2014 NATIONAL SURVEY ON DRUG USE AND HEALTH

METHODOLOGICAL RESOURCE BOOK SECTION 10: EDITING AND IMPUTATION REPORT

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland

February 2016

This page intentionally left blank

2014 NATIONAL SURVEY ON DRUG USE AND HEALTH: EDITING AND IMPUTATION REPORT

Prepared for the 2014 Methodological Resource Book (Section 10)

Contract No. HHSS283201300001C
RTI Project No. 0213757.004.106.004
Deliverable No. 49

RTI Authors:

Michael Bradshaw
Lisa Carpenter
Chuchun Chien
Susan Edwards
Glynis Ewing
Peter Frechtel
Kristen Gulledge Brown
Wafa Handley
Larry Kroutil
Jeff Laufenberg
Philip Lee
Peilan Chen Martin
Andrew Moore

RTI Project Director:

David Hunter

SAMHSA Project Officer:

Peter Tice

SAMHSA Author:

Rebecca Ahrnsbrak

For questions about this report, please e-mail Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

February 2016

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2016). *2014 National Survey on Drug Use and Health: Methodological Resource Book (Section 10, Editing and Imputation Report)*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Acknowledgments

This report was prepared for the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality, by RTI International (a registered trademark and a trade name of Research Triangle Institute). Special thanks are due to several current and former RTI staff members. Ralph Folsom, Eric Grau (formerly with RTI), and Avinash Singh (formerly with RTI) co-developed the predictive mean neighborhood (PMN) methodology. Claudia Clark edited the report, Debbie Bond formatted the report, and Amanda Lewis-Evans assisted with the flowcharts that appear in Chapter 3 (Figures 3.1 through 3.4). SAMHSA reviewers include Eunice Park-Lee and Matthew Williams.

Table of Contents

Table of Contents

Chapter **Page**

Table of Contents

Chapter **Page**

Table of Contents

Chapter **Page**

Table of Contents

Chapter **Page**

Table of Contents (continued)

List of Tables

List of Tables (continued)

List of Figures

Figure **Page**

This page intentionally left blank

1. Introduction

Conducted annually, the National Survey on Drug Use and Health (NSDUH) is the primary source of information on alcohol use, tobacco use, illicit drug use (including misuse of prescription drugs), substance use disorders, and mental health issues for the U.S. civilian, noninstitutionalized population aged 12 years or older. In the 2014 NSDUH, this population included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes) and civilians residing on military bases. The target population excluded people with no fixed household address (e.g., homeless or transient people not in shelters), residents of institutional group quarters (e.g., jails and nursing homes), children younger than 12 years of age, and active military personnel. As it has since 1999, the 2014 NSDUH utilized a 50-state, multistage cluster design that enables the Substance Abuse and Mental Health Services Administration (SAMHSA) to provide representative estimates for each state and the District of Columbia. Both direct and model-based state and substate estimates are produced on a variety of measures based on a combination of multiple years of data.

This report focuses on the editing and statistical imputation procedures that were applied to respondent data for the 2014 NSDUH. Logical editing uses data from elsewhere within the same respondent's record to reduce the occurrence of missing or ambiguous data or to resolve inconsistencies between related variables. Imputation is defined as the replacement of missing values with valid, nonmissing values. Statistical imputation usually involves some randomness to preserve the natural variability in the data.

1.1 Organization of the NSDUH Questionnaire and Overview of Content

The 1999 survey marked the transition from data collection based on paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI).[1] Since then, the NSDUH data have been collected using CAI, which allows a private mode of data collection for respondents to answer questions pertaining to drug use and other sensitive topics.

The CAI interview is organized by modules. "Modules" in the NSDUH questionnaire refer to sections that are organized together by mode of administration, content, and routing logic. For example, the alcohol module includes an initial question that asks whether respondents ever had a drink of an alcoholic beverage. Respondents who report that they have ever used alcohol are asked additional questions about their age when they first used alcohol, when they last used alcohol, and (if relevant) their use of alcohol in the 12 months or 30 days prior to the interview. Respondents who do not report lifetime alcohol use are routed to the next module, which pertains to marijuana use.

The first module consists of questions about certain demographic characteristics, including birth date (which is used to determine a respondent's age), gender, marital status, Hispanic/Latino origin, racial group, and education level (highest grade completed). Computer-assisted personal interviewing (CAPI) is used for these questions, a process in which interviewers read the questions and respondents give their answers aloud to the interviewers,

[1] CAI specifications for the 2014 NSDUH are available at http://www.samhsa.gov/data/.

who then enter the responses into the computer. The logic for determining which questions the interviewers should ask is controlled by the computer program based on the responses to previous questions that interviewers enter into the computer. Consequently, interviewers can concentrate on asking questions and recording respondent answers, without having to concern themselves with comprehending and following skip pattern instructions.

Following completion of these demographic questions, interviewers orient respondents about the use of specific features of the laptop computer and use of headphones for listening to questions. Respondents then complete a brief tutorial on answering self-administered questions, which are administered through use of audio computer-assisted self-interviewing (ACASI). ACASI is used for the majority of questions in the interview and is particularly useful for collecting data about sensitive topics such as substance use, problems associated with substance use, risk and protective factors related to substance use, and mental health issues.

In the ACASI portion of the interview, respondents can read the questions on a computer screen and also are encouraged to listen to an audio recording of the questions on headphones. Respondents then enter their answers directly into a laptop computer. These features of ACASI prevent interviewers or others in the household from knowing what questions the respondents are being asked and how they are answering. The availability of audio recordings of the questions is especially useful for respondents with limited reading ability because they can listen to the questions instead of having to read them.

Once respondents complete the ACASI portion of the interview, they turn the laptop computer back over to the interviewer. Remaining interview questions are administered through CAPI. Topics in the remainder of the interview include immigrant status (i.e., whether respondents were born in the United States), state residency in periods prior to the interview, current education status, employment status and workplace issues, household characteristics, health insurance coverage, and income.

In addition, the CAI instrument consists of core and noncore modules. Core modules, such as those pertaining to key demographic characteristics and substance use, have been designed to stay relatively constant from one year to the next to permit measurement of trends in drug use across time. Table 1.1 summarizes the content of core modules in the 2014 NSDUH interview, including the mode of administration. For prescription psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives), the questionnaire asked about "nonmedical" use. Nonmedical use is defined in NSDUH as use of a drug without a prescription of the respondent's own or use only for the experience or feeling that the drug caused. In addition, for the questions for lifetime nonmedical use of prescription drugs, respondents were shown printed "pill cards" with pictures of prescription drugs for questions in the module to aid respondents in answering the questions.

