (1994). They compared the results of alternative self-administered interview modes: (1) paper and pencil self-interview, (2) video (only) computer-assisted self-interview; and (3) audio (with respondent option for video) computer-assisted self-interview (ACASI). Their research conducted in a cognitive laboratory setting with a small sample (about 40 subjects) showed generally higher proportions reporting substance use as one progressed from paper to video (only) to full ACASI. Statistically significant ($p < 0.10$) results were observed for several of the substance use measures studied. Turner et al. (1998a) found significant mode effects when comparing answers to sensitive questions administered through ACASI when compared to paper and pencil self-administered questionnaires. The sensitive questions studied included several on drug abuse.

Aquilino, Wright, and Supple (2000) report on their analysis of 3,169 interviews conducted in 40 self-representing Primary Sample Units (PSU′s) of ISR-Temple′s national sample. Respondents were randomly assigned to either a computerized self-administered questionnaire or a paper-and-pencil self-administered questionnaire. Their questionnaire included questions about the prevalence and recency of cigarette smoking, alcohol use, and use of illicit drugs modeled after the questionnaire used for the 1995 National Household Survey on Drug Abuse. Their research focused on the effects of bystanders. They concluded that, when a parent was present during the interview, participants were less likely to report the use of alcohol or marijuana.[1] They also concluded that computerized self-administered interviewing reduced this effect relative to paper-and-pencil self-administered interviewing.

The planning and testing for a transition to CAI is discussed in an earlier report (SAMHSA, 2001). A major objective of introducing CAI technology was to improve the quality of the data by providing valid substance use reports and by avoiding the inconsistencies that arise naturally in the PAPI approach. Under the prior PAPI approach, the

[1]Gfroerer (1985) reported similar effects of parental presence on the reporting of substance use by youth.

sensitive sections of the interview were completed in private by the respondent. Instead of asking the respondent to follow skip instructions around most non-applicable questions, the respondent was asked to respond to each question but allowed to check a response indicating the question did not apply.[2] The CAI interview was programmed to automatically route the respondent to appropriate sections based on responses to gate questions. In addition, consistency checks were programmed into the interviewing process to detect inconsistent answers and solicit the respondent′s answers to additional questions intended to resolve the inconsistencies. Some of the expected benefits of the CAI approach were:

- More complete responses (fewer missing items)
- More internal consistency among responses to different questions.

Nevertheless, some inconsistencies and some missing item data remained in the CAI data. A study was undertaken to examine the first six months′ data from the 1999 NHSDA to study the effects of alternative editing and imputation strategies. The primary focus was on establishing a "usable" case rule and determining the impacts of alternative editing strategies. Only the traditional "hot deck" imputation procedure was considered in the initial efforts. As a result of this study and further review by NHSDA experts, the "flag and impute" method of editing was adopted for the 1999 CAI data (see ***Chapter 5***). To preserve the potential for measuring trends based on 1999 PAPI comparisons to PAPI data from prior years, no changes to the previous PAPI editing and imputation procedures were instituted for the 1999 PAPI data. As a result, the PAPI and CAI data editing and imputation procedures were quite different.

A usable case for PAPI required sufficient responses to the specific substance answer sheets so that at least lifetime use (yes or no) could be established for marijuana, cocaine, and alcohol. For the PAPI data, editing was used to resolve a high proportion of inconsistencies with some indication of more recent use generally interpreted as a basis for creating consistency by assuming more recent use; this was a continuation of the editing policy developed over time for prior NHSDA surveys (Guess et al., 2000). Editing and/or imputation was required for missing data and for responses that were inconsistent with reported age of first use or frequency of use. Other edits were required for questions with multiple gate questions, open-

---

[2]Some skip patterns were included in the PAPI core section answer sheets for stimulants, sedatives, tranquilizers, and pain relievers and in the non-core answer sheets.

ended responses, or with related substances (*e.g.*, cocaine and crack cocaine). Imputation for recency of use was limited to the remaining cases that could be imputed using the "sequential hot deck" method and conditioning on available known information about broader classes of recency of use inferred from the actual response data. Frequency of use was imputed using a combination of model-based and hot deck procedures (Bowman et al., 1998).

A usable case for CAI required responses to the lifetime use question for cigarettes and at least nine other substances. A new approach for editing and imputation was followed for CAI since a new time series was being established with the redesign in 1999 (see ***Chapters 5 and 6***).

The main point of this discussion of editing and imputation in the context of evaluating mode effects is that the final published statistics reflect not only administration mode but also post survey processing through the editing and imputation processes. The analysis attempts to isolate these effects.

7.2 Sample and Experimental Design

The 50-state design involved minimum sample size requirements for all states adequate to support small area estimation procedures applied at the state level. The small area estimation procedures utilized the sample data as well as state and substate statistics available from other sources. In eight higher population states, larger samples were selected to support estimates based on the sample data alone. Interviewers completed dwelling unit rosters of eligible persons using a handheld computer. The computer randomly assigned the household to either the PAPI or CAI mode and then selected the sample respondents accordingly. Interviewers did not know in advance which mode of interviewing was to be assigned at any given dwelling unit.

The national CAI sample included 66,706 respondents from 7,200 area segments selected from 900 area strata called field interviewer (FI) regions.[3] The PAPI subsample involved a subsample of 250 FI regions. Within each FI region in the PAPI subsample, eight area segments were utilized for both mode samples. Larger samples were selected in the PAPI subsample segments and selected housing units were randomly assigned by

[3]An FI region was assigned two segments per quarter and corresponded to the approximate work load for one field interviewer (FI).

mode to meet target sample sizes. Sample selection probabilities for the PAPI sample were coordinated to achieve oversampling by age (younger people were sampled at higher rates) and by race/ethnicity (Hispanic and black persons were sampled at higher rates). Sample selection probabilities for the CAI sample were coordinated to achieve oversampling by age only; the total state sample requirements were adequate to meet national sample size requirements by race/ethnicity groups (Bowman et al., 2000). The combined design is summarized in ***Table 7.1***.

Table 7.1 The 1999 NHSDA Sample Design by Mode

Units	CAI Main Sample	PAPI Subsample
First level of stratification	States	Design strata similar to those used in prior years
FI regions	900 covering the entire U.S. land area and treated as sampling strata. 12 for small sample states and 48 for large sample states	Subsample of 250 treated as primary sampling units
Area segments	7,200 nationally, 8 per FI region	2,000 nationally, 8 per PAPI FI region
Final sample sizes by mode:		
CAI	66,706 (all segments)	19,482 (PAPI segments only)
PAPI		13,809

Design-based weights were developed for both the PAPI and the CAI samples. Calibration procedures were applied to adjust for nonresponse, to control for known marginal totals, and to control extreme weights (Chen et al., 2001). The final PAPI weights also involved calibration to control for interview experience levels between the 1998 and 1999 PAPI samples in order to improve the estimates of trend (see ***Chapter 8***).

In the following sections, two sets of analyses are presented. The first set examines the official estimates produced from the PAPI and CAI samples and utilizes the entire sample under both modes; note that these analyses incorporate the adjustment of the PAPI weights for FI experience. The second set of analyses limits both samples to the 250 FI regions in which both PAPI and CAI data were collected. Although this second set of analyses utilizes a smaller CAI sample, it takes full advantage of sample matching at both the interviewer and segment levels. FI experience matching occurs naturally based on the sample and experimental design. Therefore, PAPI weights for the second set of analyses are not adjusted for FI experience.

7.3 Full Sample Analysis

7.3.1 Approach to Variance Estimation

In order to take advantage of the positive covariance between PAPI and CAI estimates induced by the use of a common set of 2,000 sample segments for all of the PAPI sample and part of the CAI sample, special design structure variables (variance estimation strata and variance estimation replicates) were required. Since FI regions serve as strata for the CAI sample and as primary sampling units for the PAPI sample, there was no direct way of capturing the covariance and using the entire CAI sample. A heuristic approach was followed in developing a new design structure which could be used to simultaneously analyze both samples. Steps in the process were:

- Base the new structure on the PAPI structure except that strata with three replicates need to have the third replicate randomly reassigned as replicate 1 or 2.
- Within the six broad design strata defined on the original sampling frame used to select the PAPI sample, associate FI regions not selected for PAPI with the variance estimation stratum used for PAPI based on rounding the cumulative size measures.[4]
- For FI regions in the PAPI sample, use the PAPI variance replicate assignment as modified above (two variance estimation replicates per variance estimation stratum).
- For FI regions not in the PAPI sample, maintain the CAI replicate assignment which assigns half of the segments (one per quarter) to each of two replicates.

While this approach to design structure variables for CAI/PAPI mode analysis does not fit either design perfectly, it does capture the total variance and allows for taking advantage of any covariance induced by the matched segment design feature. For CAI estimates, the method essentially reduces the number of denominator degrees of freedom from 900 to 123;

---

[4]Within each of six design strata, cumulative size measures divided by 2 were rounded to obtain an initial approximation of the best final stratum assignment. Additional constraints were programmed to insure that initial FI regions were assigned to stratum 1 even though their cumulative counts rounded to zero and to assign the final cases in a (previously defined) three-replicate strata to the correct PAPI stratum number.

also, since the PAPI FI regions are treated as replicates (rather than strata), the CAI variance estimate inappropriately captures some small increase due to not completely removing the population variance contribution associated with strata. For PAPI estimates, the numerator degrees of freedom are reduced from 127 to 123 as a result of collapsing four strata with three replicates each down to two replicates each; other than some possible additional instability in the variance estimator, little harm is done to the expected variance estimates for the PAPI sample.

7.3.2 Selection of Analysis Variables

For the purpose of evaluating mode effects, 18 lifetime and past year measures and 20 past month measures of substance use (see ***Table 7.2***) were selected for which estimates had been reported in the *1999 Summary of Findings Report* (SAMHSA, 2000) for both the PAPI and the CAI samples. The majority of the measures are based on data from a single questionnaire answer sheet or module. "Any illicit drug" and "Any illicit drug other than marijuana" are composite measures based on subsets of illicit drugs. "Smokeless tobacco" is a composite measure based on reports regarding snuff and chewing tobacco. Finally, there are three sets of "parent-child" type relationships: cocaine with crack; hallucinogens with PCP and LSD; and psychotherapeutics with stimulants, sedatives, tranquilizers, and pain relievers. These parent-child type relations provide additional opportunities for reported inconsistencies and require either programmed internal edits in the CAI instrument or additional editing and imputing after data collection**.** In practice, only a limited amount of programmed respondent resolution to parent-child drug inconsistencies was implemented; e.g., if a respondent reported never using cocaine, the respondent was still allowed to respond to the crack module questions and no attempt was made to resolve the inconsistency between the cocaine and crack responses within the programmed CAI instrument. In general, more opportunities for reported inconsistencies exist within the PAPI instrument for all drugs.

Results are presented for all persons 12 or older and for three age groups: 12 to 17, 18 to 25, and 26 or older. The analysis was conducted based on direct survey estimates of prevalence rates for recency of use. The statistical significance of mode differences was assessed through t-tests using the DESCRIPT procedure in SUDAAN (Shah et al., 1997). Modeling was not used to control for sample differences since the weight calibration methods adjust both PAPI and CAI weights to common sets of marginal distributions based on age, race/ethnicity, and geographic region. Further analysis by other subgroups could be productive in assessing possible interaction of mode effects with additional variables, such as education or employment.

7.3.3 Full Sample Results

The full sample analysis was focused on three comparisons between CAI and PAPI data: (1) substance use prevalence rates before editing and imputation, which provide the best assessment of the "pure" mode effect, (2) substance use prevalence rates after editing and imputation, which provide the best assessment of the overall impact of the NHSDA redesign, and (3) the proportion of cases that required some editing and/or imputation. This last comparison is of interest because it shows how the impact of editing and imputation has changed with the implementation of CAI.

Recency of substance use is measured by two questions. The first question asks the respondent to indicate whether he/she has "ever used" a substance. The second question asks the respondent to identify the most recent use of the substance: e.g., within the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months. The first question alone is sufficient to ascertain lifetime use. In the CAI mode, respondents were routed around remaining questions in the substance module if they answered "no", "don't know", or "refused" to this gate question. In the PAPI mode, respondents were asked all questions on the substance answer sheet (with categories including "never used") and would sometimes answer subsequent questions in such a way that lifetime use could be inferred. For the purposes of assessing mode effects for lifetime use, only the answers to this first question were considered under both modes. Any inference of lifetime use based on answers to subsequent questions was treated as requiring editing so that both CAI and PAPI lifetime use before edits were based strictly on direct responses to the substance gate question (or questions).

For several substances, multiple gate questions were asked to determine lifetime use. For example, lifetime use of hallucinogens was determined by asking a series of questions concerning having "ever used" PCP, LSD, peyote, mescaline, psilocybin, "Ecstasy", or some other hallucinogen not specifically listed. A "yes" answer to any one of the multiple gate questions was sufficient to identify the respondent as a lifetime user; conversely, "no" answers for all the gate questions were required to determine that the respondent was not a lifetime user.

The edited and imputed prevalence rates incorporate the editing and imputation processes (see ***Chapters 5*** and ***6***). The comparison of the "before edit" prevalence rates provides the best assessment of the "pure" mode effect, because it removes the impact of editing and imputation, which is different in CAI and PAPI. Furthermore, the lifetime prevalence comparison provides a better mode effect measure than the past year or past month comparisons. The detailed analyses in this chapter are presented for life time and past month use. Summaries of significant findings include past year measures. The need for editing and imputation is the same for both past year and past month use.

The initial focus is on past month or "current" use as these measures are the most widely used for policy and research purposes. Based on field tests and other research conducted during the transition planning process (SAMHSA, 2001), some increase was expected in the reported prevalence of substance use as a result of switching from the PAPI to the CAI mode. Based on the "before edit" data, this was generally true. It was also expected that the impact of editing would decrease with the switch from PAPI to CAI, and this was most often the case. It was assumed that the higher reported prevalence under CAI would be sustained through the editing and imputing process; in fact, the results were mixed.

Table 7.2 shows the three comparisons discussed above for reported past month use. The "before edits" and "requiring edits" columns include only 14 of the 20 measures since recency of use for the remaining six are either composite measures or are derived from additional questions. All 20 measures are shown in the edited and imputed columns.

The significantly higher reported CAI results "before edits" for inhalants and psychotherapeutics (stimulants, sedatives, tranquilizers, and pain relievers) may be partially due to the specific probing about a larger number of substances when using the CAI instrument. The way that the questions were asked could also cause the higher CAI results. While these are more feasible options under CAI than under PAPI, some of the increase in reporting might have occurred with an expanded PAPI instrument.

The other significant results in ***Table 7.2*** for "before edits" reports are for marijuana and cigarettes for persons aged 12 to 17. The initial question about lifetime cigarettes use was changed from the PAPI version

"Have you ever smoked a cigarette, even one or two puffs?"

to the CAI version

"Have you ever smoked part or all of a cigarette?"

The questions about past month cigarette use were also different between PAPI and CAI. In the PAPI version, the question asked

"How long has it been since you last smoked a cigarette?"

whereas in the CAI version

"During the past 30 days, have you smoked part or all of a cigarette?"

The marijuana questions were more similar in PAPI and CAI, but since the objectives in developing the CAI instrument were only secondarily to provide a mode comparison, some observed differences may be attributable to more than just the mode effect in its simplest sense. Nevertheless, it is noteworthy that, for the unedited past month use comparisons, CAI estimates are significantly higher than PAPI estimates for marijuana, inhalant, and cigarette use for youths aged 12 to 17, while corresponding comparisons for adults show no significant differences.

Table 7.2 also shows the proportion of cases that required some editing or imputation to determine past year or past month use under each mode. Looking only at the significant differences, PAPI generally requires more editing and imputation than CAI, as would be expected. The exceptions to this rule occur with "parent-child" sets of drugs where the CAI instrument has not attempted to reconcile inconsistencies between responses concerning a broad class of drugs (e.g., hallucinogens) and the members of the class (e.g., PCP and LSD). PCP and LSD account for six of the seven instances where the editing requirement for CAI significantly exceeded the editing requirement for PAPI. The other instance for which CAI editing and imputation requirements significantly exceed PAPI requirements is pain relievers for persons aged 12 to 17. The CAI respondent must address lifetime non-medical use of a long list of specific pain relievers and the opportunities for missing responses and errors in specifying "other" categories can lead to extensive editing in spite of the additional opportunities for inconsistency resolution under the CAI mode.

It is interesting to note that some of the initially significant mode differences before editing and imputation are no longer significant following the editing and imputation process. This is true in almost all cases for the individual psychotherapeutic drugs; sedatives for persons aged 18 to 25 and tranquilizers for persons aged 26 and older remain significantly higher under the CAI mode. The initial significant differences for marijuana, inhalants, and cigarettes (persons aged 12 to 17) lose statistical significance

Table 7.2 Percentage Reporting Past Month Use for Selected Illicit Drugs, Alcohol, and Tobacco and Percentages Requiring Edits and/or Imputation: 1999 PAPI vs. CAI Full Sample Analysis

	Aged 12 or Older					
	Percentage				Percentage Requiring Editing and/or Imputation	
	Before Edits		Edited and Imputed			
Drug	**PAPI**	**CAI**	**PAPI**	**CAI**	**PAPI**	**CAI**
Any Illicit Drug			6.95	6.25		
Marijuana and hashish	4.37	4.58	5.37	4.73	2.26[b]	1.04
Cocaine	0.52	0.62	0.83	0.70	0.68	0.55
Crack	0.12	0.16	0.21	0.19	0.24	0.33
Inhalants	0.13[b]	0.24	0.32	0.26	0.98[b]	0.49
Hallucinogens	0.38	0.39	0.73[b]	0.42	1.00	0.70
PCP	0.03	0.02	0.03	0.02	0.16[b]	0.42
LSD	0.22	0.21	0.23	0.22	0.27[b]	0.55
Heroin	0.04	0.06	0.09	0.07	0.13	0.16
Nonmedical use of any psychotherapeutic			1.46	1.79		
Stimulants	0.13[b]	0.39	0.35	0.43	0.87	0.72
Sedatives	0.03[b]	0.10	0.07	0.10	1.26[b]	0.40
Tranquilizers	0.32[a]	0.48	0.43	0.50	1.12[b]	0.48
Pain relievers	0.61[b]	1.11	0.96	1.19	1.40[b]	0.91
Any illicit drug other than marijuana			2.89	2.71		
Alcohol	46.34	45.63	51.97[b]	46.45	7.97[b]	1.70
"Binge" alcohol use			15.07[b]	20.21		
Heavy alcohol use			5.72	5.68		
Cigarettes	24.55	25.68	29.68[b]	25.76	6.41[b]	0.81
Smokeless tobacco			2.16[b]	3.42		
	Aged 12 to 17					
Any Illicit Drug			8.97	9.76		
Marijuana and hashish	5.16[a]	6.75	7.01	7.22	3.15[b]	1.57
Cocaine	0.48	0.47	0.74	0.53	0.42	0.47
Crack	0.23	0.08	0.35	0.08	0.21	0.18
Inhalants	0.45[b]	0.93	0.94	1.05	1.61	1.67
Hallucinogens	0.91	1.00	1.60	1.10	1.44	1.94
PCP	0.24	0.10	0.24	0.11	0.04[b]	1.21
LSD	0.80	0.62	0.80	0.64	0.07[b]	1.40
Heroin	0.14	0.13	0.14	0.17	0.16	0.16
Nonmedical use of any psychotherapeutic			2.44	2.86		
Stimulants	0.40	0.53	0.48	0.65	1.06	1.30
Sedatives	0.20	0.16	0.26	0.19	0.67	0.89
Tranquilizers	0.81	0.46	0.97	0.51	1.10	1.01
Pain relievers	1.37	1.86	1.81	2.14	1.40[b]	2.76
Any illicit drug other than marijuana			4.43	4.50		
Alcohol	14.23	15.68	18.98	16.54	8.12[b]	3.14
"Binge" alcohol use			7.78[b]	10.10		
Heavy alcohol use			3.56	2.41		
Cigarettes	10.96[b]	14.82	15.93	14.90	5.13[b]	1.70
Smokeless tobacco			1.24[b]	2.29		

(Continued)

Table 7.2 (Continued)

Drug	Aged 18 to 25 Percentage: Before Edits PAPI	Before Edits CAI	Edited and Imputed PAPI	Edited and Imputed CAI	Percentage Requiring Editing and/or Imputation PAPI	Percentage Requiring Editing and/or Imputation CAI
Any Illicit Drug			18.80	16.36		
Marijuana and hashish	13.46	13.84	16.40	14.22	5.24[b]	1.35
Cocaine	1.22	1.57	1.90	1.67	1.63[a]	0.77
Crack	0.19	0.24	0.40	0.26	0.39	0.36
Inhalants	0.42	0.62	0.62	0.64	1.78[a]	0.56
Hallucinogens	1.98	1.78	3.49[a]	1.87	3.55[b]	0.93
PCP	0.04	0.10	0.04	0.10	0.67	0.34
LSD	0.97	0.94	0.98	0.97	0.90	0.55
Heroin	0.09	0.12	0.10	0.13	0.36	0.19
Nonmedical use of any psychotherapeutic			3.49	3.73		
Stimulants	0.43[b]	0.95	0.72	1.05	0.88	0.68
Sedatives	0.02[b]	0.19	0.04[b]	0.20	1.02[a]	0.43
Tranquilizers	0.79	0.98	1.37	1.06	1.79[a]	0.65
Pain relievers	1.60[a]	2.43	2.33	2.61	2.70[b]	1.24
Any illicit drug other than marijuana			7.47	6.07		
Alcohol	54.66	56.18	60.16	57.24	8.39[b]	1.83
"Binge" alcohol use			31.10[b]	37.90		
Heavy alcohol use			13.02	13.25		
Cigarettes	37.51	39.58	41.01	39.67	4.62[b]	0.58
Smokeless tobacco			3.74[b]	5.74		
Aged 26 or Older						
Any Illicit Drug			4.68	4.08		
Marijuana and hashish	2.73	2.72	3.29	2.79	1.64[b]	0.92
Cocaine	0.41	0.49	0.66	0.56	0.55	0.53
Crack	0.09	0.15	0.16	0.19	0.22	0.34
Inhalants	0.03	0.08	0.18	0.09	0.76[a]	0.31
Hallucinogens	0.04	0.08	0.14	0.08	0.51	0.49
PCP			0.00	0.00	0.10[b]	0.33
LSD	0.02	0.03	0.02	0.03	0.19[a]	0.43
Heroin	0.02	0.04	0.08	0.05	0.09	0.15
Nonmedical use of any psychotherapeutic			0.98	1.31		
Stimulants	0.05[b]	0.28	0.27	0.30	0.84	0.65
Sedatives	0.01[a]	0.08	0.06	0.08	1.38[b]	0.33
Tranquilizers	0.18[b]	0.40	0.20[a]	0.40	1.01[b]	0.38
Pain relievers	0.34[b]	0.78	0.62	0.82	1.18[b]	0.61
Any illicit drug other than marijuana			1.90	1.89		
Alcohol	49.34	47.96	55.12[b]	48.73	7.88[b]	1.48
"Binge" alcohol use			13.38[b]	18.62		
Heavy alcohol use			4.79	4.85		
Cigarettes	24.23	24.83	29.66[b]	24.91	6.88[b]	0.73
Smokeless tobacco			2.03[b]	3.19		

[a]Difference between PAPI and CAI estimate is statistically significant at the .05 level.
[b]Difference between PAPI and CAI estimate is statistically significant at the .01 level.

after editing and imputation. Past month cigarette use for 12 and older and for 26 and older is significantly higher with PAPI than CAI after editing and imputation, but there is no mode difference for these age groups based on the "before edits" data.