Table 1.1 Content of Core Modules in the 2014 NSDUH Interview

Module	Content	Mode of Administration
Core Demographics	• Age • Gender • Hispanic/Latino origin and race • Marital status • Military service • Highest educational grade • Perceived health status	Interviewer administration using CAPI
Tobacco	• Lifetime use or nonuse of the following: – Cigarettes – Snuff – Chewing tobacco – Cigars – Pipe tobacco • Additional questions for lifetime users	Self-administration using ACASI
Alcohol	• Lifetime use or nonuse • Additional questions for lifetime users	ACASI
Marijuana	• Lifetime use or nonuse • Additional questions for lifetime users	ACASI
Cocaine and Crack	• Lifetime use or nonuse of the following: – Any cocaine – Crack cocaine (if lifetime cocaine user) • Additional questions for lifetime users	ACASI
Heroin	• Lifetime use or nonuse • Additional questions for lifetime users	ACASI
Hallucinogens	• Lifetime use or nonuse of seven hallucinogens, including any other hallucinogen besides the ones that had been listed • Additional questions if lifetime use reported for any of the seven specific hallucinogens	ACASI
Inhalants	• Lifetime use or nonuse of 11 specific types of inhalants for kicks or to get high, including any other inhalant besides the ones that had been listed • Additional questions if lifetime use reported for any of the 11 specific inhalants	ACASI
Pain relievers	• Lifetime use or nonuse of any of 28 specific prescription pain relievers that were not prescribed or were taken only for the experience or feeling (i.e., "nonmedical" use), including any other prescription pain reliever besides the ones shown to the respondent • Additional questions if lifetime nonmedical use reported for any of the specific pain relievers	ACASI, plus a printed "pill card" showing pictures of specific pain relievers to aid respondent recall
Tranquilizers	• Lifetime nonmedical use or no nonmedical use of any of 22 specific prescription tranquilizers, including any other prescription tranquilizer besides the ones shown to the respondent • Additional questions if lifetime nonmedical use reported for any of the specific tranquilizers	ACASI, plus a printed pill card showing pictures of specific tranquilizers to aid respondent recall

Table 1.1 Content of Core Modules in the 2014 NSDUH Interview (continued)

Module	Content	Mode of Administration
Stimulants	• Lifetime nonmedical use or no nonmedical use of any of 19 specific prescription stimulants or types of stimulants (e.g., prescription diet pills), including methamphetamine (which often is illegally manufactured) and any other prescription stimulant besides the ones shown to the respondent • Additional questions if lifetime nonmedical use reported for any of the specific stimulants or types	ACASI, plus a printed pill card showing pictures of specific stimulants to aid respondent recall
Sedatives	• Lifetime nonmedical use or no nonmedical use of any of 15 specific prescription sedatives or types of stimulants (e.g., barbiturates), including any other prescription sedative besides the ones shown to the respondent • Additional questions if lifetime nonmedical use reported for any of the specific sedatives or types	ACASI, plus a printed pill card showing pictures of specific sedatives to aid respondent recall

ACASI = audio computer-assisted self-interviewing; CAPI = computer-assisted personal interviewing.

In contrast to the core modules, the content of noncore modules can change across years to measure new or developing topics of interest or to rotate certain topics in or out of the interview. In noncore sections, therefore, questions or entire modules can, in principle, be added or deleted, or the wording of existing questions can change from one year to the next. In practice, however, many of the noncore portions of the interview have remained in the survey, relatively unchanged, from year to year.

The topics that are covered in noncore modules also can show more variation than the topics that are included for the core modules. As shown in Table 1.2, for example, there were 18 noncore ACASI modules in 2014 that covered topics such as injection drug use, perceptions of risk and availability for different substances, substance use disorders (i.e., dependence or abuse), arrests and driving under the influence of alcohol or illicit drugs, treatment for substance use problems, physical health conditions, risk and protective factors for substance use among adolescents, and mental health issues.

Table 1.2 Content of Noncore ACASI Modules in the 2014 NSDUH Interview

Module	Content
Special Drugs	• Routes of heroin administration (i.e., smoking, sniffing, injection) • Injection of methamphetamine, other stimulants, cocaine, or other drugs • Additional methamphetamine questions • General injection use behaviors • Miscellaneous drugs (e.g., Adderall®, Ambien®, cough and cold medications)
Risk/Availability	• Perceived risk of harm associated with use of cigarettes, alcohol, or specific illicit drugs • Perceived availability of illicit drugs • Miscellaneous risk behaviors (e.g., seatbelt use)
Blunts	• Use of cigars with marijuana in them (i.e., "blunts") • Use of marijuana that had been recommended by a doctor or other health care professional
Substance Dependence and Abuse	• Nicotine (i.e., cigarette) dependence • Symptoms of dependence and abuse for alcohol and illicit drugs, including nonmedical use of prescription drugs

Table 1.2 Content of Noncore ACASI Modules in the 2014 NSDUH Interview (continued)

Module	Content
Special Topics	• Arrests in the lifetime and past 12 months • Probation and parole status, past 12 months • Driving under the influence of alcohol or illegal drugs • Knowledge of state laws for marijuana possession
Market Information for Marijuana	• How respondents who used marijuana in the past 12 months obtained marijuana, including price/value information for marijuana that was purchased or obtained through a trade
Prior Substance Use	• Use of marijuana, cigarettes, alcohol, or cocaine more than 12 months ago but within the past 24 months • Last use of substances included in the core drug section (see Table 1.1) if not used in the past 30 days • Sources of prescription drugs and methamphetamine • Sequence of initiation of cigarettes, alcohol, and marijuana for adolescents aged 12 to 17
Substance Treatment	• Treatment for use of alcohol or illicit drugs (i.e., not counting tobacco) in the lifetime and past 12 months • Perceived need for treatment for use of alcohol or illicit drugs • Barriers to receiving treatment for use of alcohol or illicit drugs
Health Care	• Pregnancy status of females aged 12 to 44 • Emergency room visits and hospitalizations in the past 12 months • History of specific medical conditions • Height and weight • Outpatient medical visits in the past 12 months • Conversations about tobacco, alcohol, or illicit drug use with health care professionals in the past 12 months
Adult Mental Health Service Utilization	• Administered to respondents aged 18 or older • Use of inpatient or outpatient mental health services in the past 12 months • Payment for inpatient or outpatient mental health services • Use of prescription medication to treat a mental health condition • Unmet demand for mental health services • Use of alternative sources for mental health treatment
Social Environment	• Administered to respondents aged 18 or older • Moves in the past 5 years • Specific illegal behaviors regardless of arrest (e.g., selling illegal drugs) • Attitudes about marijuana use • Religious involvement and beliefs
Parenting Experiences	• Administered to parents when two people were selected for an interview and the second selected person was an adolescent • Beliefs about whether their child has used tobacco, alcohol, or drugs • Talks with their child about substance use • Attitudes about drug education
Youth Experiences	• Administered to respondents aged 12 to 17 • Moves in the past 5 years • Risk and protective factors for substance use • Fighting and delinquent behaviors • Attitudes and perceptions of others' attitudes about substance use • Exposure to drug prevention messages • Religious involvement and beliefs