Estimates of hallucinogen use (12 and older and 18 to 25) become significantly higher under the PAPI mode after editing and imputation. Past month alcohol use becomes significantly higher under PAPI for 12 and older and 26 and older.

The derived measures for binge alcohol and smokeless tobacco are significantly higher under the CAI mode in all four age groups. In both cases, this result can probably be attributed to differences in questionnaire structure and data processing in PAPI and CAI. Consistent with procedures in prior years, no imputation was performed for missing or inconsistent responses to the questions for which binge (or heavy) alcohol use is derived for PAPI. CAI-based estimates of binge and heavy alcohol use, however, are imputed. Smokeless tobacco questions are quite different in PAPI and CAI. The PAPI instrument includes only two questions on smokeless tobacco (ever use and recency of use), limiting the potential impact of editing. The CAI instrument, while incorporating skip instructions, includes a detailed set of questions and asks separately about chewing tobacco and snuff.

Table 7.3 shows the three comparisons (estimates before edits/imputation, estimates after edits/imputation, and requirements for editing/imputation) for lifetime use measures. Note that "binge" alcohol use and heavy alcohol use are not shown because they refer only to past month use.

The editing and imputation requirements under the CAI mode were much lower for simply determining lifetime use than for determining past year or past month use. CAI editing and imputation requirements well below 1 percent of all respondents were the rule for most substances; the exceptions were the 12-17 age group under the CAI mode for hallucinogens (1.14 percent) and for pain relievers (1.62 percent). Hallucinogens and pain relievers both have multiple gate questions and open-ended response options with edits that could not be pre-programmed into the CAI instrument.

PAPI editing and imputation requirements were significantly higher than CAI requirements for 43 out of a possible 54 substance-by-age-group measures. In only one case was the CAI editing requirement found to be higher. PAPI editing requirements ranged from less than one percent to close to four percent.

Out of 56 possible substance-by-age-group measures taken before editing and imputation, 23 showed statistically significant and higher prevalence rates under the CAI and only one was statistically significant and higher for the PAPI mode. After editing and imputation, the results were even stronger showing statistically significant and higher results for

CAI in 33 out of a possible 72 substance-by-age-group measures and statistically significant and higher results for PAPI in only two instances.

Similar comparisons were made for past year measures, and the results of the significance testing for past month, past year, and lifetime use are summarized in ***Table 7.4***. No corrections for multiple comparisons have been made, since real interest exists for estimates in each particular substance. By focusing only on effects found to be significant at the single comparison level, unnecessary evaluation of meaningless differences is eliminated. Any correction for multiple comparisons would reduce the set of significant differences even further. The "before edit" statistically significant results show all but one case (lifetime cigarettes for 26 or older) are higher in the CAI mode.

The editing requirements for past month and past year use are more often higher for PAPI than for CAI; the editing requirements for lifetime use are almost always higher for PAPI.

The results for edited and imputed prevalence measures are the most mixed for past month use as discussed above. Eight of the ten significant results showing a higher CAI estimate occur with smokeless tobacco and binge alcohol use, both of which are affected by questionnaire and processing differences. None of the six significant results showing a higher PAPI estimate occur in the 12-17 age group. Past year and lifetime use results remain mixed, but tend to show more statistically significant results with the CAI mode exceeding the PAPI mode. This is especially true for lifetime use measures, which provide the best measure of mode effect based on respondents′ answers alone.

7.4 Matched Segment Sample Analysis

Additional analyses were also conducted based on the CAI and PAPI cases in the PAPI sample of FI regions. This analysis used the full PAPI sample of 13,809 respondents, but only 19,482 of the 66,706 CAI respondents. Restricting the sample for both CAI and PAPI to the PAPI sample of FI regions allowed full advantage to be taken of the positive correlation induced by sampling persons from the same area segments. In addition, the matched segment sample should control for any possible confounding with geographic or demographic group variables. Since the same interviewer was assigned to conduct both the PAPI and CAI interviews, any possible interviewer effects are also held under control by limiting the analysis to the matched segments. Since interviewer effects are

Table 7.3 Percentage Reporting Lifetime Use for Selected Illicit Drugs, Alcohol, and Tobacco and Percentages Requiring Edits and/or Imputation: 1999 PAPI vs. CAI Full Sample Analysis

	Aged 12 or Older					
	Percentage				Percentage Requiring Editing and/or Imputation	
	Before Edits		Edited and Imputed			
Drug	**PAPI**	**CAI**	**PAPI**	**CAI**	**PAPI**	**CAI**
Any Illicit Drug			37.55	39.68		
Marijuana and hashish	34.22	34.43	34.29	34.56	0.63	0.35
Cocaine	9.94[a]	11.43	10.10[a]	11.49	0.79[b]	0.24
Crack	2.04[a]	2.63	2.04[a]	2.67	0.74[b]	0.25
Inhalants	6.08[b]	7.72	6.15[b]	7.75	0.54[a]	0.17
Hallucinogens	10.17	11.28	10.20	11.33	0.61	0.38
PCP	2.42	2.56	3.28[a]	2.57	3.65[b]	0.15
LSD	7.81	8.65	8.12	8.69	2.77[b]	0.19
Heroin	1.13	1.38	1.15	1.38	1.41[b]	0.12
Nonmedical use of any psychotherapeutic			10.38[b]	15.41		
Stimulants	4.11[b]	7.16	4.48[b]	7.20	1.50[b]	0.29
Sedatives	2.17[b]	3.48	2.46[b]	3.50	2.05[b]	0.32
Tranquilizers	3.61[b]	6.21	4.17[b]	6.27	1.65[b]	0.27
Pain relievers	5.67[b]	8.95	7.00[b]	8.99	1.82[b]	0.36
Any illicit drug other than marijuana			19.89[b]	24.09		
Alcohol	80.70	81.27	81.16	81.31	0.55[b]	0.04
Cigarettes	69.76	68.16	70.28	68.16	0.78[b]	0.00
Smokeless tobacco			15.46[b]	19.09		
	Aged 12 to 17					
Any Illicit Drug			21.97[b]	27.65		
Marijuana and hashish	16.83	18.67	16.86	18.73	0.25	0.30
Cocaine	1.56[b]	2.38	1.74	2.38	0.45[b]	0.09
Crack	0.82	0.62	0.83	0.62	0.45	0.12
Inhalants	5.89[b]	9.10	5.97[b]	9.13	0.42	0.51
Hallucinogens	4.64	5.63	4.73	5.67	0.57[a]	1.14
PCP	0.85	0.93	1.06	0.94	2.06[b]	0.22
LSD	3.04	3.76	3.34	3.78	2.00[b]	0.24
Heroin	0.26	0.44	0.27	0.44	0.82[b]	0.09
Nonmedical use of any psychotherapeutic			6.81[b]	10.88		
Stimulants	1.69[b]	3.91	2.31[b]	3.95	1.85[b]	0.70
Sedatives	0.85	0.81	1.29	0.82	1.52[b]	0.75
Tranquilizers	1.41[a]	2.45	2.15	2.47	1.96[b]	0.72
Pain relievers	4.17[b]	8.02	5.74[b]	8.16	1.90	1.62
Any illicit drug other than marijuana			12.36[b]	18.31		
Alcohol	38.34[b]	42.79	38.83[a]	42.87	0.86[b]	0.18
Cigarettes	33.09[a]	37.07	33.28[a]	37.07	0.49[b]	0.00
Smokeless tobacco			5.76[b]	9.79		

(Continued)

controlled by this design, no adjustment to the initial PAPI weights was required for this analysis. To obtain population level estimates based on the CAI sample in the matched segments, the CAI weights for the 19,482 CAI respondents in these segments were simply multiplied by the PAPI FI region (PSU) weight; no additional calibration or poststratification

Table 7.3 (Continued)

	Aged 18 to 25					
	Percentage				Percentage Requiring Editing and/or Imputation	
	Before Edits		Edited and Imputed			
Drug	**PAPI**	**CAI**	**PAPI**	**CAI**	**PAPI**	**CAI**
Any Illicit Drug			50.72	52.60		
Marijuana and hashish	46.13	46.61	46.29	46.81	0.52	0.33
Cocaine	10.34	11.85	10.84	11.88	1.20[b]	0.17
Crack	2.99	3.27	2.99	3.29	0.89[b]	0.22
Inhalants	10.75[b]	14.07	10.81[b]	14.10	0.67	0.12
Hallucinogens	19.05	19.22	19.05	19.29	0.75	0.29
PCP	2.17	2.40	2.99	2.43	3.23[b]	0.16
LSD	15.45	14.64	15.87	14.70	2.14[b]	0.17
Heroin	1.42	1.74	1.42	1.75	1.12[b]	0.11
Nonmedical use of any psychotherapeutic			14.23[b]	20.92		
Stimulants	3.90[b]	8.97	4.95[b]	8.99	2.41[b]	0.20
Sedatives	1.18	1.95	1.97	1.97	2.55[b]	0.34
Tranquilizers	5.53[b]	7.83	6.55	7.87	2.54[b]	0.30
Pain relievers	8.60[b]	15.09	10.61[b]	15.16	3.10[b]	0.37
Any illicit drug other than marijuana			29.56[a]	33.29		
Alcohol	81.23	83.85	81.85	83.91	0.75	0.05
Cigarettes	66.26	68.91	66.74	68.91	0.74b	0.00
Smokeless tobacco			19.20[a]	25.81		
	Aged 26 or Older					
Any Illicit Drug			37.48	39.15		
Marijuana and hashish	34.60	34.54	34.65	34.67	0.70	0.36
Cocaine	11.02	12.59	11.12	12.67	0.77b	0.27
Crack	2.05[a]	2.80	2.05[a]	2.85	0.75[a]	0.28
Inhalants	5.32	6.46	5.39	6.50	0.53[a]	0.13
Hallucinogens	9.44	10.72	9.47	10.77	0.60	0.29
PCP	2.68	2.81	3.63	2.82	3.94[b]	0.14
LSD	7.18	8.32	7.48	8.35	2.99[b]	0.19
Heroin	1.21	1.45	1.22	1.45	1.54[b]	0.12
Nonmedical use of any psychotherapeutic			10.22[b]	15.10		
Stimulants	4.48[b]	7.31	4.70[b]	7.35	1.30[b]	0.26
Sedatives	2.52[b]	4.10	2.70[b]	4.13	2.04[b]	0.25
Tranquilizers	3.58[b]	6.46	4.05[b]	6.52	1.46[b]	0.20
Pain relievers	5.39[b]	8.05	6.56[a]	8.07	1.59[b]	0.19
Any illicit drug other than marijuana			19.30[b]	23.34		
Alcohol	86.41	86.11	86.84	86.13	0.47[a]	0.02
Cigarettes	75.37[a]	72.29	75.94[a]	72.29	0.82[b]	0.00
Smokeless tobacco			16.16[b]	19.23		

[a]Difference between PAPI and CAI estimate is statistically significant at the .05 level.
[b]Difference between PAPI and CAI estimate is statistically significant at the .01 level.

adjustments were applied to the CAI subsample for these analyses since the primary focus is on mode comparison.

The same set of substance use measures addressed in the full sample analysis was analyzed in the matched sample analysis, but only differences between CAI and PAPI prevalence rates are displayed in ***Table 7.5*** for past month use and ***Table 7.6*** for lifetime use. Positive values in ***Table 7.5*** and ***7.6*** indicate higher prevalence in CAI, while negative values indicate

Table 7.4 Summary of Significant Findings for Full Sample Analysis of Recency of Use: 1999 CAI vs. PAPI

	Number of Significant Differences – CAI (C) vs. PAPI (P)							
	12 or Older		12 to 17		18 to 25		26 or Older	
Items analyzed	C>P	P>C	C>P	P>C	C>P	P>C	C>P	P>C
Rates based on reports before editing/imputation								
Past month	5	0	3	0	3	0	4	0
Past year	12	0	7	0	3	0	4	0
Lifetime	7	0	7	0	4	0	5	1
Rates based on edited/ imputed data								
Past month	2	3	2	0	3	1	3	2
Past year	3	2	6	0	5	0	1	2
Lifetime	10	1	9	0	6	0	8	1
Percent of cases requiring editing/imputation								
Past month/year	2	7	3	3	0	9	2	7
Lifetime	0	12	1	9	0	10	0	12

higher prevalence in PAPI. The analysis of percent requiring edits was not repeated for the matched sample.

The matched sample results prior to editing and imputation are somewhat different than the corresponding full sample results. There are fewer significant mode differences in the matched sample analysis than in the full sample analysis. Of the 56 age-by-drug comparisons for past month use, only 11 show statistically significant differences between CAI and PAPI. These are found primarily in estimates of nonmedical psychotherapeutic use, with CAI estimates higher than PAPI estimates. Among youths aged 12 to 17, the CAI-PAPI differences for unedited past month use are generally smaller in the matched sample analysis than in the full sample analysis, and mode differences are not statistically significant for marijuana or inhalants. The CAI estimate for unedited past month youth cigarette use is significantly higher than the corresponding PAPI estimate. Past year substance use before editing and imputation shows 17 statistically significant differences with higher prevalence in the CAI mode. The strongest results are shown for lifetime use with 22 statistically significant differences with CAI higher than PAPI and two statistically significant differences with PAPI higher than CAI.

After editing and imputation, the majority of statistically significant differences are the result of higher PAPI mode estimates for both past month and past year use; 21 of 29 statistically significant differences show

Table 7.5 Difference in Percentage Reporting Past Month Use for Selected Illicit Drugs, Alcohol, and Tobacco: 1999 PAPI vs. CAI Matched Sample Analysis

	Aged 12 or Older		Aged 12 to 17	
	CAI-PAPI		CAI-PAPI	
Drug	Before Edits	After Edit/ Impute	Before Edits	After Edit/ Impute
Any Illicit Drug		-1.32[a]		-1.39
Marijuana and hashish	0.00	-0.88[a]	0.27	-1.09
Cocaine	0.17	-0.14	0.02	-0.39[a]
Crack	-0.02	-0.16	-0.10	-0.30[a]
Inhalants	0.02	-0.15	0.11	-0.42
Hallucinogens	-0.01	-0.35[b]	-0.36	-1.29[b]
PCP	-0.01	-0.01	-0.28[a]	-0.27[a]
LSD	-0.05	-0.04	-0.50[a]	-0.47
Heroin	-0.03	-0.11[a]	-0.01	0.04
Nonmedical use of any psychotherapeutic		-0.18		-0.08
Stimulants	0.27[b]	-0.02	0.07	0.01
Sedatives	0.13[a]	0.03	-0.08	-0.14
Tranquilizers	-0.02	-0.12	-0.16	-0.37
Pain relievers	0.33[a]	-0.08	0.29	-0.19
Any illicit drug other than marijuana		-0.72[a]		-1.40[a]
Alcohol	-0.85	-5.58[b]	0.07	-3.85[b]
"Binge" alcohol use		4.27[b]		1.39[a]
Heavy alcohol use		-0.52		-1.10[a]
Cigarettes	0.70	-3.89[b]	2.42[b]	-2.81[b]
Smokeless tobacco		0.61[a]		0.98[b]
	Aged 18 to 25		Aged 26 or Older	
Any Illicit Drug		-1.39		-1.31[a]
Marijuana and hashish	0.95	-1.33	-0.21	-0.78
Cocaine	0.40	-0.28	0.16	-0.08
Crack	-0.02	-0.37	-0.01	-0.10
Inhalants	0.22	-0.04	-0.03	-0.14
Hallucinogens	0.12	-1.16[a]	0.01	-0.09
PCP	0.18[a]	0.18[a]	0.00	0.00
LSD	0.18	0.19	-0.02	-0.02
Heroin	-0.02	-0.03	-0.04	-0.14[a]
Nonmedical use of any psychotherapeutic		0.69		-0.35
Stimulants	0.67[b]	0.39	0.23[a]	-0.10
Sedatives	0.18	0.11	0.15[a]	0.04
Tranquilizers	0.28	-0.15	-0.06	-0.08
Pain relievers	1.17[b]	0.53	0.19	-0.18
Any illicit drug other than marijuana		-0.77		-0.63
Alcohol	0.62	-3.72[a]	-1.14	-6.05[b]
"Binge" alcohol use		7.07[b]		4.21[b]
Heavy alcohol use		0.02		-0.52
Cigarettes	2.18	-2.00	0.23	-4.34[b]
Smokeless tobacco		1.69[b]		0.38

[a]Difference estimate is statistically significant at the .05 level.
[b]Difference estimate is statistically significant at the .01 level.

higher PAPI estimates for past month use, and 15 of 23 significant differences show higher PAPI estimates for past year use. These results contrast with the results before editing discussed above. For lifetime measures, the statistically significant differences are higher for CAI in 23 of 27 instances, as was the case with the data before editing. Thus,

Table 7.6 Difference in Percentage Reporting Lifetime Use for Selected Illicit Drugs, Alcohol, and Tobacco: 1999 PAPI vs. CAI Matched Sample Analysis

	Aged 12 or Older		Aged 12 to 17	
	CAI-PAPI		CAI-PAPI	
Drug	**Before Edits**	**After Edit/ Impute**	**Before Edits**	**After Edit/ Impute**
Any Illicit Drug	-0.02	1.13		1.73
Marijuana and hashish	1.40[a]	0.00	-0.60	-0.62
Cocaine	0.04	1.30[a]	0.37	0.14
Crack	1.12[a]	0.10	-0.22	-0.24
Inhalants	0.90	1.12[a]	2.03[b]	1.95[b]
Hallucinogens	0.27	0.94	-0.29	-0.36
PCP	1.01[a]	-0.76[a]	-0.02	-0.24
LSD	0.12	0.69	-0.08	-0.35
Heroin		0.07	-0.01	-0.04
Nonmedical use of any psychotherapeutic	2.97[b]	4.10[b]		1.96[b]
Stimulants	1.09[b]	2.45[b]	2.07[b]	1.15[b]
Sedatives	2.42[b]	0.60	-0.06	-0.86[b]
Tranquilizers	2.81[b]	1.64[b]	0.95[b]	-0.26
Pain relievers		1.25[a]	2.44[b]	0.20
Any illicit drug other than marijuana	-0.47	3.48[b]		2.41[b]
Alcohol	-3.16[b]	-0.77	2.68	2.17
Cigarettes		-3.64[b]	1.21	1.00
Smokeless tobacco		2.12[b]		3.60[b]
	Aged 18 to 25		**Aged 26 or Older**	
Any Illicit Drug		2.10		0.91
Marijuana and hashish	1.04	1.03	-0.08	-0.05
Cocaine	1.37	0.98	1.57[a]	1.54
Crack	-0.40	-0.37	0.15	0.24
Inhalants	2.65[b]	2.69[b]	0.74	0.74
Hallucinogens	1.20	1.24	1.03	1.08
PCP	0.32	-0.48	0.30	-0.87
LSD	0.21	-0.17	1.31[a]	0.98
Heroin	0.47	0.46	0.08	0.03
Nonmedical use of any psychotherapeutic		6.81[b]		3.94[b]
Stimulants	5.41[b]	4.40[b]	2.69[b]	2.30[b]
Sedatives	1.06[b]	0.01	1.26[b]	0.91
Tranquilizers	2.53[b]	1.55	2.61[b]	1.92[b]
Pain relievers	6.41[b]	4.04[b]	2.26[b]	0.93
Any illicit drug other than marijuana		4.24[b]		3.52[b]
Alcohol	1.25	0.93	-1.08	-1.35
Cigarettes	0.32	-0.09	-4.25[b]	-4.77[b]
Smokeless tobacco		4.02[b]		1.62

[a]Difference estimate is statistically significant at the .05 level.
[b]Difference estimate is statistically significant at the .01 level.

comparing the results of the data before editing with the results of the data after editing and imputation, it appears that the editing and imputation cause a shift in the PAPI vs. CAI comparisons for past month and past year measures, but have little impact on comparisons of lifetime measures.

***Table** 7.7* summarizes the significant mode effect findings for past month, past year, and lifetime uses based on the matched sample analysis. After editing and imputation, the majority of significant results for past

Table 7.7 Summary of Significant Findings for Matched Sample Analysis of Recency of Use: 1999 CAI vs. PAPI

	Number of Significant Differences – CAI (C) vs. PAPI (P)							
	12 or Older		12 to 17		18 to 25		26 or Older	
Items analyzed	**C>P**	**P>C**	**C>P**	**P>C**	**C>P**	**P>C**	**C>P**	**P>C**
Rates based on reports before editing / imputation								
Past month	3	0	1	2	3	0	2	0
Past year	4	0	4	0	6	0	3	0
Lifetime	7	1	4	0	5	0	6	1
Rates based on edited / imputed data								
Past month	2	7	2	8	3	2	1	4
Past year	0	7	3	1	5	0	0	7
Lifetime	8	2	5	1	6	0	4	1

month and past year show higher PAPI measures. Of the eight cases in which the past month use estimate is higher in CAI than in PAPI, seven involve binge alcohol use or smokeless tobacco use. In contrast to the full sample analysis which found none of the PAPI estimates significantly higher than CAI estimates for youth past month use, the matched sample analysis found eight past month use measures higher in PAPI than CAI for youth. The FI experience weight calibration process, which is incorporated into the full sample analysis but not the matched sample analysis, tended to reduce the 1999 PAPI prevalence rate as shown in ***Chapter 8***; this may partially explain the difference in results between the full sample and matched sample analysis for past month use. For lifetime use measures, the majority of significant findings in both the full and matched sample analyses show higher CAI measures.

7.5 Conclusions

With its dual-sample design, the 1999 NHSDA provided a large sample for assessing the impact of mode of interview. However, the intention of the 1999 NHSDA design was not to evaluate the mode effect, but to determine the impact of the overall change in method due to the redesign of the survey in 1999, on the time series of substance use statistics. The overall change involved many aspects of the survey design and estimation procedures in addition to mode, such as the sampling plan, the questionnaire, data editing, and imputation. Isolating the "pure" mode effect with these data is difficult and is complicated further by the unexpected impact of interviewer experience on substance use prevalence estimates, as discussed in ***Chapter 8***. Nevertheless, the analyses presented in this chapter provide some important findings concerning mode effects, as well as on the comparability of pre-1999 NHSDA published estimates with estimates from the redesigned NHSDA.

When studying the reporting of sensitive or illegal behaviors, conventional thinking has been that higher reporting is closer to the truth. This evaluation continued that approach recognizing that this may not be true in every case.

The results support previous research that shows higher reporting of sensitive behaviors with ACASI than with self-administered paper answer sheets. A total of 336 comparisons of unedited estimates from PAPI and CAI were made. Of these, 112 indicate significantly higher CAI estimates, while only five indicate significantly higher PAPI estimates. Higher CAI estimates are particularly evident for lifetime prevalences, the measures impacted least by questionnaire structure differences between CAI and PAPI.

The analyses of edited and imputed NHSDA estimates show mixed results for past year and past month use. Lifetime prevalence rates, which are minimally impacted by editing and imputation, are generally higher with CAI than with PAPI. A total of 448 comparisons of edited and imputed estimates from PAPI and CAI were made. Of a total of 62 statistically significant differences in lifetime prevalence, 56 indicate a higher CAI estimate and six indicate a higher PAPI estimate. Results for past year and past month measures show variation across substances and age groups, and are also different in the full sample analysis than in the matched sample analysis. Out of 87 significant differences for past month or past year use, 41 indicate a higher CAI estimate and 46 indicate a higher PAPI estimate. For past month use of alcohol and cigarettes, the substances with the highest prevalence of use, the observed PAPI estimate is higher than the CAI estimate for every age group and for both the full and matched sample analyses, and 11 of the 16 comparisons are statistically significant. Marijuana use estimates also tend to be higher with PAPI, although most of these differences are not statistically significant.

One clear conclusion is that the CAI mode of interviewing leads to more internally consistent and complete data. Under PAPI, the need for editing and imputation to clarify recency of use is larger for most substances and provides a greater opportunity to influence estimates through the editing and imputation process. As noted in ***Chapter 5***, the rules applied to the PAPI survey in 1999 (and prior years) were more aggressive in directly inferring more recent use (e.g., past month rather than past year or lifetime). The more aggressive editing rule along with a higher editing requirement undoubtedly contributed to the higher PAPI mode estimates for past year and past month use after editing and imputation. The editing and imputation process did not have the same impact on lifetime use estimates, which continued to show generally higher prevalence estimates in the CAI mode both before and after the editing and imputation process.

In summary, the CAI methodology produces more complete data with a lower requirement for editing and higher prevalence estimates when treating only unambiguous reports as positive indications of substance use. The final past month and past year prevalence estimates, which reflect not only the mode effect, but all the aspects of the total survey design from item content and presentation through editing, imputation, and weighting, tend to be greater under PAPI for some measures, such as past month use of marijuana, alcohol, and cigarettes, but are greater with CAI for other measures, such as binge alcohol use and smokeless tobacco use. Lifetime substance use measures are more consistently higher in CAI mode both before and after editing and imputation. These significant differences in prevalence rates produced by the old and new NHSDA methodologies demonstrate why the pre-1999 NHSDA estimates should not be compared with the 1999 and later estimates from the redesigned NHSDA in describing long-term trends in substance use. Furthermore, the lack of consistency of results across substances, age groups, and particularly between the full sample and matched sample analyses suggests that the development of adjustments of the pre-1999 estimates to make them "comparable" with 1999 and later estimates is probably not feasible.

7.6 References

Aquilino, W. S., Wright, D.L., & Supple, A.J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use & Misuse*, *35*, 845-867.

Barker, P., Gfroerer, J., Caspar, R., & Lessler, J. (1998). Major design changes in the National Household Survey on Drug Abuse. In *Proceedings of the Survey Research Methods Section*, 732-737. Dallas, TX: American Statistical Association.

Bowman, K., Chen, P., Crump, C.J., Packer, L.E., Wheeless, S.C., & Witt M.B. (1998). *1997 National Household Survey on Drug Abuse: Weighting and imputation procedures*. Research Triangle Park, NC: Research Triangle Institute.

Bowman, K., Chromy, J.R., Odom, D.M., & Penne, M.A. (2000). 1999 National Household Survey on Drug Abuse: Sample design report. In RTI (Ed.), *1999 National Household Survey on Drug Abuse: Methodological resource book*. Volume 1, Section 2. Research Triangle Park, NC: Research Triangle Institute.

Chen, P., Gordek, H., Heller, D.C., Penne, M.A., Singh, A.C., & Westlake, M. (2001). Sampling weight calibration for the 1999 NHSDA. In RTI (Ed.), *1999 National Household Survey on Drug Abuse: Methodological resource book*. Volume 1, Section 3. Research Triangle Park, NC: Research Triangle Institute.

Gfroerer, J. (1985). Underreporting of drug use by youths resulting from lack of privacy in household interviews. In B. Rouse, N. Kozel, & L. Richards, Eds., *Self-report methods of estimating drug use: Meeting current challenges to validity,* 22-30. Washington, DC: National Institute on Drug Abuse.

Guess, L.L., Handley, W., Kroutil, L., Smarrella, D., Suresh, P., & Woodell, C. (2000). *1998 National Household Survey on Drug Abuse: Procedures for editing 1998 NHSDA interview data and specifications for stages of machine-editing of drug variables*. Research Triangle Park, NC: Research Triangle Institute.

O'Reilly, J.M., Hubbard, M.L., Lessler, J.T., Biemer, P.B., & Turner, C.F. (1994). Audio and video computer assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics*, *10*, 197-214.

Substance Abuse and Mental Health Services Administration. (2000). *Summary of findings from the 1999 NHSDA*. Office of Applied Studies, Series H-12 (DHHS Pub. No. SMA 00-3466). Rockville, MD: Author.

Substance Abuse and Mental Health Services Administration. (2001). *Development of computer-assisted interviewing procedures for the NHSDA*. Office of Applied Studies, Methodological Series M-3 (DHHS Pub. No. SMA 01-3514). Rockville, MD: Author.

Shah, B.V., Barnwell, B.G., & Bieler, G.S. (1997). *SUDAAN User's Manual, Release 7.5*. Research Triangle Park, NC: Research Triangle Institute.

Tourangeau, R., & Smith, T.W. (1998). Collecting sensitive information with different modes of data collection. In M. Couper et al., Eds., *Computer assisted survey information collection*, 431-454. New York: John Wiley & Sons.

Turner, C.F., Forsyth, B.H., O'Reilly, J., Cooley, P.C., Smith, T.K., Rogers, S.M., & Miller, H.G. (1998b). Automated self-interviewing and the survey measurement of sensitive behaviors. In M. Couper et al., Eds., *Computer-assisted survey information collection*. New York: John Wiley & Sons.

Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., & Sonenstein, F.L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, *280*, 867-873.

Turner, C.F., Lessler, J.T., & Devore, J.W. (1992). Effects of mode of administration and wording on reporting of drug use. In C.F. Turner, J.T. Lessler, & J.C. Gfroerer, Eds., *Survey measurement of drug use: Methodological studies*, 177-220. Washington, DC: National Institute on Drug Abuse.

8

Impact of Interviewer Experience on Respondent Reports of Substance Use

Arthur Hughes, James Chromy, Katherine Giacoletti, and Dawn Odom

The redesign of the National Household Survey on Drug Abuse (NHSDA) in 1999 included a change in the primary mode of data collection from paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI). In addition, the sample design changed from a Nation-based design to a State-based one, and a supplemental sample was collected using PAPI to measure change between 1999 and earlier years. The overall sample size increased from 25,500 in 1998 to 80,515 in 1999, including the CAI and supplemental PAPI samples. Consequently, it was necessary to hire more interviewers than in previous years, which resulted in a higher proportion of inexperienced interviewers. New interviewing staff turnover was also high in 1999, requiring additional training of newly hired interviewers and contributing to the general inexperience of the interviewing staff for both the CAI and PAPI samples. Early analysis of year-to-year change based on comparable PAPI designs showed unanticipated and unusually large increases in substance use from 1998 to 1999 and indicated, in general, that lack of NHSDA interviewing experience was associated with higher estimates of substance use prevalence. This chapter describes the analysis that was done to understand and explain this relationship.

8.1 Prior Research

Research on the effect of interviewer experience on substance use is not entirely new. Turner, Lessler, and Devore (1992) show that past year marijuana use based on data collected from interviewers with no NHSDA experience was 1.4 times as high as that of experienced interviewers. For past year cocaine use, inexperienced interviewers had rates that were almost twice as high. For alcohol and cigarettes, these ratios were generally about the same. However, this analysis was descriptive and thus did not control on demographic or geographic covariates that may help explain some of these differences.

Another study looked at interviewer experience and found little impact on survey outcome measures. In an effort to gauge the representativeness of opinion polls, Keeter, Miller, Kohut, Groves, and Presser (2000) studied the results of two national telephone surveys that used identical questionnaires but very different levels of effort: a "standard" survey conducted over a 5-day period, which used a sample of adults who were home when the interviewer called; and a "rigorous" survey conducted over an 8-week period, which used random selection from among all adult household members. The examination of a variety of variables showed very few significant differences between the two survey types (e.g., engagement in politics, social and political attitudes, including attitudes toward surveys). One variable that was consistently significant was interviewer experience. A regression analysis found interviewer experience to be the most significant predictor of survey type, where interviewers with experience were more likely to conduct interviews in the rigorous study.

8.2 Initial Investigation

Initial analysis of substance use trends based on 1998 and 1999 PAPI data indicated unexpected increases in lifetime use of any illicit substance use along with other measures such as nonmedical use of any psychotherapeutic and analgesics. To investigate this issue further, response rates, the introduction of new survey procedures at the beginning of the year, weighting, editing, and imputation were examined to see if any procedural changes or errors could have been responsible for the unexpected results. Response rates in 1999 were lower than in previous years, particularly in the first quarter, due to the large number of interviewers with no prior NHSDA interviewing experience. Weighting, editing, and imputation procedures used in the 1999 PAPI were consistent with procedures in 1998 (and recent prior years). Extensive analysis revealed no technical problems associated with any of these factors.

The level of field interviewer experience was also studied as part of this inquiry. An interviewer was classified as having some interviewing experience in a given year if he or she conducted at least one NHSDA interview in the prior year. For some substances, this analysis revealed a progressive decline in reported prevalence rates as interviewer experience increased over the survey year. Interviews conducted by experienced interviewers also showed lower prevalence rates from the outset. See ***Tables 8.1*** and ***8.2*** for estimates of lifetime use of any illicit substance and nonmedical use of any psychotherapeutic by interviewer experience and by interview order, which is a measure of experience level over the course of the survey year (1 = first interview conducted, 100 = 100th interview conducted). Estimated rates of lifetime use of any illicit substance are higher among earlier interviews than among interviews that took place later in the survey year (***Table 8.1***), and rates within a given year and interview order are higher among interviewers with no NHSDA experience. Although the trend is not as strong in 1999 for those with previous experience, NHSDA the same pattern occurs, perhaps partly because most 1999 interviewers had no NHSDA interviewing experience. Yet given the 1999 interviewers with no experience, the prevalence rates earlier in the year are still higher than they were after these interviewers gained experience within the year. Measures of interviewer experience and interview order are explained in greater detail in the next section. Based on these initial findings, more in-depth analysis of the effect of interviewer experience was done and is discussed in the remainder of this chapter.

Table 8.1 Percentage Reporting Lifetime Use of Any Illicit Substance by Survey Year, Interview Order, and Interviewer Experience†

	1998, %			1999 PAPI, %		
Interview Order	**No Experience**	**Some Experience**	**All**	**No Experience**	**Some Experience**	**All**
1-19	40.9	35.5	38.4	39.9	36.3	39.3
20-39	38.7	32.6	35.2	40.3	41.8	40.7
40-59	38.2	33.7	35.2	38.0	37.7	37.9
60-99	39.0	32.1	34.2	37.7	37.8	37.7
100 +	43.2	31.7	34.2	35.7	30.6	33.8
All interviews	40.1	33.1	35.8	38.9	37.1	38.5

†"Experience" refers specifically to NHSDA interviewing experience.

8.3 Interviewer Training and Measures of Experience

All interviewers (with and without NHSDA experience) participated in an extensive training session either at the beginning of the year (for initial hires) or later in the year (for replacement hires). The 1999 training covered all aspects of survey work, including field sampling, administrative procedures, obtaining cooperation, household screening, and the interview

Table 8.2 Percentage Reporting Lifetime Nonmedical Use of Any Psychotherapeutic by Survey Year, Interview Order, and Interviewer Experience†

	1998, %			1999 PAPI, %		
Interview Order	**No Experience**	**Some Experience**	**All**	**No Experience**	**Some Experience**	**All**
1-19	14.6	8.2	11.6	13.3	13.8	13.4
20-39	10.7	7.9	9.1	11.9	10.9	11.7
40-59	8.8	8.4	8.6	12.7	7.2	11.1
60-99	11.2	7.5	8.6	10.6	8.5	10.0
100 +	12.7	5.2	6.9	9.2	6.7	8.2
All interviews	12.2	7.4	9.2	12.0	9.7	11.4

†"Experience" refers specifically to NHSDA interviewing experience.

process itself. All interviewers were sent a home study package and were expected to spend between 24 and 32 hours of self-study prior to attending classroom training. Comparatively more classroom time (7 days) was spent on the CAI interview because it involved new computer-assisted procedures; however, all PAPI interviewers received an additional day of instruction on the unique aspects of paper-and-pencil interviewing. Similarly, all bilingual interviewers received an additional classroom day of training to become acquainted with the Spanish version of the instrument.

Table 8.3 presents counts of both interviewers and completed interviews by interviewer experience for 1998 and 1999. In 1999, some interviewers worked only on CAI cases (these are excluded from ***Table 8.3***), whereas others worked on both PAPI and CAI interviews. Almost as many NHSDA-experienced interviewers (188) participated in the 1999 PAPI survey as were carried over in 1998 (195). In 1999, all PAPI interviewers also conducted an equal or larger number of CAI interviews. As a result, it was necessary to hire many more interviewers to work in areas where both PAPI and CAI interviewing were conducted. Interviewers with NHSDA experience constituted only about 19 percent of total PAPI interviewers in 1999, compared with about 40 percent in 1998.

Generally, experienced interviewers conduct more interviews than interviewers with no previous NHSDA experience because most experienced interviewers start work at the beginning of the survey year and are more likely to remain with the project for the full year. Thus, in both 1998 and 1999, the experienced interviewers conducted more than their proportional share of the PAPI interviews. However, an even more significant difference can be seen in the number of interviews conducted

Table 8.3 PAPI Interviewers and Interviews by Interviewer Experience†

Interviewer Experience	Interviewers				Interviews			
	1998		1999		1998		1999	
	No.	%	No.	%	No.	%	No.	%
None	296	60.29	805	81.07	8,407	32.97	9,526	68.98
Some	195	39.71	188	18.93	17,093	67.03	4,283	31.02
Total	491	100.00	993	100.00	25,500	100.00	13,809	100.00

†"Experience" refers specifically to NHSDA interviewing experience.

by experienced interviewers in 1998 relative to 1999. Only about 31 percent of 1999 PAPI interviews were conducted by experienced interviewers compared with more than 67 percent of interviews conducted in 1998 (see ***Table 8.3***).

Interviewers also gain experience and learn to adapt their approaches as they continue to conduct interviews during a survey year. One measure of current-year experience is based on the ordinal number of each interview in the current year's survey. This measure can be viewed as determining both recent experience and time since last formal training. An ordinal number was assigned to each interview conducted by an interviewer in the current year based on the date (and time, if necessary) of the interview. This was assigned combining both CAI and PAPI cases for interviewers who had both types. More than 38 percent of the 13,809 interviews in the 1999 PAPI sample were conducted during the first quarter, compared with 28 percent in the second quarter and 17 percent in the third and fourth quarters.

The number of interviews in this category constituted about 21 percent of all interviews in 1998 and about 10 percent in 1999 (***Table 8.4***), so regression estimates that take into account interview order can change the estimate of annual change substantially.

Tables 8.4 and ***8.5*** show the unweighted and weighted distribution of interviews by both interviewer experience measures for 1998 and 1999. A higher proportion of 1999 PAPI cases were conducted early in the interviewers' current-year experience, compared with 1998 cases. This is not a result of combining order across CAI and PAPI cases, which would tend to increase the order number; it more likely reflects the generally smaller workload for interviewers in 1999 because of early resignations and shorter periods of employment.

Table 8.4 Distribution of PAPI Interviews by Interview Order and Interviewer Experience†

Interview Order	1998				
	No Experience		Some Experience		All
	No.	%	No.	%	%
1-19	3,049	11.96	3,368	13.21	25.17
20-39	1,928	7.56	3,162	12.40	19.96
40-59	1,213	4.76	2,607	10.22	14.98
60-99	1,217	4.77	3,491	13.69	18.46
100 +	1,000	3.92	4,465	17.51	21.43
Subtotals	8,407	32.97	17,093	67.03	100.00
Total	25,500				
	1999 PAPI				
1-19	4,083	16.01	1,226	4.81	20.82
20-39	2,276	8.93	1,030	4.04	12.96
40-59	1,146	4.49	670	2.63	7.12
60-99	1,249	4.90	737	2.89	7.79
100 +	772	3.03	620	2.43	5.46
Subtotals	9,526	37.36	4,283	16.80	54.15
Total	13,809				

†"Experience" refers specifically to NHSDA interviewing experience.

Table 8.5 Weighted Distribution of PAPI Interviews by Interview Order and Interviewer Experience† (Numbers in Thousands)

Interview Order	1998				
	No Experience		Some Experience		All
	No.	%	No.	%	%
1-19	32,819	15.02	28,663	13.12	28.14
20-39	20,182	9.24	27,765	12.71	21.95
40-59	11,216	5.13	22,728	10.40	15.53
60-99	10,856	5.00	25,683	11.76	16.76
100 +	8,616	3.94	29,912	13.70	17.64
Subtotals	83,689	38.31	134,751	61.69	100.02
Total	218,440				
	1999 PAPI				
1-19	58,427	26.42	12,913	5.83	32.25
20-39	38,886	17.59	13,394	6.06	23.65
40-59	22,859	10.34	9,590	4.34	14.68
60-99	27,332	12.36	10,590	4.79	17.15
100 +	17,084	7.73	10,048	4.54	12.27
Subtotals	164,588	74.43	56,535	25.56	100.00
Total	221,123				

†"Experience" refers specifically to NHSDA interviewing experience.

8.4 Modeling Effects of Interviewer Experience

The general approach taken to assess the impact of interviewer experience on the substance use prevalence estimates used regression models to examine trend effects for substance use prevalence before and after adjustment for interviewer experience. The effects of other covariates of substance use were removed by including them in the model. Covariates used in models in this section included (1) U.S. Bureau of the Census region (Northeast, North Central, South, and West); (2) age categories (12 to 17, 18 to 25, 26 to 34, and 35+); (3) gender; (4) age category by gender interaction; and (5) race/ethnicity (Hispanic, non-Hispanic black, and non-Hispanic non-black).

Initial regression analyses focused on unedited gate question responses. A gate question is generally the first question in each section of the questionnaire where a respondent can indicate use or nonuse of a particular substance. For example, in the marijuana section of the questionnaire, the gate question reads as follows: "Have you *ever*, even once, used marijuana or hashish?" Answers to these gate questions (either yes or no) will give an indication of some level of substance use. Various unedited gate count composites were used to apply linear regression procedures. One of these composites contained 16 items based on the total number of positive responses to the gate questions for cigarettes, smokeless tobacco, cigars, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, LSD, PCP, inhalants, analgesics, tranquilizers, stimulants, and sedatives.

All analyses of the unedited gate question responses used weighted linear regression. RTI's Survey Data Analysis (SUDAAN) software program was employed, and the analysis weights were used for both years. The sample structure was represented using standard NHSDA analysis NEST statements for variance strata and variance replicates. Because a new design was introduced in 1999, no allowance was made for any overlap in the sample structure. The 16-gate composite was used to explore some alternative models (see ***Table 8.6)***. Model 1 presents the weighted estimates by year with no adjustments for demographics or any interview experience. Estimates in the 1998 and 1999 columns represent a weighted average of the number of gate questions answered "yes." Thus, in 1998, and among the gate questions from the 16-item composite mentioned above, a weighted average of 2.877 gates were answered "yes" compared with 2.970 in 1999. The year-to-year change, less than 0.1 of a gate question, was not statistically significant (p-value = .122).

Table 8.6 Trend Effects for 16-Gate Composite under Alternative Models

Model		1998	1999	Change from 1998 to 1999		
				Estimate	SE	p-Value
1	Unadjusted mean	2.877	2.970	0.093	.060	.122
2	Unadjusted mean by age					
2.1	12-17	1.442	1.497	0.055	.063	.381
2.2	18-25	3.345	3.458	0.113	.110	.304
2.3	26-35	3.404	3.530	0.126	.094	.179
2.4	36+	2.886	2.980	0.094	.080	.241
3	Adjusted for demographics only			0.101	.056	.073
4	Adjusted for interviewer experience and demographics			0.282	.050	.000[a]
5	Adjusted for interviewer experience, interview order, and demographics			0.003	.061	.963

[a]p-Value less than .05.

Model 2 presents the same unadjusted weighted estimates within four age groups. This model shows similar small but nonstatistically significant increases within age groups. Adjusting for demographics (Model 3) did not appreciably affect the estimate of change. When adjusted for interviewer experience and demographics (Model 4), the adjusted estimate of change increased to almost 0.3 gate questions and was statistically significant. Taking into account both interviewer experience and interview order within year (Model 5) reduced the estimated change to less than 0.01 gate questions.

Table 8.7 repeats some of the data on annual change in the average number of positively answered gate questions and also shows the regression coefficients for the interviewer experience variables. Note that NHSDA experience is estimated to reduce the average number of gate questions answered "yes" by 0.240. Similarly, the number of gate questions answered "yes" decreased with increasing interview order. All estimates are differences relative to the experienced interviewers' interviews 1 through 19 (the reference class). The largest decrease in the number of gate questions answered "yes" occurs for the interview-order category of 100 or more (0.358).

Table 8.8 shows the unadjusted (Model 1) and fully adjusted (Model 5) estimates for the unedited gate questions for any illicit substance. An adjustment for interview experience reduced the magnitude of the difference even though both unadjusted and adjusted estimates of change are not statistically significant.

Table 8.7 Year, Interviewer Experience,† and Order Effects for 16-Item Composite Gate Questions

Description	Estimate	SE	p-Value
1998 unadjusted	2.877		
1999 unadjusted	2.970		
Change from 1998 to 1999			
Before adjustment (Model 1)	0.093	.043	.122
Model 5 adjustment	-0.003	.061	.963
Interviewer experience			
None (reference class)	0.000		
Some	-0.240	.059	.000[a]
Interview order			
1-19 (reference class)	0.000		
20-39	-0.124	.062	.045[a]
40-59	-0.099	.070	.161
60-99	-0.206	.071	.004[a]
100+	-0.358	.063	.000[a]

†"Experience" refers specifically to NHSDA interviewing experience.
[a] p-Value less than .05.

Table 8.8 Year, Interviewer Experience,† and Order Effects for Any Illicit Substance Gate Questions

Description	Estimate	SE	p-Value
1998 unadjusted	0.877		
1999 unadjusted	0.953		
Change from 1998 to 1999			
Before adjustment (Model 1)	0.076	.042	.073
Model 5 adjustment	0.023	.044	.603
Interviewer experience			
None (reference class)	0.000		
Some	-0.126	.042	.003[a]
Interview order			
1-19 (reference class)	0.000		
20-39	-0.092	.048	.057
40-59	-0.074	.053	.158
60-99	-0.124	.051	.016[a]
100+	-0.238	.049	.000[a]

†"Experience" refers specifically to NHSDA interviewing experience.
[a] p-Value less than .05.

Table 8.9, which shows model results for nonmedical use of any psychotherapeutic, presents a case where a statistically significant estimate of change from 1998 to 1999 is reduced by about half (from 0.040 before adjustment to 0.019 after adjustment) after adjusting for interviewer experience and is no longer statistically significant (p-value is 0.002 before adjustment and 0.162 after adjustment). All analyses of unedited gate question counts (***Tables 8.7*** through ***8.9***) show a consistent direction in the estimates of the effect of both measures of interviewer experience on the

Table 8.9 Year, Interviewer Experience,† and Order Effects for Any Psychotherapeutic Gate Questions, 1998-1999 PAPI

Description	Estimate	SE	p-Value
1998 unadjusted	0.146		
1999 unadjusted	0.186		
Change from 1998 to 1999			
Before adjustment (Model 1)	0.040	.013	.002[a]
Model 5 adjustment	0.019	.013	.162
Interviewer experience			
None (reference class)	0.000		
Some	-0.049	.013	.000[a]
Interview order			
1-19 (reference class)	0.000		
20-39	-0.032	.016	.044[a]
40-59	-0.037	.017	.029[a]
60-99	-0.040	.018	.028[a]
100+	-0.077	.015	.000[a]

†"Experience" refers specifically to NHSDA interviewing experience
[a]p-Value less than .05.

number of gate questions answered affirmatively (i.e., a decrease in gate questions answered "yes" with some NHSDA interviewing experience) and declines relative to the first 19 interviews with increasing order categories.