Table 1.2 Content of Noncore ACASI Modules in the 2014 NSDUH Interview (continued)

Module	Content
Mental Health	• Administered to respondents aged 18 or older • Psychological distress in the past 30 days or past 12 months • Impairment in carrying out activities because of psychological distress • Suicidal thoughts and behavior
Adult Depression	• Administered to respondents aged 18 or older • Symptoms of depression in the lifetime and past 12 months • Impairment in carrying out activities because of depression symptoms • Treatment for depression
Youth Mental Health Service Utilization	• Administered to respondents aged 12 to 17 • Use of mental health services in the past 12 months • Reasons for receiving mental health services from specific sources
Adolescent Depression	• Administered to respondents aged 12 to 17 • Age-appropriate questions for symptoms of depression in the lifetime and past 12 months • Age-appropriate questions for impairment in carrying out activities because of depression symptoms • Treatment for depression
Consumption of Alcohol	• Number of drinks on the last occasion of alcohol use in the past 30 days • Underage alcohol use • Alcohol use in combination with illicit drugs • Initiation of consumption of five or more drinks on an occasion (i.e., binge alcohol use) • Females' consumption of four or more drinks on an occasion

ACASI = audio computer-assisted self-interviewing.

1.2 Overall Data Quality Issues with CAI

Conversion of the NSDUH interview from PAPI (paper format) to CAI (computer format) greatly reduced or, in some cases, eliminated the following data quality problems:

- illegible responses, multiple marks, or out-of-range values;
- item nonresponse (i.e., missing data);
- incorrectly executed skip patterns; and
- inconsistencies among related variables.

For example, when a question's instructions ask respondents to choose only one answer, PAPI respondents nevertheless are physically able to mark multiple answers. This cannot occur in the CAI because the computer program will permit entry of only one response to the item. Similarly, the CAI has been 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 editing. Further, the skip patterns that are embedded in the CAI were designed to reduce the occurrence of inconsistent data by not giving respondents the *opportunity* to provide inconsistent answers. The occurrence of inconsistent data was further reduced through the use of consistency checks built into the CAI program that prompted respondents to resolve inconsistencies that occurred 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 were not completely eliminated because CAI respondents still had the option of entering a response of "don't know" or "refused" when answering a given item. Similarly, even though consistency checks were designed to reduce inconsistent reporting, the CAI program was not equipped to address every possible inconsistent report that a respondent could make. Finally, in some situations, conversion to CAI could introduce new data quality issues. As discussed in Section 2.3.2 in Chapter 2, for example, the lack of direct interviewer monitoring could allow some respondents to use the computer keyboard to enter nonsensical patterns of answers for reasons such as accidental errors or disinterest.

1.3 Organization of this Report

The material in this report combines two sections from versions of the NSDUH Methodological Resource Book (MRB) prior to the 2012 survey: Section 10, which addressed logical editing and coding of variables, and Section 11, which addressed the statistical imputation procedures that were applied to variables that underwent additional processing. These two sections have been combined since the 2012 MRB for several reasons.

First, editing and imputation are closely related, and combining the reports affords the opportunity to remove redundant information.[2] Second, structuring the report in this manner is designed to aid readers in following the "life cycle" of NSDUH data, starting with a respondent's answers to NSDUH questions, how the variables that capture these answers are edited, and, where applicable, how the data are imputed after having been edited. Third, in MRBs prior to 2012, some documentation of editing procedures appeared in Section 11 instead of in Section 10; thus, readers who were primarily interested in documentation of the editing procedures that were applied to certain variables had to check two MRB sections to locate that information rather than just one.

Preliminary coding and processing of unedited data after interviewers transmitted the data from the field as well as the general principles of logical editing are discussed in Chapter 2 of this report. The predictive mean neighborhood (PMN) imputation methodology, which is used to impute NSDUH data, is described in detail in Chapter 3. It is recommended that readers first review Chapters 2 and 3, as these two chapters provide a foundation for Chapters 4 through 11, The information in both chapters helps to set an appropriate context for readers as they review the documentation for the specific editing and imputation procedures that were employed on a particular variable set.

Following these initial chapters, Chapters 4 and 5 address the editing and imputation procedures that were applied to the CAI core and noncore demographic variables, respectively. Editing and imputation procedures for the core substance use variables are discussed in Chapter 6, and Chapter 7 addresses the editing procedures for the noncore self-administered variables. Chapter 7 also discusses the imputation procedures for cigarette dependence,[3] which differ from the procedures used for other drug variables. Chapter 8 describes the edits that were applied to the household roster, the creation of imputation-revised versions of the roster-derived household

[2] The similar themes running through Appendices B and E of this report provide a good illustration of the close relationship of editing and imputation. In particular, the edits listed in Appendix B and the logical constraints listed in Appendix E overlap substantially.

[3] This term is referred to in imputation reports prior to 2012 as "nicotine dependence."

composition variables, and the creation of respondent-level variables with individual roster information. Chapter 9 summarizes the editing and imputation procedures that were applied to the income variables. Procedures for logical editing and imputation of missing values in the health insurance variables are described in Chapter 10. The editing and imputation processing of the pair relationship variable and related household count variables are detailed in Chapter 11.