8.5 Impact on Specific Reported Measures

In addition to looking at the unedited gate question counts above, efforts were made to examine the possible effects of adjusting for interviewer experience on measures of change for reported prevalence measures of lifetime, past year, and past month use for three substance categories: any illicit substance, marijuana, and nonmedical use of any psychotherapeutic (***Table 8.10***). For modeling the effects of interviewer experience in

Table 8.10 Odds Ratios for Year, Interviewer Experience,† and Order Effects for Any Illicit Substance, Marijuana, and Nonmedical Use of Any Psychotherapeutic, 1998-1999 PAPI

All Illicit	**Lifetime**	**Past Year**	**Past Month**
Change from 1998 to 1999			
Before adjustment (Model 1)	1.12[a]	1.23[a]	1.22[a]
Model 5 adjustment	1.06	1.14[a]	1.15[a]
Interviewer experience			
None (reference class)	1.00	1.00	1.00
Some	0.84[a]	0.78[a]	0.83[a]
Interview order			
1-19 (reference class)	1.00	1.00	1.00
20-39	0.98	0.93	0.94
40-59	0.93	0.98	0.98
60-99	0.89	1.00	0.97
100+	0.84[a]	0.86	0.92
Marijuana	**Lifetime**	**Past Year**	**Past Month**
Change from 1998 to 1999			
Before adjustment (Model 1)	1.08	1.15[a]	1.13
Model 5 adjustment	1.04	1.10	1.08
Interviewer experience			
None (reference class)	1.00	1.00	1.00
Some	0.90[a]	0.82[a]	0.86
Interview order			
1-19 (reference class)	1.00	1.00	1.00
20-39	0.99	0.88	0.89
40-59	0.98	0.99	1.02
60-99	0.93	1.04	1.01
100+	0.87[a]	0.94	0.93
Any Psychotherapeutic	**Lifetime**	**Past Year**	**Past Month**
Change from 1998 to 1999			
Before adjustment (Model 1)	1.26	1.50	1.68[a]
Model 5 adjustment	1.10	1.28	1.41[a]
Interviewer experience			
None (reference class)	1.00	1.00	1.00
Some	0.69[a]	0.64	0.59[a]
Interview order			
1-19 (reference class)	1.00	1.00	1.00
20-39	0.83	0.99	1.18
40-59	0.80	0.82	0.85
60-99	0.74	0.84	0.99
100+	0.61	0.77	0.87

†"Experience" refers specifically to NHSDA interviewing experience.
[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

these tables, logistic regression was used, and the results are shown in terms of odds ratios. In these models, the response variable was substance use (dichotomous outcome, 1 = yes, 0 = no). Odds ratios highlighted in bold type and that are greater than 1 for the "change from 1998 to 1999" effect indicate that 1999 estimates are significantly higher than 1998 estimates; these odds ratios are statistically significant ($\alpha = 0.05$).

Results are shown before and after adjustment for demographics, interviewer experience, and interview order. In all cases, adjustment reduces the magnitude of the year-to-year change and, in many cases, moves it from a statistically significant change to a nonsignificant one. ***Tables 8.11***, ***8.12***, and ***8.13*** show results from age-specific models of interviewer experience for lifetime and past month use for three substance categories: any illicit substance, marijuana, and nonmedical use of any psychotherapeutic. Covariates used in these age-specific models included (1) U.S. Bureau of the Census region (Northeast, North Central, South, and West), (2) gender, (3) race/ethnicity (Hispanic, non-Hispanic black, and non-Hispanic nonblack), and (4) population density (1 million or more persons in a Metropolitan Statistical Area [MSA], 250,000 to 999,999 persons in an MSA, less than 250,000 persons in an MSA, persons not in an MSA and not in a rural area, and persons not in an MSA and in a rural area). Population density was added to take into account differences in substance use rates among large MSAs, small MSAs, and non-MSAs. As before, results are shown before and after adjustment for demographics, interviewer experience, and interview order. Similarly, across age groups and substance, the adjustment reduces the magnitude of the year-to-year change and generally moves it from a statistically significant change to a nonsignificant change.

Adjustment of model estimates of annual change in substance use (1998 to 1999) showed a consistent pattern of reducing 1999 substance use estimates relative to 1998 estimates. Thus, it appears that much of the observed annual change in prevalence is an artifact of the changing mix in interview experience across the 2 years, and caution should be used in reporting increases in substance abuse based on the otherwise comparable PAPI surveys.

8.6 Inclusion of Other Covariates

A number of potential confounders of interviewer experience were suggested after the initial experience analysis results were presented to the field staff. These factors were addressed by including them in the full regression model (Model 5). As explained below, none of the suggested covariates reduced the effect of interviewer experience on the prevalence estimates.

Table 8.11 Odds Ratios for Year, Interviewer Experience,† and Order Effects for Any Illicit Substance, by Age Category, 1998-1999 PAPI

	Lifetime				Past Month			
Description	12-17	18-25	26-34	35+	12-17	18-25	26-34	35+
Change from 1998 to 1999								
Before adjustment (Model 1)	1.23[a]	1.17[a]	1.02	1.15[a]	1.08	1.14	1.38[a]	1.30
Model 5 adjustment	1.13	1.15	0.95	1.07	1.00	1.20	1.31	1.16
Interviewer experience								
None (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Some	0.74[a]	0.87	0.84[a]	0.84[a]	0.78[a]	1.02	0.85	0.71
Interview order								
1-19 (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20-39	0.85	1.01	0.99	0.99	0.91	0.91	0.83	1.02
40-59	1.02	1.04	0.95	0.91	1.02	1.04	1.00	0.92
60-99	0.81	0.90	0.87	0.91	0.74	1.00	1.05	0.99
100+	0.90	0.89	0.75[a]	0.87	0.87	0.91	0.71	1.14

†Note: Experience refers specifically to NHSDA interviewing experience.
[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

Table 8.12 Odds Ratios for Year, Interviewer Experience,† and Order Effects for Marijuana, by Age Category, 1998-1999 PAPI

	Lifetime				Past Month			
Description	12-17	18-25	26-34	35+	12-17	18-25	26-34	35+
Change from 1998 to 1999								
Before adjustment (Model 1)	1.11	1.10	0.95	1.12	0.98	1.16	1.33	1.09
Model 5 adjustment	1.07	1.10	0.91	1.08	0.94	1.20	1.25	1.00
Interviewer experience								
None (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Some	0.83	0.90	0.91	0.89	0.85	0.96	0.83	0.81
Interview order								
1-19 (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20-39	1.01	1.00	0.96	1.00	0.98	0.93	0.79	0.86
40-59	1.29	1.06	0.99	0.94	1.11	1.13	1.08	0.85
60-99	0.93	0.98	0.87	0.93	0.96	1.11	1.18	0.84
100+	1.03	0.99	0.78	0.87	0.93	1.01	0.65	1.10

†Note: Experience refers specifically to NHSDA interviewing experience.
[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

Recent research indicates the possibility that a seasonal pattern of substance use may occur for certain substances and age groups (Huang, Schidhaus, & Wright, September 1999). To account for this effect, an additional model was analyzed, which included all variables in Model 5 plus calendar quarter, a variable often used to capture seasonality. The regression analysis showed that quarter was only marginally significant. In addition, both measures of interviewer experience remained highly significant

Table 8.13 Odds Ratios for Year, Interviewer Experience,† and Order Effects for Nonmedical Use of Any Psychotherapeutic, by Age Category, 1998-1999 PAPI

	Lifetime				Past Month			
Description	**12-17**	**18-25**	**26-34**	**35+**	**12-17**	**18-25**	**26-34**	**35+**
Change from 1998 to 1999								
Before adjustment (Model 1)	1.47[a]	1.42[a]	1.25[a]	1.21	1.64[a]	1.35	1.43	2.12[a]
Model 5 adjustment	1.26	1.32[a]	1.06	1.05	1.36	1.39	1.36	1.57
Interviewer experience								
None (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Some	0.62[a]	0.76[a]	0.65[a]	0.68[a]	0.51[a]	1.01	0.84	0.36[a]
Interview order								
1-19 (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20-39	0.67[a]	0.93	1.04	0.77[a]	0.88	1.23	0.88	1.46
40-59	0.78	0.97	0.77	0.79	0.83	0.93	0.82	0.82
60-99	0.76	0.69[a]	0.75	0.75[a]	0.55	0.88	0.59	1.51
100+	0.63[a]	0.58[a]	0.59[a]	0.65[a]	0.79	0.72	0.77	1.11

†Note: Experience refers specifically to NHSDA interviewing experience.
[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

in these models with the parameter estimates maintaining their same direction as in Model 5.

Typical behavior in most surveys is to delay more difficult interviews (respondent not at home after repeated attempts, initial nonrespondents, etc.) until the end of the data collection period. In the NHSDA, this is the last month of each calendar quarter. It has also been observed that inexperienced NHSDA interviewers tend to get assignments in areas where substance use may be highest (i.e., metropolitan areas). To capture this information in the models, an alternative time variable (month in quarter) was developed as a main effect to represent the difference in case types; for the interviewer assignments, an interaction term (interviewer experience × population density) was also included. However, in further testing, the impact of these variables was not found to be statistically significant in most models.

Based on discussions with field interviewers, the suggestion was made that more experienced interviewers receive a larger proportion of interviews in higher income areas, which tend to have lower prevalence rates. In fact, supervisors and interviewers are aware of the neighborhood income level before assigning cases to interviewers. The assumption is that these experienced interviewers are better able to produce interviews in these higher income areas. A number of segment-level surrogates for income level were used, including (1) percent owner-occupied residents,

(2) median housing unit rental value, and (3) median owner-occupied housing unit value. The distribution of interviewer experience by these income measures showed that more experienced interviewers actually complete a higher proportion of interviews in lower income areas ($\chi^2 < 0.01$). Initially, a regression analysis on the number of "yes" gate questions was performed including all covariates in Model 5 plus an income measure, resulting in Model 6. ***Table 8.14*** shows the comparison of Model 5 versus Model 6 and the parameter estimates for experience and income in Model 6. After the initial analysis, Model 6 was used to assess the effect of experience on specific prevalence rates given income in the model. These income variables were not found to be statistically significant in any of the lifetime models. However, percent owner-occupied was found to be significant in the past year and past month substance use models of any illicit substance and any illicit substance except marijuana. Yet income did not diminish the effect of interviewer experience or interview order. Both measures of experience remained highly significant with the same parameter directions as in Model 5. ***Table 8.15*** shows the effect of percent owner-occupied on lifetime, past year, and past month any illicit substance use and any illicit substance except marijuana.

Table 8.14 Year, Interviewer Experience,† Order Effects, and Income for 16-Item Composite Gate Questions, 1998-1999 PAPI

Description	Estimate	SE	p-Value
Change from 1998 to 1999			
Before adjustment (Model 1)	0.093	.043	.122
Model 5 adjustment	-0.003	.061	.963
Model 6 adjustment	-0.002	.061	.977
Interviewer experience			
Some	-0.240	.058	.000[a]
None (reference class)	0.000		
Interview order			
1-19 (reference class)	0.000		
20-39	-0.124	.062	.044[a]
40-59	-0.099	.071	.160
60-99	-0.206	.071	.004[a]
100+	-0.356	.064	.000[a]
% Owner occupied (income)			
>50 (reference cell)	0.000		
10-50	0.055	.070	.433
<10	-0.060	.134	.657

†"Experience" refers specifically to NHSDA interviewing experience.
[a] p-Value less than 0.05.

Table 8.15 Odds Ratios for Year and Income for Any Illicit Substance and Any Illicit Substance Except Marijuana, 1998-1999 PAPI

Description	Any Illicit Lifetime	Any Illicit Past Year	Any Illicit Past Month	Any Illicit Except Marijuana Lifetime	Any Illicit Except Marijuana Past Year	Any Illicit Except Marijuana Past Month
Change from 1998 to 1999						
Before adjustment (Model 1)	1.12[a]	1.23[a]	1.22[a]	1.08	1.15[a]	1.13
Model 5 adjustment	1.06	1.14[a]	1.15[a]	1.04	1.10	1.08
Model 5 + income	1.06	1.14[a]	1.16[a]	1.05	1.22[a]	1.24[a]
% Owner occupied (income)						
> 50% (reference cell)	1.00	1.00	1.00	1.00	1.00	1.00
10-50%	1.06	1.32[a]	1.40[a]	1.12	1.28[a]	1.29[a]
<10%	0.93	1.33[a]	1.50[a]	1.10	1.52[a]	1.63[a]

[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

8.7 Model-Based vs. Weight-Based Estimation Methodology and Impact on Estimates

At this point in the analysis, it was determined that changes in the mix of interviewers with NHSDA interviewing experience between 1998 and 1999 had an impact on the estimates of annual change and that an adjustment procedure had to be developed for 1999 PAPI estimates. Two approaches were considered: the first used models as described in Section 8.5, and the second recomputed the 1999 analysis weights taking into account interviewer experience and the interview order distribution observed in 1998. Estimates were generated for the following 12 substance use measures for persons aged 12 years or older and those aged 12 to 17: lifetime and past month use of any illicit substance, marijuana, nonmedical use of any psychotherapeutic, analgesics, alcohol, and cigarettes. With few exceptions, estimates from the two methods were very similar.

Because the model-based and weight-based estimates were quite similar in most cases, the decision was made to use the weight-based method because of its computational simplicity. In addition, the calculation of weights was independent of any reported substance use (or other) variables; like the original weights, the adjusted weights could be used to calculate estimates for any characteristics of interest. This approach was facilitated by introducing additional controls in the poststratification step of the weighting process. The additional controls represented two levels of NHSDA interview experience (NHSDA experience and no NHSDA experience) and three levels of the interview order number (1-19, 20-99, and 100+). As a result, the number of controls was increased from 131 to

134. The control totals were derived by using the 1998 weighted distribution as shown in ***Table 8.5*** (61.69 percent with NHSDA experience vs. 38.31 percent with no NHSDA experience; 28.14 percent with interview number 1 to 19, 54.24 percent in the category 20 to 99, and 17.64 percent in the 100+ category) on the 1999 survey population.

In view of the large discrepancies between the distributions of the interviewer characteristics over the 2 years, the bounds on the poststratification adjustment factor had to be broadened to keep the same set of covariates in the model in addition to the new interviewer experience covariates. As a result, the realized design effect for the total sample increased from 3.01 to 5.77 because, on average, the adjusted weights were about twice as large as the original weights for the NHSDA-experienced interviewer data and were half the size of original weights for data corresponding to interviewers with no NHSDA experience.

As expected, the adjusted 1999 PAPI estimates were generally lower than unadjusted estimates for the illicit substance. However, the adjusted estimates will have limited use because of significantly higher variability induced by the weight adjustment. These adjusted estimates were presented in the 1999 NHSDA Summary of Findings Report (SAMHSA, 2000).

8.8 Impact of Field Interviewer Experience on 1999 CAI Estimates

The 1999 NHSDA Summary of Findings Report (SAMHSA, 2000) stated that the large change in the distribution of experienced and inexperienced field interviewers between the 1998 and 1999 surveys was associated with unanticipated and unusually large increases in substance use rates for data collected using the PAPI method. The report also found that data collected from interviewers with NHSDA experience resulted in substance use rates that were significantly lower than those based on data collected from interviewers with no NHSDA experience. As a result, the 1999 PAPI estimates presented in the SAMHSA report were based on analysis weights that were adjusted to measures representing the 1998 FI experience distribution.

Along with fielding PAPI data, the 1999 NHSDA marked the beginning of the use of CAI methods to solicit data from more than 66,000 respondents in 50 States and the District of Columbia that year. This section focuses on the analysis of the 1999 CAI data to determine the impact of FI experience on substance use estimates. Overall, it was found that the interviewer effects remain, though not as pronounced as found in the PAPI data.

Similar to analyses of the 1998 and 1999 PAPI data, field interviewer experience for the 1999 CAI data was defined in two ways: (1) a two-level overall experience variable (no NHSDA experience, some NHSDA experience) and (2) by interview order, which is a measure of experience level over the course of the survey year (i.e., 1 = first interview conducted, 100 = 100th interview conducted). Here interview order was defined in terms of a five-level variable (1-19, 20-39, 40-59, 60-99, and 100+). For the 1999 CAI, interviewers with no experience were simply those who did not have any NHSDA interviewing experience prior to the 1999 survey.

Tables 8.16 and ***8.17*** present the distribution of CAI field interviewers and interviews in 1999 according to interviewer experience. More than 86 percent of the CAI interviewer workforce had no NHSDA experience. The large number of inexperienced interviewers in 1999 was due to extensive hiring to work the sample of 66,706, which had expanded threefold from 1998. These inexperienced interviewers were responsible for about 78 percent of the CAI interviews (Table 8.16). Note that over half of the interviews were conducted by field interviewers (FIs) before their 40^{th} interview in either survey year. The weighted and unweighted distributions in ***Table 8.17*** are similar. Overall, the 1999 FI workforce and collected data were dominated by inexperienced interviewers.

Tables 8.18 and ***8.19*** compare 1999 CAI and PAPI weighted estimates of lifetime use of any illicit substance and nonmedical use of any psychotherapeutic by interviewer experience and interview order. Both the 1999 PAPI and CAI estimates show a decreasing trend as the interview order increases; also, estimates within a given year and interview order were higher among interviewers with no NHSDA experience than among those with some experience. However, the decline among PAPI interviewers was generally larger than among CAI interviewers: for example, a 38.8 percent decrease overall in rates of lifetime use of any nonmedical psychotherapeutic substance between the 1-19 and 100+ interview order group (from 13.4 to 8.2 percent) (***Table 8.19***), compared with 15.8 percent (from 15.8 percent to 13.3 percent in the CAI group). Estimates of lifetime use of any illicit substance also declined for both PAPI and CAI overall, although at a slower rate between the lowest and highest interview order groups among CAI interviewers.

Table 8.16 Interviewers and Interviews by Interviewer Experience† Distribution, 1999 CAI

Interviewer Experience	CAI Interviewers		CAI Interviews	
	No.	%	No.	%
None	1,544	86.40	52,322	78.44
Some	243	13.60	14,384	21.56
Total	1,787	100.00	66,706	100.00

†"Experience" refers specifically to NHSDA interviewing experience.

Table 8.17 Unweighted and Weighted Distribution of CAI Interviews by Interview Order and Interviewer Experience,† 1999 CAI

Interview Order	No Experience		Some Experience		Total
	No.	%	No.	%	%
	Unweighted CAI				
1-19	18,713	28.05	2,999	4.50	32.55
20-39	12,088	18.12	2,656	3.98	22.10
40-59	7,902	11.85	2,262	3.39	15.24
60-99	8,505	12.75	3,076	4.61	17.36
100 +	5,114	7.67	3,391	5.08	12.75
Subtotals	52,322	78.44	14,384	21.56	100.00
Total	66,706				
	Weighted CAI				
1-19	66,339	30.00	14,760	6.68	36.68
20-39	39,169	17.71	12,646	5.72	23.43
40-59	22,925	10.37	8,582	3.88	14.25
60-99	22,507	10.18	11,166	5.05	15.23
100 +	12,416	5.61	10,613	4.80	10.41
Subtotals	163,355	73.87	57,768	26.13	100.00
Total	221,123				

†"Experience" refers specifically to NHSDA interviewing experience.

To investigate the effects of adjusting for interview experience on various measures of change, a logistic regression model was used with the results shown as odds ratios. RTI's SUDAAN was employed and the analysis weights were used. The sample structure was represented using standard NHSDA analysis NEST statements for variance strata and variance replicates. The substance use measures modeled were lifetime, past year (not shown), and past month use of any illicit substance, marijuana, any illicit substance except marijuana, nonmedical use of any psychotherapeutic, and analgesics (***Tables 8.20 - 8.24***). In these models, the

Table 8.18 Percentage Reporting Lifetime Use of Any Illicit Substance by Interview Order and Interviewer Experience,† 1999 PAPI and 1999 CAI

	1999 PAPI			1999 CAI		
Interview Order	**No Experience**	**Some Experience**	**All**	**No Experience**	**Some Experience**	**All**
1-19	39.9	36.3	39.3	41.5	39.5	41.1
20-39	40.3	41.8	40.7	40.8	39.4	40.5
40-59	38.0	37.7	37.9	38.9	35.4	38.0
60-99	37.7	37.8	37.7	40.7	34.8	38.7
100+	35.7	30.6	33.8	37.1	35.8	36.5
All interviews	38.9	37.1	38.5	40.5	37.3	39.7
% change from 1-19 to 100+ interviews	-10.5	-15.7	-14.0	-10.6	-9.4	-11.2

†"Experience" refers specifically to NHSDA interviewing experience.

Table 8.19 Percentage Reporting Lifetime Nonmedical Use of Any Psychotherapeutic by Interview Order and Interviewer Experience,† 1999 PAPI and 1999 CAI

	1999 PAPI			1999 CAI		
Interview Order	**No Experience**	**Some Experience**	**All**	**No Experience**	**Some Experience**	**All**
1-19	13.3	13.8	13.4	16.0	14.8	15.8
20-39	11.9	10.9	11.7	16.5	16.4	16.5
40-59	12.7	7.2	11.1	15.6	11.4	14.4
60-99	10.6	8.5	10.0	16.2	13.3	15.2
100+	9.2	6.7	8.2	14.2	12.2	13.3
All interviews	12.0	9.7	11.4	16.0	13.9	15.4
% change from 1-19 to 100+ interviews	-30.8	-51.4	-38.8	-11.3	-17.6	-15.8

†"Experience" refers specifically to NHSDA interviewing experience.

response variable was a dichotomous measure of substance use (1 = yes, 0 = no).