This report also contains 12 appendices. Appendix A of this report is identical to Appendix A of the 2014 NSDUH editing and imputation evaluation report (Center for Behavioral Health Statistics and Quality, 2015c). It contains a number of tables that quantify the amount of imputation and logical assignment (i.e., editing) that selected analytic variables underwent during imputation processing in 2014. Appendix B provides a summary of data issues and the specific edits that were applied in response to these issues for the noncore demographic, core drug use, and noncore self-administered variables. Appendix C provides details on the handling of "OTHER, Specify" responses to the race and Hispanic/Latino group questions so that the data could be summarized in a meaningful way. The covariates in each of the imputation models are listed in Appendix D. The tables in Appendix D also include (1) the starting list of covariates for each model and (2) descriptions of each level and identification of the reference level for categorical covariates. Appendix E provides details on each final hot-deck step in the imputation procedures. The quality control measures used in the imputation procedures are summarized in Appendix F. Reasons that interviewers gave for overriding consistency checks in the household roster are presented in Appendix G, along with evaluations of their legitimacy and the resulting actions in editing the roster. The rules for determining pair relationships are defined in Appendix H. The conditions used for reconciling differing multiplicity counts between pair members are described in Appendix I, and the conditions used for reconciling differing household-level person counts between pair members are described in Appendix J. Appendix K details the priority conditions for creating household-consistent covariates. Appendix L contains detailed information about household-level and person-level eligibility and the completeness criteria used to construct the household-level and person-level files.

1.4 Changes from the 2013 Survey to the 2014 Survey

This section summarizes the changes in editing and imputation procedures implemented on the NSDUH since the 2013 survey year. It is intended for analysts who have had previous experience with the data. Those with little or no prior experience with NSDUH data are encouraged to review Chapters 2 and 3 before reviewing this section. Before making a change to the editing and imputation process, a careful assessment of the impact of the change is done to determine what, if any, effect the change would have on the estimates. The impact assessment involves applying the new process to the previous year's data and comparing the results. For variables whose creation has a stochastic component, the new process is run at least three times to assess whether the results are directly attributable to the new process or whether they could be due to random variation.

Overall, the changes implemented to the editing and imputation procedures between the 2013 and 2014 surveys were minor. These changes are described in three sections below. The first section describes changes to the imputation procedures that were implemented to accommodate changes to the questionnaire. The second section describes changes to the imputation procedures that reduced the time required to process the data, with little to no impact on the final imputation-revised variables. Finally, the third section lists a few minor corrections and improvements to the imputation procedures that also had little impact on the final imputation-revised variables. The changes described in this chapter were implemented during the normal editing and imputation process.

1.4.1 Changes to Accommodate Questionnaire Changes

In the 2014 questionnaire, a question was added (QP03a) to identify an available proxy if one is not identified during the first attempt. To accommodate this change, a new edited variable (PRXRETRY) was created and existing SAS® programs were updated to reflect the new routing procedure. In the 2014 data, there were no positive responses to QP03a.

1.4.2 Changes to Streamline Imputation Procedures

There was one measure implemented in 2014 to streamline the imputation process. A change was made to the imputation procedures for the drug variables to utilize resources more efficiently and ultimately to reduce the amount of time required to process these variables. During the prediction (PRD) modeling step of PMN, the full NSDUH sample of approximately 67,500 records is typically used in SUDAAN® procedures to properly account for the complex sample design. However, because many of the models include small domains (e.g., past month users of heroin among respondents aged 12 to 17), this method for specifying the sample design is inefficient and can be accomplished with a much smaller dataset. Beginning in 2014, the imputation procedures for the drug use variables were migrated to a new system that utilizes a design dataset to inform SUDAAN of the NSDUH sample design. This dataset is much smaller and contains only one observation for each replicate within a stratum (1,500 observations in 2014). Prior to implementation, tests were conducted using 2013 data to ensure that the new method for specifying the design did not result in any changes to the output of the modeling procedures. Because of time constraints, this new method was only implemented for the drug use variables in 2014. Similar updates are planned for other modules (e.g., income) in future survey years as time and budget allow.

1.4.3 Changes Involving Minor Corrections and Enhancements

Eight changes were made to the 2014 editing and imputation procedures that involved minor corrections or enhancements. None of these changes had a significant impact on the estimates.

The first change was related to the hot-deck programs for both 12-month and 30-day frequency. Rather than modeling the exact number of days that a particular drug was used in the past 12 months or in the past 30 days, the PRD modeling step of PMN proceeds as follows.

1. Apply the empirical logit transformation to the proportion of days in the reference period that the respondent used the drug. For discussion of the empirical logit transformation as applied to the 12-month and 30-day frequency-of-use variables, see Sections 6.3.5.2.2 and 6.3.5.4.2, respectively.
2. Fit the prediction model using linear regression and use the transformed variable from step 1 as the response variable.

3. Back-transform the predicted means from step 2 into proportions.
4. Use the proportions from step 3 as the predicted means in the hot-deck step.

In rare cases, this method produced predicted means that were outside the interval [0, 1], and the delta constraint in the hot-deck step was not designed to handle such cases. The delta constraint looks for potential donors whose predicted means are within 5 percent of the recipient's predicted means in such a way that it is insensitive to whether the 0 or the 1 is treated as the "success." In some instances where the recipient's predicted mean was outside the interval [0, 1], the delta constraint eliminated all potential donors from the neighborhood, which led to the delta constraint being dropped. In other cases where the recipient's predicted mean was outside the interval [0, 1], the delta constraint did not eliminate *any* donors from the neighborhood. To correct this situation, the delta constraint was revised to use the absolute value of the predicted mean to ensure that all item nonrespondents whose predicted means were outside the interval [0, 1] were treated in the same manner as item nonrespondents whose predicted means were inside the interval [0, 1]. Between 2011 and 2013, across both 12-month frequency and 30-day frequency for all drugs, 40 item nonrespondents had predicted means that fell outside the [0, 1] interval. Because of the small number of cases affected, this change was made for 2014 processing, but no revisions were made to prior years of data.

The second change made during the 2014 processing cycle was related to "OTHER, Specify" responses for race. Each year, approximately 100 to 150 respondents enter a country rather than an actual race. For these cases, that country's census information is used to probabilistically assign a race, as described in Section 4.2.7.2.1. The distribution of race within each country was last updated in 2006. For the 2014 processing cycle, updated information was obtained as available, and the new distributions were used to assign a value to race when the respondent reported a country in the "OTHER, Specify" response.

The third change made for 2014 was also related to "OTHER, Specify" responses for race. Prior to 2014, respondents who entered "Pakistan," "Sri Lanka," "Bangladesh," "Nepal," "Bhutan," or "South Asia" were classified as "Asian Indian." However, beginning in 2014, these respondents were classified as "Other Asian" to be accurate and more consistent with the manner in which such responses are treated by the U.S. Census Bureau. This change affected 32 to 51 respondents each year from 2002 to 2013. For these cases from years prior to 2014, the affected variables, NEWRACE1 and IRNWRACE, were corrected and redelivered to the NSDUH data files.