Significant odds ratios that are less than 1 represent estimates that are significantly lower than those from the reference class (at the $\alpha = 0.05$ level of significance), after controlling on a set of covariates. The covariates used are:

- Interviewer experience (no NHSDA experience, some NHSDA experience)
- Interview order (1-19, 20-39, 40-59, 60-99, and 100+)
- Age of respondent (12-17, 18-25, 26-34, 35+)

Table 8.20 Odds Ratios for Prior Interviewer Experience and Order Effects for Any Illicit Drug, by Age Category, 1999 CAI Only

	Lifetime				Past Month			
Description	**12-17**	**18-25**	**26-34**	**35+**	**12-17**	**18-25**	**26-34**	**35+**
Prior interviewer experience								
No NHSDA (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Some NHSDA	0.90[a]	0.92	0.93	0.88	0.89	0.89	1.05	1.11
Interview order								
1-19 (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20-39	0.96	0.94	0.99	0.99	0.92	0.90	0.99	1.01
40-59	0.89[a]	0.95	0.83	0.87	0.91	1.00	0.88	0.72
60-99	0.86[a]	0.92	0.86	0.97	0.84[a]	1.04	0.89	0.80
100+	0.82[a]	0.85[a]	0.85	0.84	0.86	0.88	1.00	0.77

[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

Table 8.21 Odds Ratios for Prior Interviewer Experience and Order Effects for Marijuana, by Age Category, 1999 CAI Only

	Lifetime				Past Month			
Description	**12-17**	**18-25**	**26-34**	**35+**	**12-17**	**18-25**	**26-34**	**35+**
Prior interviewer experience								
No NHSDA (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Some NHSDA	0.89[a]	0.94	0.98	0.90	0.85	0.90	1.02	0.92
Interview order								
1-19 (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20-39	0.98	0.92	0.99	0.94	1.00	0.88	1.01	0.92
40-59	0.96	0.97	0.84	0.88	1.00	0.99	0.87	0.66
60-99	0.89	0.90	0.89	1.00	0.93	1.03	0.89	0.79
100+	0.97	0.82[a]	0.91	0.90	0.97	0.88	1.03	0.76

[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

- U.S. Bureau of the Census region (Northeast, North Central, South, and West)
- Gender of respondent
- Race/ethnicity of respondent (Hispanic, non-Hispanic black, and non-Hispanic nonblack, all other races
- Population density (1 million or more persons in an MSA, 250,000 to 999,999 persons in an MSA, less than 250,000 persons in an MSA, persons not in an MSA and not in a rural area, and persons not in an MSA and in a rural area).

Table 8.22 Odds Ratios for Prior Interviewer Experience and Order Effects for Any Illicit Substance Except Marijuana, by Age Category, 1999 CAI Only

	Lifetime				Past Month			
Description	**12-17**	**18-25**	**26-34**	**35+**	**12-17**	**18-25**	**26-34**	**35+**
Prior interviewer experience								
No NHSDA (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Some NHSDA	0.97	0.89[a]	0.80[a]	0.86	0.97	0.91	1.11	1.74[a]
Interview order								
1-19 (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20-39	0.97	1.01	0.93	1.05	0.89	1.07	1.06	0.98
40-59	0.93	1.01	0.91	0.91	0.94	1.28[a]	1.13	0.72
60-99	0.89[a]	1.02	0.92	0.95	0.85	1.09	0.64	0.71
100+	0.81[a]	0.94	0.80[a]	0.81	0.80	0.87	0.82	0.93

[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

Table 8.23 Odds Ratios for Prior Interviewer Experience and Order Effects for Nonmedical Use of Any Psychotherapeutic, by Age Category, 1999 CAI Only

	Lifetime				Past Month			
Description	**12-17**	**18-25**	**26-34**	**35+**	**12-17**	**18-25**	**26-34**	**35+**
Prior interviewer experience								
No NHSDA (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Some NHSDA	0.96	0.90	0.90	0.83[a]	0.85	0.92	1.22	1.88[a]
Interview order								
1-19 (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20-39	1.04	1.05	0.92	1.11	0.86	1.14	1.16	1.05
40-59	0.92	1.06	0.79[a]	0.91	0.93	1.25	0.75	0.91
60-99	0.89	0.97	0.91	1.03	0.83	0.89	0.80	0.70
100+	0.80[a]	0.91	0.86	0.82	0.77	0.84	1.15	0.76

[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

Most notable are the generally lower odds ratios for experienced interviewers compared to those with no experience. The exception appears to be among persons age 35 or older, where odds ratios are higher for various past month drug use measures. Compared to the PAPI analysis (using exactly the same model on the 1998 and 1999 PAPI data), the CAI odds ratios comparing experienced to inexperienced interviewers are higher and

Table 8.24 Odds Ratios for Prior Interviewer Experience and Order Effects for Analgesics, by Age Category, 1999 CAI Only

	Lifetime				Past Month			
Description	**12-17**	**18-25**	**26-34**	**35+**	**12-17**	**18-25**	**26-34**	**35+**
Prior interviewer experience								
No NHSDA (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Some NHSDA	0.95	0.95	0.86	0.97	0.82	0.91	1.33	2.03[a]
Interview order								
1-19 (reference class)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20-39	1.09	1.12	0.91	1.07	0.85	1.38[a]	1.00	0.78
40-59	0.90	1.12	0.86	0.78	0.84	1.45[a]	0.78	0.49
60-99	0.88	1.03	1.01	0.88	0.84	1.09	0.70	0.54
100+	0.72[a]	0.99	0.88	0.79	0.82	0.95	0.87	0.44

[a]Odds ratio is statistically different from 1.00 at the 0.05 level of significance.

closer to 1.00. For example, the PAPI odds ratios for nonmedical use of any psychotherapeutic substance during the lifetime vary from 0.62 to 0.76 among the four age groups (***Table 8.13***); in CAI, odds ratios vary from 0.83 to 0.96 among the same four age groups (***Table 8.23***). In these same two tables, PAPI and CAI odds ratios differ significantly among persons age 35 or older for past month nonmedical use of any psychotherapeutic substance (0.36 for PAPI vs.1.88 for CAI). It is not clear what may be causing this amount of variation to occur; so, further studies will be needed to develop an understanding of this phenomenon.

8.9 Conclusions

The analysis presented in this chapter indicates that the uneven mix of experienced and inexperienced NHSDA field interviewers in 1999 had some effect on estimated substance use rates for that year. Overall, the effect on 1999 CAI prevalence rates is smaller in magnitude than the effect on 1999 PAPI rates, which is an indication that the CAI methods are playing a role in reducing the effects of FI experience on substance use rates. However, because the mechanism of these effects is unknown, additional studies will be undertaken to increase our understanding of this phenomenon. In the meantime, analyses of interviewer effect as seen in this chapter will continue to be presented in subsequent reports. These findings have resulted in an added emphasis—in training and in the field—on encouraging experienced and new FIs to follow the interview protocol.

8.10 References

Huang, L., Schidhaus, S., & Wright, D. (1999, September). *Seasonality of substance use: National Household Survey on Drug Abuse 1992-96* (working paper prepared for the Substance Abuse and Health Services Administration, Office of Applied Studies).

Keeter, S., Miller, C., Kohut, A., Groves, R.M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, *64*, 125-148.

Substance Abuse and Mental Health Services Administration. (2000). *Summary of findings from the 1999 NHSDA*. Office of Applied Studies, Series H-12 (DHHS Pub. No. SMA 00-3466). Rockville, MD: Author.

Turner, C.F., Lessler, J.T., & Devore, J.C. (1992). Effects of mode of administration and wording on reporting of drug use. In C. Turner, J. Lessler, & J. Gfroerer, Eds., *Survey measurement of drug use: Methodological studies*. National Institute on Drug Abuse (DHHS Pub. No. ADM 92-1929). Rockville, MD: National Institutes of Health.

9

Changes in NHSDA Measures of Substance Use Initiation

Lisa Packer, Dawn Odom, James Chromy, Teresa Davis, and Joseph Gfroerer

The National Household Survey on Drug Abuse (NHSDA) data are used to generate a number of annual estimates relating to the initiation of substance use. Using the responses to retrospective questions about age of first use, annual estimates are generated for the incidence rate of first substance use, for the number of initiates to substance use, and for the average age of first use. Estimates of new initiates and average age of first use are reported for all lifetime substance users aged 12 or older. These use initiation measures are important since they can capture the rapidity with which new substance users arise in specific population subgroups and can identify emerging patterns of substance use.

The redesign of the NHSDA in 1999 introduced some changes in the questions about initiation as well as the method of administration. In the presence of these changes, the overall data processing and estimation methodologies were reviewed and, in some cases, revised. The revisions to the editing and imputation procedures are summarized in ***Chapters 5*** and ***6***. Limitations of the existing methodology for computing incidence rates were found and, as a result, a new incidence rate methodology was developed. The definition of initiation of daily cigarette use was modified and an adjustment to the program logic in the calculation for the incidence of first daily use of cigarettes was also made.

This chapter is organized in three sections addressing the impact of methodological change on substance use initiation measures. ***Section 9.1*** describes the old and new incidence rate estimation methods and evaluates its impact; the impact of the editing and imputing changes is evaluated in conjunction with the method impact. ***Section 9.2*** focuses on the questionnaire wording and administration mode effects. **Section *9.3*** focuses on all the issues associated with initiation of first daily use of cigarettes. ***Section 9.4*** summarizes the chapter and presents some recommendations.

9.1 Impact of Imputation and Revised Incidence Rate Calculation Method

Prior to 1999, a respondent's year of first use of each substance was determined based on the reported age of first use, the date of the interview, and the date of birth of the respondent. The year of first use was used to develop estimates of annual initiates and to develop the respondent-level numerator and denominator inputs to the incidence rate calculation and was then dropped from the permanent data file. The new combined editing and imputation procedures flag more inconsistencies, impute for both missing and inconsistent reports, and retain the imputed date of first use consistent with reported age of first use and other substance use measures. The availability of an imputed date of first use for each lifetime substance user enabled consideration of a more precise approach to calculation of substance use incidence rates.

9.1.1 Background

For diseases, the incidence rate for a population is defined as the number of new cases of the disease, *N,* divided by the person time, *PT*, of exposure or

$$IR = \frac{N}{PT} \ .$$

The person time of exposure is measured as the net time that individuals in the population during an observed period of time are at risk of developing the disease. The person time of exposure ends at the time of diagnosis (e.g., Greenberg, Daniels, Flanders, Eley, and Boring, 1996, pp. 16-19). Similar conventions are applied for defining substance use incidence.

In the 1998 and earlier NHSDA instruments, the only questions about initiation of substance use ask the respondents to report their age of first use for specified substances. Using this age of first use information, along with date of birth and interview date, an approximate date of first use was calculated for each substance a respondent reported ever using. In the redesigned NHSDA computer-assisted interviewing (CAI) instrument, additional questions about initiation of substance use were included. Recent initiates (persons who first used at their current age or at their current age minus one) were asked to report not only their age of first use for specified substances, but also the month and year of first use. As a result, the exact month of first use for specified substances was known completely or in part (sometimes month or year were not reported) for 7 to 16 percent of the substance users (depending on the substance) in the 1999 NHSDA sample. Another difference in the basic questions asked involved an additional prompt used for substances in the paper and pencil interview (PAPI) questionnaire only. The PAPI modules for alcohol, marijuana, and heroin used the following additional probe following the age of first use question: "If you're not sure how old you were, make your best guess."

In addition to the new questions about recent substance initiation in the 1999 CAI, the questionnaire also changed due to the routing logic used in the CAI instrument, which helped automatically resolve data inconsistencies between related items. For example, respondents were asked their age of first substance use and were prompted to review their response if the reported age of first use was inconsistent with their reported current age. With more exact data available in the 1999 survey for a portion of the sample, both the age of first use data and the incidence methodology were examined to determine if refinements could be made to increase the validity, quality, and consistency of the incidence estimates.

This examination led to several methodology changes used in the calculation of the 1999 NHSDA substance incidence rates. First, missing age of first use data were imputed, which resulted in consistent and nonmissing age of first use data for all users. Prior to the 1999 data, respondents with a missing age of first use were simply excluded from the calculation of incidence numbers and rates. Second, the assignment of the date of first use was refined so that the assigned date is now consistent with other reported related information such as substance use recency and frequency data. Last, the improved age of first substance use and date of first substance use data allows the determination of a more exact person time of exposure during the targeted period.

9.1.2 Old Incidence Rate Calculation Method and Motivation for Change

Prior to 1999 CAI, the only initiation data collected were the respondents' age of first substance use. An approximate date of first use was calculated to be consistent with the respondents' reported age of first use and their interview and birth date. No attempt was made to ensure consistency with other reported substance use measures. A date of first use was randomly assigned within the 365 days corresponding to the respondents' age when they first began substance use. However, only the year of first use was retained and used in the analysis. Moreover, no imputation was performed on the age of first use data, and the incidence estimates excluded respondents with missing data.

The old methodology computed age-specific rates by considering the first use of substance *d* in age group *a* during year *y*. These rates can be described as the number of new users per 1,000 potential new users. The incidence rate is estimated as a weighted ratio estimate:

$$IR(d,y) = \frac{\sum_i w_i \, I_i(d,a,y)}{\sum_i w_i \, e_i(d,a,y)} ,$$

where the indicator functions for the numerator and denominator, respectively, $I_i(d,a,y)$ and $e_i(d,a,y)$, are defined in ***Table 9.1*** (Johnson, Gerstein, Ghadialy, Choi, and Gfroerer, 1996).

These individual indicator functions are created for each age group by year so that the ratio estimates can be easily calculated. The I_i can have the value of 0 or 1. If the respondent had a birthday in age interval *a* during year *y* and first used substance *d* in that year *y*, then they are given a value of 1. Otherwise, the respondent is given a value of 0 for I_i. The denominator indicator function (e_i) indicating person-years of exposure can have values of 0, 1, or 0.5. If a respondent had no birthday in the age interval *a* during year *y* or if the respondent reported using the substance *d* prior to year *y*, then e_i was set to 0. If a respondent had a birthday during the age interval *a*, never used the substance *d* during the year *y*, and indicated that they had not initiated use of the substance, then e_i was assigned a value of 1. These cases are still at risk for initiation. If a respondent had a birthday during the age interval *a*, had never used prior to year *y*, and initiated use of substance *d* during year *y*, e_i was assigned a value of 0.5. The value of 0.5 assumed that new users initiated use approximately halfway through the year at age *a* during year *y*, so they were considered exposed to risk during the first half of this period.

Table 9.1 Indicator Functions Using the Old Methodology (limited to $y_2 = y_1 + 1$)

Indicator and Value	Conditions (Verbal)	Conditions (Symbolic)
$I_i(d,a,y) = 1$	(1) Respondent *i* had a birthday in the age interval *a* during year *y*, and (2) Respondent *i* first used substance *d* in year *y*†.	(1) $y_1 \in [y_{b,i} + a_1, y_{b,i} + a_2)$ and (2) $y_{fu,d,i} \in [y_1, y_2)$
$I_i(d,a,y) = 0$	In all other cases	(1) $y_1 \notin [y_{b,i} + a_1, y_{b,i} + a_2)$, or (2) $y_{fu,d,i} \notin [y_1, y_2)$
$e_i(d,a,y) = 0$	(1) Respondent *i* did not have a birthday in the age interval *a* during year *y*, or (2) Respondent *i* reported using substance *d* prior to year *y*.	(1) $y_1 \notin [y_{b,i} + a_1, y_{b,i} + a_2)$ or (2) $y_{fu,d,i} < y_1$
$e_i(d,a,y) = 1$	(1) Respondent *i* had a birthday in the age interval *a* during year *y*, and (2) Never used substance *d* prior to year *y*, and (3) Did *not* initiate use of substance *d* during year *y*.	(1) $y_1 \notin [y_{b,i} + a_1, y_{b,i} + a_2)$, and (2) $y_{fu,d,i} \geq y_1$, and (3) $y_{fu,d,i} \geq y_2$.
$e_i(d,a,y) = 0.5$	(1) Respondent *i* had a birthday in the age interval *a* during year *y*, and (2) Never used substance *d* prior to year *y*, and (3) *Did* initiate use of substance *d* during year *y*.	(1) $y_1 \in [y_{b,i} + a_1, y_{b,i} + a_2)$ and (2) $y_{fu,d,i} \geq y_1$, and (3) $y_{fu,d,i} < y_2$.

† Johnson et al. (1996) give the conditions of this indicator as " $I_i(d, a, y)$ is a 0-1 variable which takes on the value 1 if and only if respondent 'i' first used substance 'd' at age 'a' in year 'y'." However, the indicator was defined in the actual computer programs as stated in the table.

It helps to think of both *a* and *y* as half-open intervals with associated end points as follows:

$$[a_1, a_2) = \{a: a_1 \leq a < a_2\} \text{ and } [y_1, y_2) = \{y: y_1 \leq y < y_2\}$$

For example, if the unit of time measure is years, persons aged 12 to 17 can be defined by setting a_1 to 12 and a_2 to 18. Similarly, the year 1998 can be defined by setting y_1 to 1998 and y_2 to 1999. A 2-year interval of 1997 through 1998 can be similarly defined by setting y_1 to 1997 and y_2 to 1999 if desired. To determine the necessary indicator functions, $a_{fu,d,i}$ was defined as respondent *i*'s age at first use of substance *d*, $y_{fu,d,i}$ as respondent *i*'s year of first use for substance *d*, and $y_{b,i}$ as respondent *i*'s birth year. Although only substance *d* users have a valid age at first use, for computational purposes, we can set $y_{fu,d,i}$ at a high number well outside the range under consideration for persons who have never used substance *d*. ***Table 9.1*** describes computation of the indicator values in detail.

This method can provide a good approximation to the incidence rate under many conditions. One problem with the method is that some persons represented in the numerator for a certain age group *a* actually initiated in the younger age group *a-1*. Likewise, these persons were also being classified in the denominator for age group *a* and excluded from the denominator for age group *a-1*. For example, an individual born on 6/15/1980 who initiated substance use at age 17 on 2/15/1998 would be included in the 18- to 25-year-old age group in year 1998. However, it is clear that he/she was 17 when initiation occurred in 1998. ***Figure 9.1*** illustrates this point.

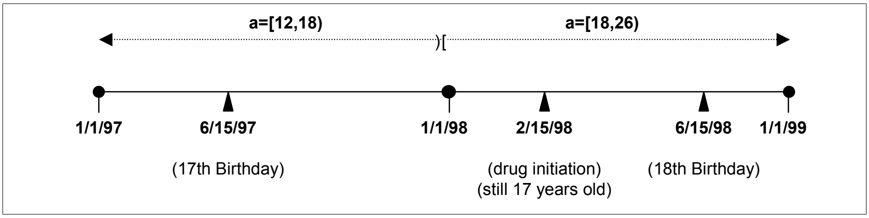

Figure 9.1 Line diagram illustrating the old methodology.

The indicators for this case would be: $I_{(d,12\text{-}17,1998)} = 0$; $e_{(d,12\text{-}17,1998)} = 0$; $I_{(d,18\text{-}25,1998)} = 1$; $e_{(d,18\text{-}25,1998)} = 0.5$.

9.1.3 New Incidence Rate Calculation Method

To overcome the shortcomings of the old methodology, one must account for the fact that age and calendar time do not intersect in whole years. Beginning with the 1999 data analysis, two changes occurred for the age of first use data. First, the NHSDA questionnaire asked year and month of first use for recent initiates. Second, if a respondent was deemed a substance user and did not answer the age of first use question, the response was statistically imputed to eliminate missing data. An exact date of first use, that is the month, day, and year of first use, was then assigned by randomly picking a date within the 365 days corresponding to the respondent's age at first substance use. However, for recent users, certain constraints were placed on the 365-day interval. If the respondent provided a month and/or year of first use, this information was used to limit the interval. For example, if a respondent reported first use in May 1999, then the interval was limited to the days between May 1, 1999, and May 31, 1999. Another limitation occurs when the recency of frequency intersects the 365-day interval. For example, if the respondent reported use in the past year but not in the past 30 days, then the 365-day interval was truncated and the date of first use was not allowed to fall in the past 30 days. By using this date of first use in conjunction with the birth date, the computation of the period of exposure can be determined exactly in terms of whole days.

Having exact dates of birth and first use also allows the determination of person time of exposure during the targeted period t. Let the target time period for measuring incidence be specified in terms of dates; e.g,. for the period 1998, the specification would consist of

$$t = [t_1, t_2) = [Jan\ 1, 1998, Jan\ 1, 1999),$$

a period that includes January 1, 1998, and all days up to but not including January 1, 1999. The target age group can also be defined by a half open interval as $a = [a_1, a_2)$. For example, the age group 12 to 17 would be defined by $a = [12, 18)$ for persons at least age 12, but not yet age 18.

If person i was in age group a during period t, the time and age interval, $L_{t,a,i}$, can then be determined by the intersection

$$L_{t,a,i} = [t_1,t_2) \cap [DOB_i MOB_i YOB_i + a_1, DOB_i MOB_i YOB_i + a_2)$$

where the time of birth in terms is defined as the day (DOB_i), month (MOB_i), and year (YOB_i). Either this intersection will be empty ($L_{t,a,i} = \varnothing$) or it is designated by the half open interval $L_{t,a,i} = [m_{1,i}, m_{2,i})$ where

$$m_{1,i} = Max\{t_1, (DOB_i MOB_i YOB_i + a_1)\}$$

and

$$m_{2,i} = Min\{t_2, (DOB_i MOB_i YOB_i + a_2)\}.$$

The date of first use, $t_{fu,d,i}$, is also expressed as an exact date. An incident of first substance d use by person i in age group a occurs in time t if $t_{fu,d,i} \in [m_{1,i}, m_{2,i})$. The indicator function $I_i(d, a, t)$ used to count an incident of first use is set to 1 when $t_{fu,d,i} \in [m_{1,i}, m_{2,i})$ and to 0 otherwise. The person time exposure measured in years and denoted by $e_i(d,a,t)$ for a person i of age group a depends on the date of first use. If the date of first use precedes the target period ($t_{fu,d,i} < m_{1,i}$), then $e_i(d,a,t) = 0$. If the date of first use occurs after the target period or if person i has never used substance d, then

$$e_i(d,a,t) = \frac{m_{2,i} - m_{1,i}}{365}.$$ [1]

If the date for first use occurs during the target period $L_{t,a,i}$, then

$$e_i(d,a,t) = \frac{t_{fu,d,i} - m_{1,i}}{365}.$$

Note that both $I_i(d,a,t)$ and $e_i(d,a,t)$ are set to zero if the target period $L_{t,a,i}$ is empty; that is, person i is not in age group a during time t. The incidence rate is then estimated as a weighted ratio estimate:

$$IR(d,a,t) = \frac{\sum_i w_i I_i(d,a,t)}{\sum_i w_i e_i(d,,a,t)}$$

where the w_i are the analytic weights. It is important to note that the weighted ratio estimator is the same function that was used in previous years. However, the indicators going into this ratio estimate changed in the 1999 data analysis.

Table 9.2 shows the logic for the new methodology in terms of time of first use, $t_{fu,d,i}$, expressed in days. For those more accustomed to thinking in terms of age at first use (also in days), $a_{fu,d,i}$, the date of birth provides a link between the two: $t_{fu,d,i} = t_{b,i} + a_{fu,d,i}$.

Using the same example that was presented in ***Section 9.1.2***, an individual born on 6/15/1980 who initiated substance use at age 17 on 2/15/1998 would be included in the 12- to 17-year-old age group, at the time of their initiation in 1998. The new methodology accounts for the fact that this person's age does not exactly intersect calendar time in whole years. ***Figure 9.2*** illustrates this point for t = [1/1/1998, 1/1/1999).

The indicators for this case would be: $I(_{d,12\text{-}17,1998}) = 1$; $e(_{d,12\text{-}17,1998}) = 46/365 = 0.126$; $I(_{d,18\text{-}25,1998}) = 0$; $e(_{d,18\text{-}25,1998}) = 0$, and the gray area shows the exposure time.

[1] If year y is a leap year, the denominator is set to 366.