The fourth change was an enhancement to the imputation procedures for the variable for respondents born outside the United States (i.e., "age of entry in the United States"). In the response propensity and PRD modeling steps of PMN for age of entry, the models include education level, employment status, and marital status as covariates. Typically these covariates would not be used during the imputation process for respondents who were 12 to 17 years old, but they would be used for other age groups. However, because of small domain sizes for this particular variable, all age groups are combined into a single group for imputation purposes. Prior to 2014, respondents who were 12 to 17 years old and in the domain defaulted to the reference cell. However, beginning in 2014, the 12- to 17-year-old respondents were put into the most logical category. For employment status and marital status, the reference cell was the most logical category (i.e., "other/not in labor force" for employment status and "never been married" for marital status), so no change was made for these covariates. For education level, 12- to 17-year-old respondents had previously defaulted to the reference cell of "college graduates," but beginning in 2014 they were classified as "less than high school." This change is expected to have minimal impact because very few cases (e.g., 15 cases in 2013) require imputation for the age-of-entry variable.

The fifth change was related to editing the household composition roster. Each respondent is asked about the age, gender, and relationship to the respondent of every member of the household. It is not uncommon for these variables to be missing, particularly when the respondent is younger than 17. In households where two individuals are NSDUH respondents, the other pair member often has valid data for those missing values. Beginning in 2013, the other pair member's data were used to edit missing roster ages for those household members with missing information on age whenever possible. In 2014, missing gender and relationship codes were also edited based on the other pair member's data regarding the age, gender, and when appropriate, relationship of the household member with missing data whenever possible.

The sixth change was made to the algorithm for assigning the parent indicator in parent-child pairs when the parent is not easily determined. Prior to 2014, the following logic was used.

1. If one pair member has never been married and the other is either married, widowed, or divorced/separated, then the respondent who is/was married was identified as the parent without considering the ages of the pair members.
2. If neither pair member had ever been married, then the second respondent to be interviewed was identified as the parent without considering the ages of the pair members.
3. If both pair members are married, widowed, or divorced/separated, then the older pair member was identified as the parent.

Beginning in 2014, the process for assigning the parent in these cases was revised so that if neither pair member had ever been married, then the older pair member was identified as the parent. Quality control procedures were also put in place to manually inspect the rosters for all cases where the younger pair member was identified as the parent in parent-child pairs to assess the reasonableness of the parent indicator.

The seventh change involved updating the hot-deck procedures of PMN to accommodate new missingness patterns and to add attempts to find a donor by dropping likeness constraints when necessary. In 2014, a new missingness pattern was identified for cocaine and crack recency/frequency where the recipient was a past month user of cocaine with a known 30-day cocaine frequency but was missing all other variables in the imputation set. In order to find a donor for this particular case, the hot-deck step had to be updated, and the corresponding documentation was also updated to reflect this change. The hot-deck steps and documentation for health insurance and the first stage of pair imputations were also updated in 2014 to add attempts to find donors for particular missingness patterns.

The eighth change to the procedures in 2014 involved the continuing migration of hot-deck programs to the PMN Hot-Deck Common Code. See Ault et al. (2011) for a description of the reasons for these modifications. By the time the 2014 data were processed, all the final hot-deck programs had been migrated, except for recency and frequency for pain relievers and stimulants. The remaining migrations are planned for 2015 processing as time and budget allow.

2. Procedures and General Principles for Editing the Computer-Assisted Data

2.1 Introduction

This chapter provides an overview of the procedures and general principles for editing the computer-assisted interviewing (CAI) data from the National Survey on Drug Use and Health (NSDUH). Logical editing typically uses data from elsewhere within the same respondent's record in a deterministic manner to reduce the occurrence of missing or ambiguous data or to resolve inconsistencies between related variables.[4] In contrast, statistical imputation procedures (see Chapter 3) apply probabilistic (stochastic) statistical methods to identify *another* respondent (i.e., a "donor") whose data are used to replace (1) missing values in the "recipient" respondent's record with nonmissing values from the donor or (2) ambiguous responses in the recipient respondent's record with more specific information from the donor.

As an example of ambiguous data, the CAI logic requires respondents to report that they have used a particular substance (e.g., marijuana) at least once in their lifetime in order to be asked when they last used it. However, a respondent can report lifetime use but not provide a definitive answer for when he or she last used the substance. In the subsequent imputation procedures for this recipient record, a donor is identified who specifically reported last using the substance in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago. Here, the ambiguous answer in the recipient's record for most recent use is replaced with one of the more specific responses supplied by the donor record.

Section 2.2 describes procedures for initially processing the transmitted NSDUH interview data to get it in a form for further data processing. Subsequent editing and coding steps are described in Section 2.3. The final section in this chapter, Section 2.4, discusses additional principles of data processing and editing that were applied once the transmitted interview data had been processed and cases with questionable data had been identified.

2.2 Initial Processing of Transmitted NSDUH Interview Data

The collected interview data were transmitted from the field as ASCII files, and daily SAS® datasets were created from these files. This daily processing included the following activities as part of the creation of initial unedited interview data files:

---

[4] There are a few situations where data from outside the respondent's record is used in logical editing. For example, some editing procedures involve data from the screener where an eligible member of the dwelling unit reports basic information about all members of the dwelling unit. Further, in situations where two members of the same dwelling unit are selected for the survey and complete the interview, data from the second respondent's record may be used in logical editing of some variables in the first respondent's record, or vice versa. This procedure allows use of information from both respondents to determine relationships among household members.

- assignment of standard NSDUH missing data codes (Section 2.2.1),
- remapping of responses to "enter all that apply" questions (Section 2.2.2), and
- identification of "usable" cases (Section 2.2.3).

Each day's processed SAS dataset was merged with the transmitted data to date until the end of the quarter (e.g., January through March = Quarter 1) when a cumulative data file (Section 2.2.4) was produced that contained all transmitted cases from that quarter.

2.2.1 Assignment of Standard NSDUH Missing Data Codes

A key activity in the initial processing of transmitted, unprocessed interview data involved the assignment of standard NSDUH missing data codes. The Blaise program for the CAI instrument uses a code of 8 (or 98 or 998, etc.) to denote responses of "refused" and a code of 9 (or 99 or 999, etc.) to denote responses of "don't know." However, in the NSDUH, a code of 98 is used to indicate when a variable is blank (i.e., not answered), and a code of 99 is used in editing procedures to indicate when the variable corresponding to a question was legitimately skipped. Therefore, the codes for missing data that were supplied by the Blaise program were replaced with the standard NSDUH codes for "don't know" (DK) and "refused" (REF). Assignment of codes as part of the editing procedures to indicate when a question was legitimately skipped is discussed in Section 2.4.2.

The following standard codes for missing data were relevant to the 2014 CAI data, depending on the number of digits for a given variable:

- 94 (or 994 or 9994, etc.) = DON'T KNOW (DK);
- 97 (or 997 or 9997, etc.) = REFUSED (REF); and
- 98 (or 998 or 9998, etc.) = BLANK (i.e., nonresponse [NR]).

Blanks were the default code for missing data if questions were not answered because of the CAI skip logic or if respondents broke off the interview before reaching a particular question. Section 2.4.3 describes situations in which variables retained codes of blank when respondents answered "don't know" to a question that governed CAI skip logic. Section 2.4.2 describes situations in which codes of blank in the unedited data could be replaced with codes to indicate that the questions had been legitimately skipped.

Codes for missing data in most unedited variables were two digits in length (i.e., 94, 97, or 98). For some variables, however, these values were part of the allowable range of responses. Questions that asked respondents to report the age when they first used a particular drug, for example, had an allowable range of up to 110 years. For the variables corresponding to these age-related questions, the codes for missing data were three digits in length (i.e., 994, 997, or 998).

Finally, the CAI logic governed whether respondents were asked additional questions about a topic based on their age, gender, or answers to other preceding questions. For example, questions in the interview about pregnancy applied only to females aged 12 to 44 (Section 7.4.9). These pregnancy questions were skipped if interviewers reported that a respondent was male or if the CAI program recorded that a female respondent was aged 45 or older. When questions were skipped because the criteria in the CAI program were not satisfied for administering the questions, the unedited variables corresponding to the skipped questions retained a code of "blank" (i.e., 98, 998, etc.). In subsequent editing (described in Section 2.4.2), these variables were examined more closely to determine whether the questions had been legitimately skipped (i.e., they were not applicable) or whether they should retain codes of "blank."

2.2.2 Remapping of Responses to "Enter All that Apply" Questions

A second activity associated with the initial processing of transmitted interview data involved the remapping of responses to "enter all that apply" type questions, which allowed respondents to choose as many responses from a given list as applied. Respondents who wanted to report more than one answer from a list did so by typing the numeric codes that corresponded to the applicable responses and by separating each entry with a space.

The CAI program captured information from these "enter all that apply" questions as separate variables in the order that respondents keyed their answer choices. For example, the transmitted CAI data included 18 separate variables for question PR04A to accommodate reports of lifetime nonmedical use of pain relievers from the list. Consequently, an "enter all that apply" variable in the transmitted data could have a different meaning across respondents, depending on which answer a respondent chose first and the number of answers that the respondent chose from the list. For example, if a respondent reported only nonmedical use of the pain reliever OxyContin® from the list in question PR04A, this response would be captured in the *first* variable from the transmitted data. The 17 remaining "unused" variables from question PR04A would be blank. If another respondent chose codeine and morphine as his or her first and second responses from the list and then chose OxyContin®, then the first variable from PR04A would be occupied by the response for codeine, the second variable would be occupied by morphine, the third would be occupied by OxyContin®, and the remaining variables would be blank.

If these "enter all that apply" variables were allowed to remain in the form in which they came from the transmitted CAI data, then these variables would have retained the information about the order in which respondents chose their answers. However, this variable structure makes it difficult to analyze the data. For example, it would be more straightforward for information about lifetime nonmedical use of OxyContin® to be captured in a single variable for subsequent use in creating edited and imputed variables for this measure. Otherwise, 18 different variables would need to be checked to identify reports of nonmedical use of OxyContin® across all of the possible combinations of answers to question PR04A. Therefore, remapping of the responses to these "enter all that apply" questions as part of the processing of transmitted data involved reassigning answers so that a nonmissing value in a given variable had one, and only one, meaning across all respondents, regardless of the number of answers that respondents chose from a list or the order in which respondents keyed their answers. For example, a discrete variable was created as part of the remapping process that captured all reports of nonmedical use of OxyContin® from respondents who chose this drug as part of any of their answers to question PR04A. If respondents did not report nonmedical use of OxyContin®, the remapped variable for OxyContin® was assigned a code of 98 (i.e., blank).

In addition to choosing one or more applicable responses from a list, respondents could use function keys to answer "don't know" or "refused" as their first response to these "enter all that apply" types of questions. In situations where respondents answered "don't know," it would be reasonable to infer that the respondent did not know which particular item on the list applied to him or her. For example, if a respondent answered question PR04A as "don't know," this was inferred to mean that the respondent did not know whether he or she had ever misused codeine, Demerol®, Dilaudid®, and so on, through Ultram®. In this case, a code of "don't know" was propagated to each of the recoded "enter all that apply" variables as part of the daily processing of the transmitted data. Similarly, if a respondent *refused* to answer question PR04A, a refusal code was propagated to all of the recoded variables on that list as part of the daily processing of the transmitted data.

2.2.3 Identification of Usable Cases

Once standard missing data codes had been assigned and the responses to "enter all that apply" questions had been remapped, the third key step in the preliminary processing of transmitted NSDUH data established the minimum item response requirements necessary for cases to be retained for weighting and further analysis (subsequently referred to as "usable" cases). These rules were 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 needed for any given record. In addition, requiring lifetime use or nonuse to be fully defined for at least one substance (i.e., no missing data) can allow data for that substance to be used in statistical imputations for other substances with missing data for lifetime use or nonuse.

In order for a case to be considered usable in NSDUH, both of the following requirements must be met.[5] The term "gate question" is used in connection with the usable case criteria because an affirmative response to these questions (e.g., "Have you ever, even once, used marijuana or hashish?") opens the "gate" to a series of other questions on use of the drug, and a negative response closes the "gate" and leads to the skipping of all other questions on use of that drug.

1. The lifetime cigarette gate question CG01 had to have been answered as "yes" or "no" so that lifetime use or nonuse could be fully defined for at least one substance. Data about lifetime use or nonuse of cigarettes is used in subsequent statistical imputations for other substances where lifetime use/nonuse is undefined.
2. At least 9 of the following 13 additional gate questions had to have answers of "yes" or "no": (1) snuff, (2) chewing tobacco, (3) cigars, (4) alcohol, (5) marijuana, (6) cocaine (in any form), (7) heroin, (8) hallucinogens, (9) inhalants, (10) pain relievers, (11) tranquilizers, (12) stimulants, and (13) sedatives.[6]

[5] The historical background and considerations for establishing usable case rules for the CAI data are discussed in a methodological chapter on editing the 1999 CAI data (Kroutil & Myers, 2002).