Table 9.2 Indicator Functions Using the New Methodology

Indicator and Value	Conditions (Verbal)	Conditions (Symbolic) in Terms of Intersection Set $L_{t,a,i}$
$I_i(d,a,t) = 1$	Respondent *i* first used substance *d* at age *a* in time interval *t*.	$L_{t,a,i} \neq \varnothing$, and $t_{fu,di,i} \in [m_1,m_2)$
$I_i(d,a,t) = 0$	In all other cases	$L_{t,a,i} \neq \varnothing$, and $t_{fu,d,i} \notin [m_1,m_2)$
$e_i(d,a,t) = 0$	(1) Respondent i's age was not in age interval *a* during any part of period *t*, or (2) Respondent i reported using substance *d* prior to period *t* or prior to age *a*.	*(1)* $L_{t,a,i} \neq \varnothing$, and (2) $t_{fu,d,i} < m_1$
$e_i(d,a,t) = 0$ $\frac{m_2 - m_1}{365}$	(1) Respondent *i*'s age was in age interval *a* during at least some part of period *t* , and (2) Respondent *i* never used substance *d* prior to period *t* and did *not* initiate use during this period.	(1) $L_{t,a,i} = [m_1,m_2)$, and (2) $t_{fu,d,i} \geq m_2$.
$e_i(d,a,t) = 0$ $\frac{t_{fu,d,i} - m_1}{365}$	(1) Respondent *i*'s age was in age interval *a* during at least some part of period *t*, and (2) Respondent *i did* initiate use during period *t*.	(1) $L_{t,a,i} = [m_1,m_2)$, and (2) $t_{fu,d,i} \in m_2)$.

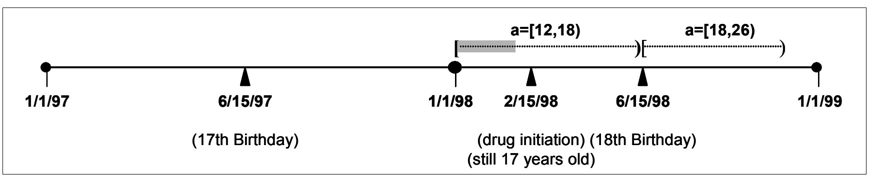

Figure 9.2 Line diagram illustrating the new methodology.

Table 9.3 shows more examples of defining the indicators using the new methodology by presenting examples with specified birth and first use dates.

9.1.4 Impact on Incidence Rate Estimates

Incidence rate estimates are impacted by both the new editing and imputation procedures and the incidence rate calculation method. To sort out the separate impacts of these two changes, age-specific incidence rates

Table 9.3 Examples of the Indicators Using the New Methodology

			t = [1 *Jan* 1998, 1 *Jan* 1999)					
			a = [12, 18)			a = [18, 26)		
i	Date of Birth, $t_{b,i}$	Date of 1st Use, $t_{fu,d,i}$	$L_{t,a,i}$	$I_i(d,a,t)$	$e_i(d,a,t)$	$L_{t,a,i}$	$I_i(d,a,t)$	$e_i(d,a,t)$
1	7/15/80	9/1/98	[1/1/98, 7/15/98)	0	196/365 = 0.54	[7/15/98, 1/1/99)	1	48/365 = 0.13
2	7/15/80	9/1/99	[1/1/98, 7/15/98)	0	196/365 = 0.54	[7/15/98, 1/1/99)	0	170/365 = 0.47
3	7/15/80	Never used	[1/1/98, 7/15/98)	0	196/365 = 0.54	[7/15/98, 1/1/99)	0	170/365 = 0.47
4	7/15/80	1/1/98	[1/1/98, 7/15/98)	1	0/365 = 0.00	[7/15/98, 1/1/99)	0	0
5	7/15/80	9/1/96	[1/1/98, 7/15/98)	0	0	[7/15/98, 1/1/99)	0	0
6	7/15/60	2/1/98	∅	0	0	∅	0	0
7	9/25/72	1/1/89	∅	0	0	[1/1/98, 9/25/98)	0	0
8	9/25/72	1/10/98	∅	0	0	[1/1/98, 9/25/98)	1	10/365 = 0.03
9	9/25/72	1/1/99	∅	0	0	[1/1/98, 9/25/98)	0	268/365 = 0.73

were computed from the 1999 CAI data using three methods for each of the 16 substances for which age of first use data were collected: (1) new methodology using imputed data, (2) new methodology using edited data, and (3) old methodology using imputed data. Sixteen detailed tables were produced showing these estimates; one for each substance. Only 3 of the 16 detailed tables are presented; annual estimates by age group for marijuana, cocaine, and cigarettes are shown in ***Tables 9.4***, ***9.5***, and ***9.6***, respectively. Note that age-specific incidence rates are based on numbers of new initiates per 1,000 years of exposure.

Editing and Imputation Effects. The effect of the new editing and imputation procedures can be evaluated using the new incidence rate calculation method and comparing results in the first two columns based on fully imputed data with the results from the middle two columns based on edited data only. The annual estimates in ***Tables 9.4, 9.5***, and ***9.6*** show mixed results depending on the year.

For marijuana (***Table 9.4***), 11 statistically significant differences occur for age 12 to 17, and 7 of the 11 are higher with imputation; at age 18 to 25, all six significant differences favor the imputed data. The general tendency is for incidence rates based on fully edited and imputed data to be higher

Table 9.4 Marijuana Annual Age-Specific Rates† of First Use (per 1,000 Person-Years of Exposure): 1999

	New Method—Imputed Variable		New Method—Edited Variable		Old Method—Imputed Variable	
Year	12-17	18-25	12-17	18-25	12-17	18-25
1965	9.6	7.5	8.9	6.7	9.1	7.7
1966	19.3	32.1	19.5[b]	32.1	17.9	32.6
1967	20.7	30.8	20.6	30.9	19.1	29.4
1968	20.3	43.5	20.2	41.2	16.3	44.6
1969	34.1	56.7	34.5[b]	56.4	29.2[b]	55.7
1970	55.8	47.7	55.0	46.7	49.1[b]	50.9
1971	46.7	49.7	46.2	48.5	41.8[a]	54.0[a]
1972	59.7	50.2	59.7	47.8[a]	54.3[b]	52.7
1973	59.0	39.4	59.6[a]	36.8	53.3[b]	43.5[a]
1974	66.6	53.0	65.6	52.4	56.6[b]	62.2[b]
1975	65.3	53.7	63.6	53.4	58.3[b]	56.8
1976	76.5	60.5	75.4	60.1	63.2[b]	72.3[b]
1977	86.5	51.2	86.5	50.0	77.9[b]	57.8[a]
1978	84.2	49.6	81.7	48.8	75.0[b]	56.7
1979	86.3	54.5	84.2	54.4	75.7[b]	63.2[b]
1980	75.9	48.2	75.5	44.6	66.6[b]	53.3
1981	51.8	35.1	51.3	35.0	50.1	36.3
1982	59.0	36.3	59.5[b]	35.1	51.9[b]	40.7[a]
1983	55.3	33.1	53.1[a]	32.5	48.5[b]	38.8[b]
1984	58.5	38.5	58.5	38.3	52.1[b]	43.7[b]
1985	58.4	38.7	57.8	37.2	51.8[b]	44.9[b]
1986	53.2	29.8	53.3	28.4[a]	47.6[b]	33.7[a]
1987	56.1	37.3	54.3	36.7	49.8[b]	42.4[b]
1988	55.7	31.6	54.6[a]	31.4	49.4[b]	33.1
1989	46.7	26.8	46.2	25.8	40.2[b]	32.2[b]
1990	48.4	29.0	48.2	28.6	42.5[b]	32.7[a]
1991	46.1	31.9	45.6	31.4	39.5[b]	37.5[b]
1992	51.0	30.5	50.4	29.5[a]	45.7[b]	34.2[b]
1993	60.0	36.7	59.7	36.3	53.4[b]	41.5[b]
1994	74.3	42.1	72.9[b]	41.2	67.2[b]	47.1[b]
1995	78.3	46.1	76.7[b]	45.6	70.8[b]	53.1[b]
1996	89.9	44.1	87.7[b]	42.7[b]	80.0[b]	52.8[b]
1997	90.0	45.1	87.3[b]	44.5[a]	79.6[b]	53.8[b]
1998	82.6	46.5	79.2[b]	45.6[b]	73.5[b]	54.5[b]

† The numerator of each rate is the number of persons in the age group who first used the substance in the year, while the denominator is the person time exposure measured in thousands of years.
[a] Difference between the estimate and New Method—Imputed is statistically significant at the .05 level.
[b] Difference between the estimate and New Method—Imputed is statistically significant at the .01 level.

Table 9.5 Cocaine Annual Age-Specific Rates† of First Use (per 1,000 Person-Years of Exposure): 1999

	New Method—Imputed Variable		New Method—Edited Variable		Old Method—Imputed Variable	
Year	12-17	18-25	12-17	18-25	12-17	18-25
1965	—‡	0.55	—‡	0.55	—‡	0.53
1966	0.11	1.49	0.11	1.50	—‡	0.86
1967	0.33	2.72	0.33	2.72	0.33	2.63
1968	1.57	4.26	1.58	4.20	1.56	4.19
1969	1.86	3.37	1.87	3.39	1.83	3.24
1970	1.99	6.82	2.01	6.85[a]	1.66	6.70
1971	3.77	6.92	3.80	6.95	2.23	8.07
1972	2.56	9.13	2.58	9.08	2.50	9.03
1973	5.18	7.31	5.23[a]	7.34[a]	3.34	8.77
1974	4.58	16.32	4.62[a]	15.55	2.89	17.43
1975	7.06	18.44	7.13[b]	18.25	4.84[a]	18.40
1976	6.24	19.04	5.94	18.31	4.41	19.65
1977	13.01	22.12	12.56	21.87	10.97[a]	21.60
1978	8.28	21.19	7.31	21.27	6.76	20.22
1979	7.51	24.03	7.54[a]	23.36	5.65	23.65
1980	7.44	22.96	7.04	21.66	6.00	22.89
1981	8.44	19.29	8.48[b]	18.60	6.77	19.70
1982	11.71	24.00	11.76[b]	22.67	8.86[b]	24.79
1983	12.37	31.86	11.82	30.78	9.59[a]	31.97
1984	10.55	27.39	10.10	27.42	9.17[a]	28.36
1985	10.42	22.94	10.45[b]	22.41	9.21	23.51
1986	12.44	16.26	12.47[b]	15.63	8.68[b]	18.66[a]
1987	11.71	19.90	11.56	19.01	9.84[b]	19.26
1988	11.93	19.04	11.37	18.43	11.36	16.64
1989	10.23	15.24	10.19	14.84	7.27[b]	15.00
1990	7.65	15.33	7.64	15.29	5.31[b]	15.98
1991	5.64	12.54	5.65[b]	12.55[a]	4.00[b]	12.57
1992	5.63	11.47	5.60	10.96	4.57[b]	11.31
1993	7.45	12.13	7.38	11.93	6.50[b]	12.42
1994	6.97	9.25	6.79	9.15	5.74[b]	9.58
1995	9.04	16.78	8.86	16.13[a]	7.46[b]	17.72[a]
1996	12.59	13.97	12.07[a]	13.46[a]	10.63[b]	15.68[b]
1997	14.07	18.92	13.40[b]	18.39[a]	11.62[b]	20.87[b]
1998	14.35	19.24	13.22[b]	18.91[b]	11.23[b]	21.39[b]

† The numerator of each rate is the number of persons in the age group who first used the substance in the year, while the denominator is the person time exposure measured in thousands of years.

‡ —Low precision; no estimate reported.

[a] Difference between the estimate and New Method—Imputed is statistically significant at the .05 level.

[b] Difference between the estimate and New Method—Imputed is statistically significant at the .01 level.

Table 9.6 Cigarette Annual Age-Specific Rates† of First Use (per 1,000 Person-Years of Exposure): 1999

Year	New Method—Imputed Variable		New Method—Edited Variable		Old Method—Imputed Variable	
	12-17	18-25	12-17	18-25	12-17	18-25
1965	119.0	63.0	121.7[a]	62.3	107.3[b]	68.3
1966	104.7	72.0	100.8	72.7	96.9[b]	74.8
1967	120.4	99.5	120.2	95.0	110.7[b]	99.0
1968	109.1	79.1	108.3	72.6	100.0[b]	83.2
1969	101.4	69.9	100.0	70.7	96.5	63.4
1970	121.0	93.9	121.4	91.3	109.2[b]	101.6
1971	121.1	74.4	118.9	73.1	111.7[b]	78.5
1972	121.8	89.4	119.1	81.5	111.8[b]	92.7
1973	131.2	80.8	130.8	80.2	117.1[b]	95.1[a]
1974	131.1	73.7	127.7	71.2	121.1[b]	76.0
1975	136.5	72.9	135.3	68.1	121.8[b]	82.0
1976	151.5	77.1	148.0	75.2	143.0[a]	84.4
1977	128.0	57.4	124.3	57.4	118.5[b]	61.4
1978	145.6	64.6	142.3[a]	63.7	134.8[b]	70.7
1979	124.7	74.2	125.3	74.4	116.2[b]	78.0
1980	104.3	74.6	103.3	71.7	98.9[b]	74.9
1981	105.4	55.1	103.9	55.2	97.8[b]	58.5
1982	108.5	55.6	106.1[a]	54.6	97.0[b]	63.1[b]
1983	109.1	55.8	109.5	53.2	97.8[b]	65.2[b]
1984	111.8	59.2	108.5[b]	59.2	105.5[b]	62.4
1985	122.0	39.8	120.6	38.7	114.7[b]	41.8
1986	116.6	50.7	116.3	49.6	106.2[b]	58.1[b]
1987	117.9	50.1	116.1	50.3[b]	107.4[b]	59.4[b]
1988	111.7	48.0	109.7	46.7	104.0[b]	53.8[b]
1989	111.6	56.7	110.7	55.7	97.6[b]	68.9[b]
1990	108.9	46.8	108.7	46.4	100.4[b]	52.3[a]
1991	101.5	53.8	99.7[b]	52.5	91.8[b]	59.8[b]
1992	112.3	54.9	112.1	53.4	102.1[b]	60.9[b]
1993	115.0	52.5	113.6[a]	52.0	106.4[b]	57.9[b]
1994	137.8	71.4	136.0[b]	70.7	126.4[b]	78.1[b]
1995	143.5	67.9	140.7[b]	66.1[a]	131.4[b]	80.5[b]
1996	154.0	78.3	151.7[b]	76.7[a]	142.3[b]	89.4[b]
1997	157.6	65.3	152.5[b]	64.4[a]	141.8[b]	78.1[b]
1998	138.7	66.5	132.8[b]	64.9[a]	125.2[b]	77.0[b]

† The numerator of each rate is the number of persons in the age group who first used the substance in the year, while the denominator is the person time exposure measured in thousands of years.

a Difference between the estimate and New Method—Imputed is statistically significant at the .05 level.

b Difference between the estimate and New Method—Imputed is statistically significant at the .01 level.

than those based on the edit-only approach. This is confirmed in a summary table (***Table 9.7)*** that averages the annual estimates for all 16 substances and then computes the relative impact of the new editing and imputation procedure relative to the old method when both use the new incidence rate computation method.[2] The averages for marijuana at age 12 to 17 are 58.3 with imputation and 57.4 with editing only, a relative increase of 1.5 percent. For persons aged 18 to 25, the averages are 40.8 with imputation and 39.9 with editing only, a relative increase of 2.3 percent.

Table 9.7 Average (1965-1998) Imputation and Editing Effects on Incidence Rate Estimates by Age Group: 1999 CAI with New Incidence Rate Method

	Average Incidence Rate with Imputation		Average Incidence Rate with Editing Only		Relative Imputation and Editing Effect, %	
Substance	**12-17**	**18-25**	**12-17**	**18-25**	**12-17**	**18-25**
Marijuana	58.3	40.8	57.4	39.9	1.47	2.28
Cocaine	7.7	15.1	7.5	14.7	2.41	2.50
Crack	1.6	3.1	1.5	3.0	2.69	3.12
Heroin	0.9	1.6	0.9	1.6	2.55	0.81
Hallucinogens	13.7	13.1	13.4	12.7	2.48	3.38
Inhalants	11.9	5.5	11.0	5.2	8.02	5.80
Pain relievers	7.8	7.9	7.1	7.1	10.12	11.05
Tranquilizers	4.9	6.0	4.6	5.8	5.83	4.78
Stimulants	7.4	7.9	6.9	7.5	6.54	6.30
Methamphetamines	4.0	5.1	3.8	4.8	5.28	5.72
Sedatives	3.3	3.6	3.2	3.5	3.13	4.30
Alcohol	129.6	199.8	128.5	196.8	0.86	1.52
Cigarettes	122.2	66.0	120.5	64.5	1.43	2.44
Daily cigarettes	40.7	38.9	40.2	37.5	1.38	3.81
Smokeless tobacco	24.6	10.3	23.5	9.8	4.57	5.20
Cigars	27.2	36.4	26.2	34.8	3.85	4.54

For cocaine (***Table 9.5***), 19 statistically significant differences are identified across the two age groups. Of these, 12 show higher estimates for edited data and seven for imputed data. In the most recent years, the significant differences are larger and favor the fully imputed data; the significant differences favoring imputed data tend to be very small. The NHSDA design oversamples persons who are currently age 25 or younger. These oversampled persons reached an age of 12 in 1985 or later; data for substance initiation prior to 1985 must therefore be based entirely on

[2]Suppressed estimates were excluded from the averages.

persons who were currently (in 1999) aged 26 to 59.[3] Since persons aged 26 or older in 1999 were sampled at a lower rate, the number of cases contributing to the numerator estimates (initiations) and to the denominator estimates (person time at risk) for the age-specific incidence rates decreases rapidly from 1999 to 1985. Prior to 1971, the numerator estimate for cocaine incidence rates was based on five or fewer sample cases; in 1993 and later years, the numerator estimate was based on 100 or more sample cases. Because the data for the most recent years are based on estimates that benefit from larger samples, the estimates from the most recent years should be better indicators of any impact of methodological change. The multiyear averages for cocaine shown in ***Table 9.7*** show small increases in incidence rates due to imputation for both age groups (about 2.5 percent for both age groups).

Because incidence rates for cigarettes (Table 9.6) are much higher, the estimates do not suffer from small sample sizes nearly as much as the cocaine estimates. Of 16 differences that tested as statistically significant across the two age groups, 14 show higher incidence rate estimates for imputed data. The multiyear (1965-1998) averages shown in ***Table 9.7*** confirm the positive impact of imputation for both age groups with relative increases of 1.4 percent at age 12 to 17 and 2.4 percent at age 18 to 25.

Incidence rates based on fully edited and imputed data tend to be higher than those based on the edit-only approach for marijuana, cocaine, and cigarettes. This is confirmed in the ***Table 9.7*** summary for all 16 substances. While the relative increases in incidence rates are usually small, they appear to be consistent across all 16 substances studied. This increase was expected because previously respondents with missing age of first use data were dropped from the analysis even though they were known substance users. The largest relative increases in incidence rates due to the new editing and imputation rules occurred among some of the substance classes that are made up of a large number of different substances (i.e., pain relievers, inhalants, stimulants, tranquilizers, and methamphetamines). Because of the many separate probes used in determining use of these classes of substances, it may be that respondents were less sure about how to answer the age-at-first-use question resulting in more missing data.

Effects of the Method of Incidence Rate Calculation. A second set of comparisons looks at the differences between the old and new method

[3]Persons 60 or older in 1999 could not have initiated substance use while 12 to 25 in 1965 or later.

of calculating age-specific rates using imputed data in both methods. This comparison illustrates the difference in the two calculation methodologies holding the editing and imputation constant (at the fully edited and imputed level). These differences based on 1999 CAI data are shown by years for three substances in ***Tables 9.4***, ***9.5,*** and ***9.6*** by comparing the first two columns (New Method—Imputed Variables) with the last two columns (Old Method—Imputed Variables).

For marijuana at age 12 to 17, the incidence rate is higher for the new methodology for all 34 years (1965-1998), and the difference is statistically significant in 29 years. The age 18 to 25 incidence rate is lower for the new methodology in 32 of the 34 years, and the difference is significantly lower in 22 of those years. For many substances, including marijuana, the mean age of first use is around the ages of 17 to 18,[4] which is the borderline between the two computed incidence age groups. The new methodology removes some of these borderline cases from the calculation of the 18 to 25 age-specific rates and correctly places them into the calculation of the 12 to 17 age-specific rates. While both the numerator (new initiates) and the denominator (exposure time) are influenced by the change in method, the main impact is through the classification of initiates by age group in the numerator. Under the new method, new initiates are assigned to an age group based on their attained age at the date of initiation. Under the old method, new initiates were assigned to an age group based on their age at their birthday during the current year (see ***Tables 9.1*** and ***9.2***). Much of the marijuana initiation occurs near age 17; the average age of initiation for marijuana is about 18 (***Table 9.8***). Under the old methodology, many of the 17-year-old initiates were being counted in the 18 to 25 age-specific rate. However, the new methodology places them in the correct age group based on their attained age at the date of initiation. For marijuana, this results in an increase of almost 13 percent for incidence rates at age 12 to 17 and a decrease of over 10 percent at age 18 to 25.

Cocaine incidence rates show a similar pattern. Fewer statistically significant differences are noted due to changing to the new method: 17 significant increases at age 12 to 17 and 5 significant decreases at age 18 to 25. The changes in the multiyear averages (***Table 9.8***) show a 22 percent increase at age 12 to 17 and about a 2 percent decrease at age 18 to 25. The average age of initiation for cocaine is 21, which would indicate that many

---

[4]Mean age of first use is another estimate that measures substance use trends and is categorized by the year of first use.

Table 9.8 Average (1965-1998) Method of Computation Effect on Incidence Rate Estimates: 1999 CAI with Fully Edited and Imputed Data

Substance	Mean AFU†	Average Incidence Rate New Method		Average Incidence Rate Old Method		Relative Method of Computation Effect, %	
		12-17	18-25	12-17	18-25	12-17	18-25
Marijuana	18.1	58.3	40.8	51.7	45.5	12.79	-10.28
Cocaine	21.1	7.7	15.1	6.3	15.4	21.69	-2.18
Crack	25.2	1.6	3.1	1.4	2.7	14.13	12.40
Heroin	22.1	0.9	1.6	0.8	1.6	19.16	-0.58
Hallucinogens	18.8	13.7	13.1	11.4	14.6	20.47	-10.19
Inhalants	17.1	11.9	5.5	11.1	5.9	7.75	-6.71
Pain relievers	21.2	7.8	7.9	7.2	8.1	9.22	-2.00
Tranquilizers	22.6	4.9	6.0	4.4	6.0	12.26	0.17
Stimulants	19.0	7.4	7.9	6.5	8.2	12.91	-3.36
Methamphetamines	19.3	4.0	5.1	3.3	5.4	21.38	-5.66
Sedatives	20.6	3.3	3.6	2.9	3.9	12.63	-5.93
Alcohol	16.9	129.6	199.8	114.1	203.7	13.60	-1.90
Cigarettes	15.4	122.2	66.0	112.1	72.0	9.03	-8.34
Daily cigarettes	18.3	40.7	38.9	35.7	41.5	13.90	-6.22
Smokeless tobacco	17.3	24.6	10.3	23.3	11.0	5.51	-6.67
Cigars	20.7	27.2	36.4	24.5	36.3	11.02	0.20

†AFU = Age of first use.

persons initiate their use of cocaine (especially crack cocaine) at higher ages. The increase at age 12 to 17 can be associated with correctly classifying persons in the 12 to 17 age group by their date of initiation when the old method would have counted them as being age 18. The smaller decrease in incidence rates for age 18 to 25 is most likely due to the repetition of the same classification error correction for persons at age 25 who would have been treated as being age 26 following the old method. The combined effects of reducing the number of 18 to 25 initiates at the 17 to 18 boundary is partially offset by the increase in the number of age 18 to 25 initiates at the 25 to 26 boundary.

Cigarette incidence rates behave very similarly to the marijuana rates as a result of the change in the rate calculation method showing 33 statistically significant increases at age 12 to 17 and 16 statistically significant differences at 18 to 25. ***Table 9.8*** shows the average increase across years for age 12 to 17 is about 9 percent and the average decrease for age 18 to 25 is about 8 percent.

Multiyear averages (1965-1998) are compared for the two methods for all 16 substances in ***Table 9.8***. ***Table 9.8*** also shows the mean age of first use (AFU) averaged over all the years for each substance. In general, the different substances follow patterns of change due to the method of incidence rate computation similar to those for marijuana, cocaine, and cigarettes; the one exception is crack cocaine, which shows increased incidence rates both at age 12 to 17 and age 18 to 25. Crack cocaine also exhibits the highest average age of initiation at 25 years of age. In this case, more reclassification occurs at the 25 to 26 boundary than at the 17 to 18 boundary, resulting in increases in incidence rates for both age groups.

9.1.5 Impact on Other Initiation of Use Measures

In addition to incidence rate estimates, SAMHSA routinely reports trends in the number of new initiates and mean age for first use for each year. These estimates are not affected by the method of calculation used for incidence rates, but may be influenced by the change in the editing and imputation procedures. ***Tables 9.9, 9.10,*** and ***9.11*** show the impact of the new editing and imputation for 1999 CAI data on the annual number of initiates and the mean age of first use for marijuana, cocaine, and cigarettes. Estimates based on 1994-1998 combined PAPI data are also presented in these tables as an indication of the overall impact of interview mode and revised editing and imputation procedures.[5]

Comparisons of the estimated number of initiates based on edited vs. imputed 1999 CAI show an increase for the imputed data for all three substances: 26 significant differences showing higher estimates from imputed data for marijuana, nine for cocaine, and 31 for cigarettes. ***Table 9.12*** shows that multiyear average numbers of initiates increased 2.4 percent for marijuana, 3.4 percent for cocaine, and 2.9 percent for cigarettes with the use of the new editing and imputing procedures. Comparisons of the estimated number of initiates based on 1994-1998 PAPI data with editing only and the estimated number of initiates based on 1999 CAI data with imputation show only four significant differences over the three substances with three having higher 1999 imputed estimates.

---

[5]This comparison is partially confounded with respondent recall effects for surveys conducted in different years. Sample sizes for 1999 PAPI data were not adequate to permit a meaningful comparison with 1999 CAI. Note also that, for each year beginning with 1994, only initiation prior to that year could be estimated using the PAPI data.

Table 9.9 Comparison of Number of Marijuana Initiates (in Thousands) and Mean Age at First Marijuana Use, 1994-1998 PAPI Versus 1999 CAI

	Initiates (1,000s)			Mean Age		
Year	1994-1998 PAPI	1999 CAI Edited Data	1999 CAI Imputed Data	1994-1998 PAPI	1999 CAI Edited Data	1999 CAI Imputed Data
1965	601	442	478	18.95	21.77	21.61
1966	977	1,229	1,234	20.05	18.68	18.68
1967	1,423	1,199	1,210	19.76	18.92	18.91
1968	1,621	1,470	1,533	18.97	18.89	18.91
1969	2,245	2,301	2,317	19.19	19.43	19.43
1970	2,611	2,501[a]	2,585	19.21	18.20	18.34
1971	2,710	2,403	2,456	18.78	17.87	17.84
1972	2,861	2,676[b]	2,747	18.62	18.16[a]	18.24
1973	2,897	2,610[a]	2,697	18.28	19.03	19.15
1974	2,966	2,873[b]	2,938	18.50	17.85	17.82
1975	3,128	2,923[a]	2,989	18.51	18.90[a]	18.84
1976	2,786	3,216	3,267	18.69	18.38	18.34
1977	2,889	3,195[a]	3,251	18.95	18.03	18.07
1978	2,846	2,959[a]	3,046	17.77	18.14	18.10
1979	2,654	2,983[b]	3,052	18.22	18.22	18.18
1980	2,499	2,564[a]	2,680	18.41	18.13	18.38
1981	2,115	1,820[a]	1,840	17.94	18.29	18.26
1982	1,964	2,056[a]	2,090	18.19	18.85	18.98
1983	2,143	1,889[b]	1,954	17.85	18.90[a]	18.77
1984	2,010	2,029	2,040	19.19	18.56	18.54
1985	1,775	1,890[a]	1,938	17.85	18.05	18.03
1986	1,845	1,604[b]	1,633	19.32	17.15	17.18
1987	1,756	1,708[b]	1,763	17.92	17.48	17.45
1988	1,565	1,595[a]	1,620	17.49	17.49[a]	17.47
1989	1,371	1,353[a]	1,388	17.87	17.21	17.29
1990	1,423	1,451[a]	1,470	17.66	17.42	17.44
1991	1,415	1,519[b]	1,545	17.47	17.78	17.76
1992	1,644	1,544[b]	1,578	17.60[b]	16.49	16.49
1993	1,983	1,930[b]	1,972	16.96	17.33	17.45
1994	2,380	2,235[b]	2,293	16.90	16.61	16.59
1995	2,409	2,359[b]	2,421	16.57	16.65[b]	16.59
1996	2,462	2,532[b]	2,616	16.62	17.24[b]	17.18
1997	2,114[b]	2,493[b]	2,571	17.09	17.16[b]	17.08
1998	NA[†]	2,345[b]	2,437	NA[†]	17.63[b]	17.52

[†]NA = not available.
[a]Difference between the estimate and 1999 CAI Imputed is statistically significant at the .05 level.
[b]Difference between the estimate and 1999 CAI Imputed is statistically significant at the .01 level.

The impact of the 1999 editing and imputation procedures on estimates of average age of first use are small and mixed. Comparisons against edited 1999 CAI data show one significant difference with higher imputed data and seven with higher edited data for marijuana, three with higher edited data for cocaine, and eight with higher edited data for cigarettes. Multiyear

Table 9.10 Comparison of Number of Cocaine Initiates (in Thousands) and Mean Age at First Cocaine Use, 1994-1998 PAPI Versus 1999 CAI

	Initiates (1,000s)			Mean Age		
Year	1994-1998 PAPI	1999 CAI Edited Data	1999 CAI Imputed Data	1994-1998 PAPI	1999 CAI Edited Data	1999 CAI Imputed Data
1965	—†	—†	—†	—†	—†	—†
1966	—†	—†	—†	23.62	—†	—†
1967	33	—†	—†	—†	—†	—†
1968	82	147	148	20.91	18.77	18.81
1969	211	136	136	19.20	18.37	18.37
1970	319	257	257	19.54	20.52	20.52
1971	343	325	325	19.21	20.12	20.12
1972	260	370	372	19.29	19.79	19.79
1973	485	370	381	20.12	19.60	19.90
1974	667	689	714	21.07	22.40	22.39
1975	790	838	847	21.37	21.13	21.11
1976	698	861	909	21.32	21.13	21.25
1977	915	1,130	1,156	21.57[a]	20.18	20.14
1978	1,026	923	997	21.10	20.87	20.99
1979	1,026	1,135	1,161	21.87	21.87	21.83
1980	1,301	1,013[a]	1,105	21.27	21.51	21.73
1981	1,375[b]	911	937	22.28	20.81	20.87
1982	1,413	1,130[a]	1,175	21.75	20.94	21.00
1983	1,169[b]	1,671[a]	1,760	21.99	22.91	23.14
1984	1,208	1,240	1,259	21.86	21.29	21.30
1985	1,182	1,139	1,156	22.44	22.22	22.20
1986	1,156	886	905	22.95	21.79	21.78
1987	1,012	1,040[a]	1,069	23.06	22.19	22.18
1988	767	856[a]	883	21.43	21.39	21.33
1989	752	755	783	22.36	22.47	22.55
1990	605	698	726	22.42	21.89	22.17
1991	469	544	544	21.45	21.91	21.91
1992	489	604	619	21.62[a]	24.08	24.09
1993	551	602	614	22.30	21.69	21.74
1994	542	464	496	21.43	20.74	21.36
1995	655	691[b]	712	19.63	20.40	20.40
1996	670	725[b]	751	18.97	21.12[a]	21.00
1997	730	825[b]	855	20.27	19.34[b]	19.25
1998	NA‡	883[b]	918	NA‡	20.44[b]	20.27

† — Low precision; no estimate reported.
‡ NA = not available.
[a] Difference between the estimate and CAI Imputed is statistically significant at the .05 level.
[b] Difference between the estimate and CAI Imputed is statistically significant at the .01 level.

averages shown in ***Table 9.12*** show a 0.02 percent increase in average age of first use for marijuana, a 0.26 percent increase for cocaine, and a 0.12 percent decrease for cigarettes. In general, the relative impact of the 1999 imputation procedures on estimated average age of first use is small relative to the impact on estimates of initiates or of incidence rates.

Table 9.11 Comparison of Number of Cigarette Initiates (in Thousands) and Mean Age at First Cigarette Use, 1994-1998 PAPI Versus 1999 CAI

	Initiates (1,000s)			Mean Age		
Year	1994-1998 PAPI	1999 CAI Edited Data	1999 CAI Imputed Data	1994-1998 PAPI	1999 CAI Edited Data	1999 CAI Imputed Data
1965	3,085	2,668	2,701	15.98	15.60	15.55
1966	2,859	2,461[a]	2,563	16.05	15.48	15.45
1967	3,263	3,184	3,275	15.76	15.66	15.70
1968	3,219	2,668[a]	2,770	15.34	15.03	15.09
1969	3,336	2,767[a]	2,844	15.37	15.05	15.00
1970	3,471	3,317[a]	3,380	15.60	15.11	15.17
1971	3,350	3,112[b]	3,218	15.06	15.37	15.33
1972	3,719	3,326[b]	3,500	15.26	15.68	15.73
1973	3,384	3,533[a]	3,597	15.29	15.13	15.13
1974	3,560	3,419[b]	3,564	15.12	15.23	15.24
1975	3,600	3,465[a]	3,588	15.34	15.00	15.07
1976	3,442	3,808[b]	3,941	15.33	15.94	15.89
1977	3,451	2,993[b]	3,101	15.56	15.12	15.07
1978	3,114	3,374[b]	3,472	15.63	15.62	15.60
1979	3,015	3,123[a]	3,139	15.72	15.73	15.72
1980	2,808	2,590[b]	2,655	15.57	15.70	15.71
1981	2,750	2,623[b]	2,697	15.71	16.79	16.81
1982	2,689	2,510[b]	2,596	15.51	15.35	15.32
1983	2,719	2,537[a]	2,580	15.29	15.29	15.32
1984	2,690	2,615[b]	2,691	15.58	15.60[a]	15.56
1985	2,749	2,467[b]	2,528	15.57	15.22	15.19
1986	2,761	2,573[b]	2,614	15.52	16.00	16.00
1987	2,548	2,431[b]	2,477	16.11[a]	15.31[a]	15.28
1988	2,493	2,360[b]	2,421	15.41	15.79	15.76
1989	2,521	2,456[b]	2,513	16.07	15.33	15.33
1990	2,575	2,284[b]	2,324	15.43[a]	15.03	14.99
1991	2,505	2,309[b]	2,388	15.87	15.37[a]	15.31
1992	2,678	2,616[b]	2,662	15.66[a]	15.09	15.10
1993	2,845	2,618[b]	2,693	15.88[a]	15.15[b]	15.09
1994	3,154	3,207[b]	3,291	15.86	15.49[b]	15.43
1995	3,293	3,227[b]	3,338	15.38	15.56[b]	15.50
1996	3,108[a]	3,410[b]	3,508	15.77	15.51[b]	15.46
1997	NA†	3,038[b]	3,169	NA†	15.47[b]	15.41
1998	NA†	NA†	NA†	NA†	NA†	NA†

† NA = not available.
[a] Difference between the estimate and CAI Imputed is statistically significant at the .05 level.
[b] Difference between the estimate and CAI Imputed is statistically significant at the .01 level.

Comparisons of the average age of first use based on 1994-1998 PAPI data with editing only and the average age for first use based on 1999 CAI data with imputation show only seven significant differences, with six higher estimates for 1994-1998 data with editing only and one higher estimate with the fully imputed 1999 CAI data. With so few significant differences and no correction for multiple comparisons, there is little evidence for

Table 9.12 Average (1965-1998) Imputation and Editing Effects on Estimates of Number of Initiates and Average Age of First Use: 1999 CAI

	Imputed and Edited Data		Edited Data		Relative Effects, %	
Substance	Initiates	AFU†	Initiates	AFU	Initiates	AFU
Marijuana	2,166	18.1	2,115	18.1	2.44	0.02
Cocaine	796	21.1	770	21.1	3.40	0.26
Crack	273	25.2	274	25.1	-0.26	0.40
Heroin	101	22.1	97	22.3	3.93	-1.04
Hallucinogens	736	18.8	713	18.8	3.20	0.08
Inhalants	503	17.1	465	17.1	8.29	-0.16
Pain relievers	553	21.2	491	21.0	12.56	0.80
Tranquilizers	411	22.6	386	22.5	6.41	0.43
Stimulants	440	19.0	413	19.1	6.56	-0.49
Methamphetamines	289	19.3	272	19.3	5.92	-0.02
Sedatives	219	20.6	209	20.6	4.98	-0.14
Alcohol	3,764	16.9	3,645	16.9	3.25	0.05
Cigarettes	2,964	15.4	2,881	15.5	2.85	-0.12
Daily cigarettes	1,760	18.3	1,696	18.2	3.79	0.44
Smokeless tobacco	958	17.3	907	17.3	5.62	0.07
Cigars	1,755	20.7	1,654	20.6	6.06	0.43

†AFU = Age of first use.

concluding any differences between the 1994-1998 PAPI data and the 1999 CAI data with respect to average age of first use.

A summary of the editing and imputation impact on these two measures is shown in ***Table 9.12*** for 16 substances. The pattern of relative impacts on estimates of the number of new initiates is very similar to relative impacts of editing and imputation on incidence rate estimates shown in ***Table 9.7***. For most substances, the new editing and imputation methods increase the estimated number of new initiates from 3 to 12 percent. The one exception is for crack, which shows a very small decrease. As with the incidence rates, the largest increases occur for substance classes that include a large number of different substances such as pain relievers, inhalants, stimulants, and tranquilizers.

The impact of the new editing and imputation procedures on the estimates of average age of first use is mixed and usually small. The largest impact is a little over 1 percent decrease in average age of first use for heroin; this difference is still only about 0.2 years.

9.2 Impact of Questionnaire Mode Change on Estimates of Substance Use Initiation

Did switching from the PAPI to the CAI questionnaire in 1999 impact the incidence rate estimates? In ***Chapter*** *7*, Chromy et al. discuss the reasons that changes in questionnaire mode could affect the substance estimates, including the fact that the CAI instrument allows for more internal consistency and more complete responses. In addition, the format of the CAI questionnaire gives the respondent more privacy when answering sensitive questions. However, ***Chapter*** *7* investigates mode effect only on recency-of-use data, not age-of-first-use data. Therefore, the impact of mode on the incidence rate estimates is investigated to a limited extent in this chapter.

Table 9.13 shows the missing data patterns observed for marijuana, cocaine, and cigarettes in advance of any editing or imputation. PAPI data are based on 1994-1998 data in order to provide more adequate sample sizes for rare measures. The "Inconsistent Data" columns refer to inconsistency that resulted in editing or imputation of the age-of-first-use variable; if inconsistencies were resolved by editing or imputing other variables (e.g., recency or frequency of use), they were excluded from the counts in order to provide more comparable basic data for comparing PAPI to CAI. Nevertheless, the 1999 CAI data show a higher level of inconsistent data, probably reflecting the more comprehensive editing for inconsistencies within the whole substance module employed with the 1999 CAI data. This increased rigor in the edit process produced an increase in inconsistencies in spite of the programmed consistency checks within the CAI instrument. Additional questions regarding year and month of first use were also added to the 1999 CAI instrument. The programmed consistency checks in the CAI instrument were primarily concerned with inconsistency among reported age of first use, reported date of birth, and date of interview. Checks for consistency with reported recency and frequency of use were implemented in the post-data-collection processing as part of the effort to produce an imputed exact age of first use consistent with all other reported (or imputed) variables in the substance module.

The combined Don't Know, Refused, and blank responses reflect the totally missing data. Blank responses to the age-at-first-use question on the CAI instrument should not occur for persons who report at least lifetime substance use and complete the entire module since they only have the options to report an age or give a Don't Know or Refused response. The

Table 9.13 Weighted Percentages of Different Types of Missing Patterns of Age of First Use Data for Lifetime Substance Users

	Missing Data Patterns, %							
	Inconsistent Data		Don't Know		Refused		Blank	
Substance	1994-1998 PAPI	1999 CAI	1994-1998 PAPI	1999 CAI	1994-1998 PAPI	1999 CAI	1994-1998 PAPI	1999 CAI
Marijuana	0.19	0.54	0.10	1.17	0.01	0.56	0.14	0.06
Cocaine	0.35	0.58	0.06	1.91	0.00	0.67	0.34	0.22
Cigarettes	0.17	0.42	0.11	2.53	0.01	0.19	0.34	0.00

small proportion with blank age-at-first-use responses is most likely the result of Refused or Don't Know responses to the substance module gate question, which subsequently were imputed as at least lifetime users. The higher proportions of CAI respondents with Don't Know or Refused responses is somewhat surprising and far outweighs the opposite relationship for blank responses. The largest difference occurs within the Don't Know category. One can only speculate that once respondents became familiar with the option to choose Don't Know or Refused, they used this option rather than guessing.

Tables 9.14, 9.15, and ***9.16*** display the 12 to 17 and 18 to 25 age-specific incidence rates for PAPI and CAI for marijuana, cocaine, and cigarettes. Annual estimates are provided for PAPI combined data for 1994-1998, for 1999 PAPI data,[6] and for 1999 CAI data. For both PAPI and CAI, edited data and the old incidence methodology were used to compute these estimates. For comparability, non-imputed edited data were used since the PAPI data do not have imputed versions of the age of first use. In addition, significance for the difference in rates between CAI and PAPI is shown in these tables. Only a small number of years for each substance show a significant difference between the CAI and PAPI estimates.

For marijuana (***Table 9.14***), only 14 individual year mode effect estimates are statistically significant and their directions are mixed; six have higher PAPI annual incidence rates and eight have higher CAI incidence rates.

[6]The weights applied to the PAPI analysis are the initially computed and calibrated weights without any adjustment to match the distribution of field interviewer experience to prior years.

Table 9.14 Comparison of Mode Effect: Marijuana Annual Age-Specific Rates of First Use (per 1,000 Person-Years of Exposure), PAPI and CAI Data

	PAPI 1994-1998		PAPI 1999 Old Method—Edited		CAI 1999 Old Method—Edited	
Year	12-17	18-25	12-17	18-25	12-17	18-25
1965	8.7	13.7[a]	11.8	13.8	8.4	6.9
1966	13.9	23.5	7.4[a]	38.8	18.1	32.6
1967	15.6	38.8	18.4	36.7	18.9	29.5
1968	20.1	45.2	35.0[a]	47.0	16.2	42.5
1969	31.7	54.1	29.0	50.9	29.5	55.3
1970	35.1	64.3	27.1[a]	50.1	48.3	50.5
1971	40.8	65.9	41.6	49.3	41.4	52.8
1972	48.4	64.1	44.3	69.4	54.5	50.5
1973	60.2	57.7[a]	52.4	68.0	54.1	40.8
1974	57.6	61.7	68.7	58.0	55.4	61.7
1975	67.8	57.8	48.7	43.7	56.7	56.4
1976	59.5	52.4[a]	74.0	61.0	62.2	72.1
1977	66.7	50.2	34.5[b]	58.5	77.8	56.8
1978	75.2	49.9	62.2	61.1	74.9	53.1
1979	60.6	59.0	63.5	45.6	73.7	62.9
1980	59.2	56.0	58.0	47.9	66.4	49.9
1981	54.3	43.1	54.4	69.2[a]	49.6	36.3
1982	48.2	42.3	55.8	35.5	52.2	39.6
1983	56.4[a]	45.1	52.1	39.9	46.3	38.2
1984	53.1	38.4	54.3	28.4[a]	52.2	43.6
1985	48.8	38.6	52.6	37.9	51.2	43.4
1986	48.4	41.3[a]	38.4	43.4	47.7	32.3
1987	48.4	40.5	41.6	29.8	48.1	41.7
1988	44.9	36.9	39.8	29.3	48.5	32.9
1989	37.0	32.8	38.1	40.4	40.0	31.1
1990	36.9	36.6	38.9	33.8	42.3	32.3
1992	44.5	37.0	44.7	25.4	45.2	33.5
1993	55.1	45.9	65.3	39.1	53.1	41.2
1994	72.8	47.9	76.7	41.7	66.0	46.1
1995	74.1	52.6	72.5	47.7	69.1	52.5
1996	79.3	52.1	56.3[b]	56.8	77.9	51.3
1997	64.4[a]	47.1	84.4	57.5	76.9	53.2
1998	NA[†]	NA[†]	55.2[a]	53.5	70.2	53.2

[†]NA = not available.
[a] Difference between PAPI and CAI is statistically significant at the .05 level.
[b] Difference between PAPI and CAI is statistically significant at the .01 level.

For cocaine (***Table 9.15***), even fewer significant differences in annual estimates are noted; two have higher PAPI estimates, and five have higher CAI estimates. No corrections for multiple comparisons were made in the significance tests in these tables, so it is not appropriate to make judgments about an overall mode effect based on a few significant differences in

Table 9.15 Comparison of Mode Effect: Cocaine Annual Age-Specific Rates of First Use (per 1,000 Person-Years of Exposure), PAPI and CAI Data

	PAPI 1994-1998		PAPI 1999 Old Method—Edited		CAI 1999 Old Method—Edited	
Year	12-17	18-25	12-17	18-25	12-17	18-25
1965	0	0.5	—†	3.2	—†	0.5
1966	0.4	0.9	0.2	1.1	—†	0.9
1967	0.6	0.8	2.8	0	0.3	2.6
1968	0.5	2.0	—†	0	1.6	4.1
1969	1.3	6.1	0.7	0	1.8	3.3
1970	2.3	8.2	0.2	8.7	1.7	6.7
1971	4.6	6.7	0.3	7.9	2.3	8.1
1972	2.9	5.7	2.7	9.6	2.5	9.0
1973	4.0	11.8	4.9	15.3	3.4	8.8
1974	3.5	17.6	7.8	8.2	2.9	16.7
1975	5.4	15.6	2.8	9.9	4.9	18.2
1976	5.1	14.1	8.7	17.8	4.1	19.0
1977	4.7[a]	20.9	4.1	32.5	10.5	21.4
1978	6.9	21.2	8.6	17.1	5.7	20.3
1979	7.7	19.1	6.1	28.9	5.7	23.0
1980	7.6	28.2	8.2	34.5	5.6	21.7
1981	9.1	26.2[a]	7.4	21.3	6.8	19.1
1982	9.2	28.8	5.1	30.1	8.9	23.5
1983	8.1	23.6	11.8	21.7	9.6	30.5
1984	8.9	23.5	11.1	22.7	9.1	28.2
1985	6.4	26.0	6.4	20.5	9.2	23.0
1986	7.8	24.6[a]	8.8	21.9	8.7	18.0
1987	7.3	21.2	6.7	19.0	9.7	18.4
1988	6.0	18.1	9.7	13.0	10.8	16.1
1989	6.0	16.3	3.7	12.7	7.2	14.6
1990	5.7	12.7	10.0	4.6[b]	5.3	15.9
1991	4.1	10.9	4.8	21.2	4.0	12.6
1992	5.9	9.8	8.2	5.8[a]	4.6	10.8
1993	5.7	11.7	3.2[a]	7.4	6.4	12.2
1994	7.3	11.2	7.0	13.7	5.6	9.5
1995	9.4	13.9	7.3	10.2[b]	7.3	17.1
1996	11.1	14.3	9.5	21.8	10.1	15.2
1997	10.8	15.9	8.4	19.5	11.1	20.2
1998	NA‡	NA‡	9.3	18.0	10.5	20.7

† — Low precision; no estimate reported.
‡ NA = not available.
[a] Difference between PAPI and CAI is statistically significant at the .05 level.
[b] Difference between PAPI and CAI is statistically significant at the .01 level.

annual estimates. This is particularly true when the directions of the effects are not consistent.