[6] Crack cocaine was not included in the usable case rule because the logic for asking about crack cocaine was dependent upon the respondent having answered the lifetime cocaine use question as "yes." In addition, although the CAI instrument asked about pipe tobacco, this was not included in the usable case rule because there was only one other question about pipe tobacco in addition to the gate question.

For cigarettes through heroin, respondents are asked a single "yes/no" question for their lifetime use or nonuse. Respondents who initially refuse to answer the gate question are asked a follow-up question to encourage them to reconsider their refusal. The usability criterion for these substances is met if these respondents change their initial refusal to an answer of "yes" or "no" (i.e., they provide a response that would no longer be considered to be a missing value).

For hallucinogens through sedatives, lifetime use or nonuse for the overall category is determined by asking multiple "yes/no" questions about lifetime use or nonuse of specific drugs within the broader category (e.g., LSD within hallucinogens). Consequently, these questions are referred to as "multiple" gate questions. If any of these multiple gate questions are answered as "yes," then the respondent logically is a lifetime user for the overall category (e.g., hallucinogens).

For these multiple gate drug categories, the criterion for usability was considered to have been met if at least one lifetime gate question in the series was answered as "yes" or "no" (e.g., for hallucinogens, if at least one question in the series LS01A through LS01H was answered as "yes" or "no"). This rule was adopted for the multiple gate drug categories because requiring lifetime use or nonuse to be known for the overall category would have placed a more stringent usability requirement on data for nonusers than for users. Specifically, unambiguous identification of lifetime nonusers for the overall category required them to answer "no" to *every* gate question in the multiple gate series because respondents could have been lifetime users of drugs that had missing data. Consequently, respondents who answered some multiple gate questions as "no" and also had some responses of "don't know" or "refused" would fail a usability rule for multiple gate drugs that required lifetime use or nonuse for the overall category to be known unambiguously. In contrast, respondents' status as lifetime users for the overall category was known if they answered "yes" to at least one drug in the series, even if they had given responses of "don't know" or "refused" for other questions in the series.

The types of follow-up questions that were administered in response to initial refusals varied for these modules with multiple gate questions.

- For hallucinogens, respondents were administered individual follow-up questions after refusals to report lifetime use or nonuse of the specific hallucinogens LSD, PCP, and Ecstasy. However, respondents were not asked a follow-up question to determine lifetime use or nonuse of any hallucinogens (i.e., regardless of which specific ones) if they refused to answer all gate questions for hallucinogens (including continued refusal to report lifetime use or nonuse of LSD, PCP, or Ecstasy).
- For inhalants, pain relievers, tranquilizers, and sedatives, respondents who refused to answer all gate questions in a module were asked a follow-up question to determine lifetime use[7] or nonuse of any drugs in the overall category. Respondents were not asked follow-up questions if they refused to answer a specific gate question (e.g., lifetime nonmedical use of Vicodin®, Lortab®, or Lorcet® in question PR03 for pain relievers) but they did not refuse to answer all gate questions in that module.

[7] For pain relievers, tranquilizers, and sedatives, this refers to nonmedical use.

- For stimulants, respondents were administered a follow-up question if they refused to answer the gate question about lifetime nonmedical use of methamphetamine, Desoxyn®, or Methedrine®. Respondents also were administered a follow-up question to determine lifetime nonmedical use or nonuse of any stimulants if they refused to answer all gate questions for stimulants (including a refusal to answer the question about methamphetamine on follow up).

Despite these variations in how refusal follow-up questions were administered in these multiple gate drug modules, the usability principle that was described previously for modules with multiple gate questions applied to these follow-up questions: if respondents refused to answer a question (or series of questions), but then answered "yes" or "no" to the follow-up probe, then they were considered to have met the usability criterion for that module. In particular, if a respondent changed a refusal for lifetime use of Ecstasy to an answer of "yes" or "no," then the respondent was considered to have met the usability criterion for hallucinogens, regardless of whether he or she had missing data for other gate questions in the hallucinogens module.

Table 2.1 lists the follow-up questions in the core modules that were administered when respondents initially refused a gate question, including follow-up questions that were administered when there was a single gate question (i.e., for cigarettes through heroin) and follow-up questions that were administered in modules with multiple gate questions. Table 2.1 also lists the implications for the usable case criteria according to how respondents answered these follow-up questions.

Table 2.1 Effects on the Potential Usable Status of a Case Based on Responses to Follow-Up Questions for Refusals to Gate Questions in the Core Drug Modules

Module or Drug (if Applicable)	Question Number for Follow-Up Question	Consequence if Follow-Up Response Is "Yes" or "No"	Consequence if Follow-Up Response Is DK or REF
Cigarettes	CGREF1	Meets usable case criterion for cigarettes	Not a usable case
Snuff	CGREF3	Meets usable case criterion for snuff	Does not meet usable case criterion for snuff[1]
Chewing Tobacco	CGREF2	Meets usable case criterion for chewing tobacco	Does not meet usable case criterion for chewing tobacco[1]
Cigars	CGREF4	Meets usable case criterion for cigars	Does not meet usable case criterion for cigars[1]
Alcohol	ALREF	Meets usable case criterion for alcohol	Does not meet usable case criterion for alcohol[1]
Marijuana	MJREF	Meets usable case criterion for marijuana	Does not meet usable case criterion for marijuana[1]
Cocaine	CCREF	Meets usable case criterion for cocaine	Does not meet usable case criterion for cocaine[1,2]
Heroin	HEREF	Meets usable case criterion for heroin	Does not meet usable case criterion for heroin[1]

Table 2.1 Effects on the Potential Usable Status of a Case Based on Responses to Follow-Up Questions for Refusals to Gate Questions in the Core Drug Modules (continued)

Module or Drug (if Applicable)	Question Number for Follow-Up Question	Consequence if Follow-Up Response Is "Yes" or "No"	Consequence if Follow-Up Response Is DK or REF
Hallucinogens/LSD	LSREF1	Meets usable case criterion for hallucinogens	Does not affect usable case status for hallucinogens if another gate question (or follow-up question) is answered as "yes" or "no"
Hallucinogens/PCP	LSREF2	Meets usable case criterion for hallucinogens	Does not affect usable case status for hallucinogens if another gate question (or follow-up question) is answered as "yes" or "no"
Hallucinogens/Ecstasy	LSREF3	Meets usable case criterion for hallucinogens	Does not affect usable case status for hallucinogens if another gate question (or follow-up question) is answered as "yes" or "no"
Inhalants	INREF[3]	Meets usable case criterion for inhalants	Does not meet usable case criterion for inhalants[1]
Pain Relievers	PRREF[3]	Meets usable case criterion for pain relievers	Does not meet usable case criterion for pain relievers[1]
Tranquilizers	TRREF[3]	Meets usable case criterion for tranquilizers	Does not meet usable case criterion for tranquilizers[1]
Stimulants/Methamphetamine	STREF1	Meets usable case criterion for stimulants	Does not affect usable case status for stimulants if another gate question (or follow-up question) is answered as "yes" or "no"
Stimulants	STREF2[3,4]	Meets usable case criterion for stimulants	Does not meet usable case criterion for stimulants[1]
Sedatives	SVREF[3]	Meets usable case criterion for sedatives	Does not meet usable case criterion for sedatives[1]

DK = don't know; REF = refused.