A more consistent pattern for the significant annual mode effects is shown for cigarettes (***Table 9.16***), with 11 statistically significant differences all with higher incidence rates based on PAPI interviews.

Table 9.16 Comparison of Mode Effect: Cigarette Annual Age-Specific Rates of First Use (per 1,000 Person-Years of Exposure), PAPI and CAI Data

	PAPI 1994-1998		PAPI 1999 Old Method—Edited		CAI 1999 Old Method—Edited	
Year	12-17	18-25	12-17	18-25	12-17	18-25
1965	104.7	115.0[b]	87.5	105.3	109.6	68.0
1966	92.3	119.0[a]	71.2	58.4	93.3	75.7
1967	112.2	122.5	114.1	125.2	110.4	95.1
1968	100.2	114.8[a]	88.4	205.8[a]	99.4	77.6
1969	112.3	117.2[b]	163.6[b]	80.1	95.0	63.2
1970	113.0	106.7	120.1	130.4	110.5	97.7
1971	116.1	92.1	112.0	96.9	109.7	77.3
1972	125.8	107.0	135.4	111.5	109.4	85.9
1973	116.2	82.0	124.1	99.3	116.6	94.7
1974	124.3	84.9	120.0	95.0	117.8	73.6
1975	121.4	94.7	128.5	107.3	120.9	77.1
1976	122.4	81.1	147.0	75.4	139.8	82.8
1977	126.7	86.5[a]	128.5	83.4	114.9	61.4
1978	113.7	75.8	116.8	89.4	132.0	69.1
1979	109.2	86.4	104.7	71.2	116.8	78.2
1980	107.0	70.0	96.4	73.4	98.1	72.4
1981	105.5	66.3	134.3[a]	76.3	96.3	58.6
1982	103.5	61.9	98.6	69.7	94.7	61.9
1983	104.6	63.0	91.2	87.1	98.0	63.2
1984	100.9	70.3	89.3	68.7	102.5	62.6
1985	108.9	65.4[b]	94.9	60.1	113.4	40.8
1986	106.2	76.3[a]	109.0	60.1	105.9	57.0
1987	97.7	65.4	108.9	52.4	106.1	59.2
1988	103.0	62.6	86.3	66.3	102.4	52.7
1989	99.4	62.1	85.5	67.5	97.7	66.8
1990	99.9	68.8[a]	101.6	70.5	100.1	52.0
1992	111.0	64.3	90.5	67.3	102.0	59.7
1993	118.2	67.9	110.9	70.0	105.1	57.4
1994	128.7	80.8	118.9	66.2	124.6	77.3
1995	146.2	76.9	116.8	63.3	128.9	78.7
1996	133.0	81.5	128.6	103.2	139.9	87.9
1997	NA†	NA†	134.5	78.8	136.7	76.9
1998	NA†	NA†	NA†	NA†	NA†	NA†

†NA = not available.
[a] Difference between PAPI and CAI is statistically significant at the .05 level.
[b] Difference between PAPI and CAI is statistically significant at the .01 level.

The 1999 PAPI annual estimates are based on relatively small sample sizes and have large variances, so estimates based on 1994-1998 NHSDA were also compared to CAI estimates in ***Tables 9.14, 9.15***, and ***9.16***. Another way to deal with high variability in annual estimates is to average the annual estimates over the period 1965 through 1998. This is the approach that was taken in ***Tables 9.17*** and ***9.18***, which look at mode

Table 9.17 Average Mode Effects for Estimated Incidence Rates: 1999 CAI and PAPI Samples (Old Method—Edited Data)

Substance	CAI Sample		PAPI Sample		Relative Effect (CAI vs. PAPI), %	
	12-17	18-25	12-17	18-25	12-17	18-25
Marijuana	50.9	44.5	48.3	45.1	5.41	-1.22
Cocaine	6.2	15.0	6.1	16.1	0.71	-6.98
Crack	1.3	2.8	1.8	3.3	-25.83	-16.68
Heroin	0.7	1.6	1.1	1.4	-34.07	14.89
Hallucinogens	11.1	14.2	10.3	13.8	7.55	2.77
Inhalants	10.2	5.6	8.4	5.8	21.19	-3.14
Pain relievers	6.4	7.3	4.9	5.5	32.56	33.36
Tranquilizers	4.1	5.8	2.8	3.8	46.41	50.24
Stimulants	6.1	7.7	4.4	4.7	40.70	64.67
Sedatives	2.8	3.7	2.4	2.7	17.65	39.12
Alcohol	113.0	201.1	104.9	212.1	7.73	-5.19
Cigarettes	110.5	70.5	110.6	84.4	-0.09	-16.47

Table 9.18 Average Mode Effects for Estimated Number of Initiates and Average Age of First Use: 1999 CAI and PAPI Samples (Old Method—Edited Data)

Substance	CAI Sample		PAPI Sample		Relative Effect (CAI vs. PAPI), %	
	Initiates (1000s)	Mean AFU†	Initiates (1000s)	Mean AFU	Initiates (1000s)	Mean AFU
Marijuana	2,115	18.1	2,139	18.4	-1.15	-1.91
Cocaine	770	21.1	788	20.5	-2.40	2.65
Hallucinogens	713	18.8	724	18.7	-1.51	0.63
Inhalants	465	17.1	452	17.2	2.89	-0.16
Pain relievers	491	21.0	366	21.6	34.13	-2.41
Stimulants	413	19.1	286	18.2	44.29	4.84
Alcohol	3,645	16.9	3,688	17.1	-1.14	-1.05
Cigarettes	2,881	15.5	2,975	15.5	-3.13	-0.16
Daily cigarettes	1,696	18.2	1,718	17.9	-1.28	1.73

†AFU = Age of first use.

effects based on comparisons of 1999 CAI and 1999 PAPI based on multiyear averages.[7] No comparisons are provided for methamphetamines,

[7]For some years, data were suppressed due to small sample sizes or poor precision. The suppression rules used in 1999 are documented in Appendix B to the Summary of Findings from the 1999 National Household Survey on Drug Abuse (Office of Applied Studies, 2000). Averages are taken over the period 1965 to 1998, but years with suppressed data are excluded from the averages.

smokeless tobacco, or cigars, since age-at-first-use questions specific to these substances were not asked in the PAPI questionnaire. In ***Table 9.18***, comparative data for crack, tranquilizers, and sedatives are excluded because too few annual estimates were available to compute meaningful averages.[8]

Before trying to interpret the results in ***Tables 9.17*** and ***9.18***, it is useful to consider the differences in the question presentation in CAI and PAPI modes. The PAPI mode utilizes a self-administered answer sheet, and the respondent is asked to respond to the age of first use even if they indicate never having used the substance in their lifetime in the previous question; the respondent is asked to specify an age of first use or check a box indicating "I have never [used the substance] in my life." The CAI mode uses computer-driven branching. If the respondent answers No, Don't Know, or Refused to the substance gate question on ever use, the computer immediately skips to the next substance module. In both modes, but particularly in CAI, the initial positive responses to the substance lifetime use question identify the only respondents that should report a substance age at first use. Based on this gate question approach, we might expect that multiyear incidence rates would relate to the reporting of lifetime use.

The results from ***Chapter 7*** on reported lifetime use before any editing and imputation are therefore relevant to mode comparisons relating age of first use and the measures derived from it. ***Table 7.6*** shows some of the details of lifetime use reporting prior to editing or imputation based on the matched sample; a limited number of statistically significant differences between CAI and PAPI are noted. The ***Table 7.6*** results for 12 to 17 and 18 to 25 can be compared to the multiyear incidence effects in ***Table 9.17***. Pain relievers, tranquilizers, and stimulants show significantly higher CAI lifetime use at both age groups in ***Table 7.6*** and correspondingly higher CAI multiyear average incidence rates in ***Table 9.17***. Inhalants show significantly higher lifetime use rates in ***Table 7.6*** for both age groups, with consistent incidence rate mode effects at age 12 to 17 and a small opposite effect at age 18 to 25 in ***Table 9.17***. ***Table 7.6*** shows statistically significant higher lifetime use of sedatives under CAI for age 18 to 25 only, and ***Table 9.17*** shows a large effect for persons 18 to 25.

No statistical testing was done for ***Tables 9.17*** or ***9.18***, but the results in these two tables appear to relate to ***Table 7.6***, where statistical testing of

---

[8]Somewhat arbitrarily, comparative averages for substances with fewer than 20 unsuppressed annual estimates (out of a possible 34 years of data) were excluded from the summary tables.

mode effects was feasible. As was noted in ***Chapter** 7*, some of the observed differences for inhalants and psychotherapeutics may be partially attributable to the specific probing for a larger number of substances when applying the CAI mode.

The mean annual number of initiates per year in ***Table 9.18*** can be compared to the lifetime use rates for persons 12 or older in ***Table 7.6***. Statistically significant higher CAI lifetime use rates are shown for cocaine, inhalants, pain relievers, and stimulants. The average number of annual initiates (***Table 9.18***) also shows higher CAI estimates for inhalants, pain relievers, and stimulants, but a small reverse effect for cocaine. A statistically significant lower CAI lifetime use rate is shown for cigarettes in ***Table 7.6*** and corresponds with a lower CAI initiates number estimated in ***Table 9.18***. There is not an obvious relationship of age-at-first-use estimates to lifetime use rates; ***Table 9.18*** shows small differences in both directions.

Given the above evidence of a relationship between mode effects for incidence rates and estimated numbers of initiates and mode effects for lifetime use, the conclusions from ***Chapter** 7*, which predominantly show higher estimates with CAI than with PAPI, seem to carry over to mode effects examined here. Even though the statistical results are mixed, there is evidence of some overall increased reporting of substance use initiation under the CAI mode, which increases estimates of incidence rates and of numbers of initiates. There was no appreciable effect on mean age of first use.

9.3 Discussion of Cigarette Daily Use Estimates

Two additional factors influenced the difference between the incidence estimates for daily cigarettes use from the 1998 PAPI and 1999 CAI NHSDA surveys. First, the daily cigarette smoking questions changed slightly in the 1999 CAI questionnaire. Second, a programming logic adjustment was made to more accurately count the at-risk population that comprises the denominator of the incidence rates.

9.3.1 Wording and Mode Effects

The general format remained constant between the PAPI and the CAI questionnaire; that is, the format consisted of two questions, the first asked if the respondent ever smoked cigarettes daily and the second asked when they first began daily smoking. However, two specific changes occurred.

As with all 1999 CAI questions, the daily cigarette question contained skip logic so that, if respondents did not identify themselves as ever being daily smokers, they were not asked when they first began smoking daily. This skip logic was not present in any of the PAPI surveys. Another change occurred in the questionnaire wording. The PAPI survey asks the respondent "Has there ever been a period in your life when you smoked cigarettes every day?" However, the 1999 CAI survey asks "Has there ever been a period in your life when you smoked cigarettes every day *for at least 30 days*?" This slight definition change in the daily smoking question could affect whether respondents consider themselves daily smokers. By imposing this time restriction of at least 30 days in the 1999 CAI, it is conceivable that a reduction would occur in the number of daily smokers. ***Tables 9.19*** and ***9.20*** show the 1999 PAPI and CAI estimates for daily smoking incidence rates, numbers of initiates to daily smoking, and mean age of first starting to smoke daily. Estimates under both modes use edited data and the old incidence rate methodology.[9] Note that the observed changes reflect the combined impacts of the mode change and the question wording change.

The annual incidence rates (***Table 9.19***) are quite variable. In two years, significantly higher PAPI rates are identified at age 12 to 17. Only one significant difference is identified for persons 18 to 25. It suggests a higher rate with the CAI mode and wording. Taking the average over the 34 years reported, the incidence rate decreases from 55.7 for PAPI to 54.2 for CAI at age 12 to 17 and decreases from 83.2 for PAPI to 75.4 for CAI at age 18 to 25. Estimates of initiates and mean age of first use (***Table 9.20***) are also quite variable. Few significant differences in annual estimates are noted. Based on 1965-1998 averages, the number of initiates increases slightly from 1,718,000 persons per year for PAPI to 1,756,000 persons per year for CAI. The mean age of first beginning daily smoking also increases slightly from 17.9 for PAPI to 18.2 for CAI. Although there is no large consistent set of changes, it does appear that the new wording may cause respondents to report a later start date for daily smoking. Reporting a later initiation of daily smoking could decrease incidence rates for younger age groups while still maintaining about the same number of annual initiates

[9]Both the CAI and PAPI estimates shown in Table 9.19 contain the programming error that is addressed in Section 9.3.2 in order to isolate the effect of the questionnaire and mode change.

Table 9.19 Mode and Wording Effects: Annual Age-Specific Initiation of Daily Cigarette Use Rates (per 1,000 Person-Years of Exposure)†

	12-17		18-25	
Year	1999 PAPI	1999 CAI	1999 PAPI	1999 CAI
1965	30.6	38.8	42.1[b]	113.9
1966	22.5	40.2	94.6	82.4
1967	47.4	55.3	145.2	77.6
1968	37.3	47.0	130.5	102.4
1969	40.6	51.6	137.9	102.9
1970	65.9	56.7	92.8	86.4
1971	87.4[a]	47.2	86.2	78.3
1972	70.0	54.6	127.1	91.5
1973	60.8	51.4	128.5	84.0
1974	63.0	63.1	121.9	84.7
1975	47.7	61.6	121.0	96.3
1976	77.8	54.2	122.8	63.5
1977	95.0	60.6	47.2	108.7
1978	42.3	60.2	100.0	75.3
1979	35.6	57.5	77.5	90.6
1980	59.1	56.4	67.6	73.2
1981	40.5	55.1	71.6	76.6
1982	64.3[a]	37.6	83.5	71.8
1983	58.3	44.2	60.5	51.0
1984	36.6	50.2	56.4	63.8
1985	68.1	54.1	43.1	55.9
1986	41.4	52.8	54.5	51.4
1987	54.6	55.2	55.0	56.9
1988	62.3	52.6	73.6	48.9
1989	50.9	51.6	67.5	58.5
1990	57.7	44.6	50.5	65.2
1991	48.0	46.9	60.3	60.1
1992	44.8	48.5	58.1	60.5
1993	59.4	58.9	67.7	57.8
1994	59.4	65.2	60.8	66.7
1995	75.0	71.7	63.2	75.8
1996	64.6	73.0	81.4	75.9
1997	73.3	73.8	87.9	81.0
1998	52.7	52.0	91.8	75.6

† Both the CAI and PAPI estimates shown in this table contain the programming error in order to be able to examine the effect of the questionnaire and mode change.
a Difference between PAPI and CAI is statistically significant at the .05 level.
b Difference between PAPI and CAI is statistically significant at the .01 level.

across all age groups. The estimated increase in the mean age of first use also supports this conclusion.

9.3.2 Programming Logic Adjustment

The second factor had a greater effect on the trend estimates. In the computer programs used before the 1999 analysis, the program logic

Table 9.20 Comparison of Number of Daily Cigarette Initiates (in Thousands) and Mean Age at First Daily Cigarette Use, 1999 PAPI Versus 1999 CAI

	Initiates (1000s)		Mean Age	
Year	1999 PAPI	1999 CAI	1999 PAPI	1999 CAI
1965	921[a]	1,675	18.1	17.8
1966	1,140	1,603	18.8	18.8
1967	2,002	1,816	18.2	18.1
1968	1,744	1,972	18.0	18.7
1969	2,018	2,057	17.5	18.5
1970	2,206	1,936	16.1	17.2
1971	2,369	1,756	16.9	18.1
1972	2,416	2,144	17.4	18.2
1973	2,270	1,955	18.1	17.5
1974	2,200	2,165	17.8	17.3
1975	1,957	2,344	18.3	17.8
1976	2,375	1,812	17.0	18.1
1977	1,844	2,458	16.5[a]	18.4
1978	1,573	1,829	18.0	17.5
1979	1,563	2,091	20.2	18.5
1980	1,574	1,836	17.7	18.2
1981	1,315	1,813	17.5	18.2
1982	1,814	1,490	17.7	18.8
1983	1,682	1,352	18.4	17.8
1984	1,115[a]	1,508	17.7	18.1
1985	1,612	1,467	18.8	17.7
1986	1,350	1,384	19.0	18.1
1987	1,550	1,478	19.4	19.1
1988	1,639	1,343	17.5	18.4
1989	1,434	1,449	18.5	18.1
1990	1,435	1,381	18.1	18.3
1991	1,402	1,485	18.2	20.0
1992	1,366	1,472	17.9	18.6
1993	1,579	1,587	17.8	18.1
1994	1,557	1,760	17.2	18.0
1995	1,732	1,923	16.9[a]	18.5
1996	1,891	1,889	17.5	18.4
1997	2,008	1,950	16.7	18.8
1998	1,758	1,664	18.1	19.0

[a] Difference between 1999 PAPI and CAI is statistically significant at the .05 level.

excluded all lifetime cigarette users who never used cigarettes daily from the analysis. This reduced the estimate of the population at risk (the denominator of the incidence rate estimate) but did not reduce the estimate of new initiates to daily smoking (the numerator of the incidence rate estimate). The effect of this program logic error was substantial and artificially increased incidence rate estimates in prior years; for example, ***Table 9.21*** shows that, for 1998, the 12 to 17 incidence rate estimate of 35.3 initiations of daily smoking per 1000 person-years of exposure using the old incidence rate methodology without the programming error would have been estimated as 52.0 initiations of daily smoking per 1,000 person-years of exposure with the programming error.

Table 9.21 Effect of the Program Logic Error for Cigarette Daily Use for the Annual Age-Specific Initiation Rates (per 1,000 Person-Years of Exposure): 1999 CAI

Age →	**12-17**				**18-25**			
Imputation →	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Incidence rate method →	Old	Old	New	New	Old	Old	New	New
Program logic error→	Yes	No	No	No	Yes	No	No	No
1996	72.8	45.7	51.4	50.0	75.9	38.6	33.2	31.6
1997	73.8	47.3	54.1	51.6	81.0	40.8	34.7	32.7
1998	52.0	35.3	40.7	37.3	75.6	37.2	32.7	30.8

Table 9.21 shows the rates for year of first use 1996-1998 computed using the old method with and without the programming error (first two columns under each age group). Two additional columns are included to show the impact of the incidence rate method (compare second and third columns under each age group) and the impact of the new editing and imputation approach (compare third and fourth columns under each age group).

The major differences shown in ***Table 9.21*** are decreases associated with correcting the program logic error. Daily cigarette use is influenced by the editing and imputation methodology and by the method of incidence rate calculation in ways similar to other substances discussed in ***Section 9.1***. The new incidence rate calculation method increases incidence rates for age 12 to 17 and decreases them for age 18 to 25. The new editing and imputation methodology increases incidence rates at both age groups.

9.4 Summary and Conclusions

The impact of revised editing and imputation procedures in 1999 was evaluated. Although the estimates for individual years were quite variable, the overall average impact of the new editing and imputation procedures was to increase incidence rates for both age groups (12 to 17 and 18 to 25, see ***Table 9.7***) and to increase the estimated number of new initiates (***Table 9.12***). The largest impacts were observed for pain relievers and other substances that use multiple gate questions before presenting the age-of-first-use question. Estimates of the average age of initiation did not appear to be consistently changed in either direction by the change in editing and imputation (***Table 9.12***).

The impact of the new method of incidence rate calculation was also studied. The number of new initiates occurring at age 17 was presumably quite high for almost all substances. The new incidence rate calculation rules treated respondents as 17 year olds right up to (but not including) their 18th birthday. The old rule classified respondents as 18 years old for the entire year in which their 18th birthday occurred. Thus, the new calculation method had the effect of increasing the estimates of time at risk and the number of initiates for 17 year olds, but because the number of initiates is high at age 17, the overall impact was greater on the numerator than the denominator. As a result, the incidence rates for persons aged 12 to 17 increased and the incidence rates for persons 18 to 25 usually decreased somewhat with the new method.

Mode effects could not be cleanly isolated because of some accompanying changes in the question routing process and supplementary questions on date of first use for recent users that were implemented in conjunction with the implementation of CAI. Within this limitation, comparable data from PAPI and CAI were studied. One somewhat surprising result was that the level of missing or inconsistent data actually increased with the introduction of CAI (***Table 9.13***). However, this may have resulted because of the increased number of checks employed to identify inconsistent data in the post-survey processing. The increase in the proportion of missing age-at-first-use data may have been facilitated by the respondent's option to answer Don't Know or Refused in CAI. A similar increase in missing data with CAI was noted for substance lifetime use in ***Chapter*** *7*.

A pattern of mode effects similar to that observed (***Chapter*** *7*) for reported lifetime substance use was found,with generally higher reporting of initiation in CAI than in PAPI. This makes sense since a respondent must be at least a lifetime user in order to legitimately report an age at first use; in the CAI instrument, respondents are routed around the age-at-first-use question if they do not answer affirmatively to the associated lifetime use question. As discussed in ***Chapter*** *7*, some of these differences between PAPI and CAI estimates may be due to contextual changes other than just mode of presentation. There was no apparent mode effect on NHSDA estimates of average age-at-first-use of substances.

The special analyses of initiation of daily cigarette smoking identified a large reduction in the estimates based on the correction to the program logic. Otherwise, the pattern of effects due to imputation and incidence rate methodologies was similar to that observed for other substances.

A minor difference in the definition of daily cigarette use in CAI and PAPI appears to have resulted in slightly higher estimates of average age at first daily smoking with CAI than with PAPI.

The larger NHSDA sample sizes available since 1999 will help make the study of the initiation of substance use more feasible and more precise. The revisions and corrections introduced in the CAI questionnaire, in the coordinated editing and imputation procedures, and in rate computation methodology in 1999 should also increase the utility of the survey data for these purposes. Based on the analyses reported in this chapter, any comparisons of 1999 and subsequent years' data with data from 1998 and prior years' data should either be avoided or tempered with an understanding of the methodological effects reported above.

9.5 References

Greenberg, R.S., Daniels, S.R., Flanders, W.D., Eley, J.W., & Boring, J.R. (1996). *Medical epidemiology*. Norwalk, CT: Appleton & Lange.

Johnson, R.A., Gerstein, D.R., Ghadialy, R., Choi, W., & Gfroerer, J. (1996). *Trends in the incidence of drug use in the United States, 1919-1992*. Substance Abuse and Mental Health Services Administration (DHHS Pub. No. SMA 96-3076). U.S. Govt. Printing Office, 1996.

Office of Applied Studies, Substance Abuse and Mental Health Services Administration. (2000, August). *Summary of findings from the 1999 NHSDA* (DHHS Pub. No. SMA 00-3466). Rockville, MD: Author.