[1] Overall status as a usable case will still be met if the usable case criteria are met for a sufficient number of other substances or modules.

[2] The interview includes a follow-up probe if the gate question for crack cocaine is refused (CKREF) but crack cocaine is not included in the criteria for identifying usable cases.

[3] Follow-up probe is asked if the respondent refused to answer all gate questions for that module.

[4] Respondents who are routed to STREF2 also had refused the methamphetamine follow-up question STREF1.

The usable case rule was a necessary, but not sufficient, requirement for a case to be considered a final respondent. Cases that had sufficient data to meet the usable case criteria could still be treated as nonrespondents if their interview data appeared to be of poor data quality, as evidenced by potential response pattern problems (Section 2.3.2).

2.2.4 Creation of Cumulative Quarterly Unedited Data Files

Following daily processing of the data, each day's SAS dataset was merged with the cumulative data that had been transmitted up to that point in the quarter. At the end of the quarter, a complete data file was produced that contained all cases that had been transmitted during the quarter. Each quarterly data file then underwent additional initial cleaning and processing (prior to the editing procedures) to modify or correct field errors, such as erroneous ID entries by the field staff. The cleaned-up (but otherwise unedited) SAS datasets from the first two quarters (also known as 6-month data) and from all four quarters (also known as 12-month data) served as the usual starting points for the subsequent logical editing procedures that are described in Chapters 4 through 11.[8]

2.3 Preliminary Editing and Coding of Processed Interview Data

In addition to procedures that were described in Section 2.2 following receipt of transmitted data from the field, preliminary coding and processing of unedited interview data encompassed the following activities:

- coding of "OTHER, Specify" data (Section 2.3.1),
- investigation of response patterns in records that otherwise met the usable case criteria (Section 2.3.2), and
- edits to "date-dependent" variables (if applicable) when the interview date was judged to be questionable (Section 2.3.3).

The first two of these activities could occur or did occur prior to creation of the cumulative quarterly unedited data files (Section 2.2.4). Edits to date-dependent variables were not performed until final interview dates had been created for respondents (Section 4.2.1).

Note that a code to denote "bad data" (i.e., to indicate an inconsistency or some other problem in the original data) could be assigned to variables during any of these activities. The following codes were assigned to denote "bad data": 85 (or 985, or 9985, etc.) = BAD DATA Logically assigned. Codes for bad data were treated as missing values. Any assignment of bad data codes was done in subsequent editing steps, not as part of the nightly processing of transmitted data.

2.3.1 Coding of "OTHER, Specify" Data

This activity took alphanumeric (text) answers that respondents or interviewers had typed (e.g., specific other drugs used, specific other payment sources of treatment) and converted them into numeric codes. These alpha answers (and the numeric codes resulting from them) are referred to as "OTHER, Specify" data.

---

[8] Edits are run on preliminary data from the first two quarters of a given survey year to identify any updates that need to be made to the programs for use with the full data from all four quarters. Running the edits on data from the first two quarters is particularly useful for testing the programs for any sections of the CAI instrument that are new or have changed since the previous survey year.

Coding of the "OTHER, Specify" variables was accomplished through computer-assisted procedures.[9] "OTHER, Specify" responses were first converted to all capital letters because respondents could use different combinations of uppercase or lowercase characters to provide an otherwise identical response. If an exact match was found between what the respondent keyed and an entry in the data dictionary (e.g., "ALCOHOL"), the computer-assisted procedures assigned the appropriate numeric code (e.g., 807 for alcohol). The system could also accommodate commonly encountered misspellings (e.g., "ALCHOHOL").

Typed answers that the respondent provided that did not match an existing data dictionary entry were reviewed by analysts on a flow basis during the quarter to determine whether an existing code should be assigned to the response or a new code should be created. Based on these decisions, new entries were added to the relevant dictionaries on a quarterly basis—including new dictionary entries corresponding to existing codes—for use in daily processing of data that were transmitted from the field in subsequent quarters. Analysts could also decide not to add a particular response to the data dictionary, in which case the response would be output for review and coding on a case-by-case basis if it was reported in a future quarter or survey year. In addition, analysts reviewed the codes that were assigned through the computer-assisted process to verify that these codes were being assigned correctly. Over time, these procedures have reduced the turnaround time and burden on analysts for producing the coded "OTHER, Specify" data.

Particularly for drugs and tobacco brands (Sections 2.3.1.1 and 2.3.1.2, respectively), which used the same data dictionaries for a number of "OTHER, Specify" variables, matching a written response to a numeric code in the data dictionary would result in that code being assigned *no matter where* respondents typed that answer. For example, a given entry in the data dictionary for drugs would be assigned the same code if it appeared in the hallucinogens module one year and appeared in the pain relievers module another year. Similarly, an entry in the data dictionary for tobacco brands would be assigned the same code regardless of the type of tobacco where the response appeared. In some situations, however, the same response could have different meanings depending on the context. For example, a given tobacco brand name with no other associated information could apply to a brand of cigarettes or to a brand of cigars, depending on whether it was specified as a cigarette brand or as a cigar brand.

As with the "enter all that apply" data that were discussed previously, respondents could answer the "OTHER, Specify" questions as "don't know" or "refused," which were then reassigned to the respective codes of 9994 or 9997, as described in Section 2.2.1. Respondents could also type in an equivalent response to "don't know" (e.g., "no idea") or "refused" (e.g., "too personal"). These equivalent responses were assigned the relevant codes for missing data (Section 2.2.1) as part of the coding procedures. For typed responses that were nonsensical or otherwise nonresponsive to the request to specify additional information, codes for bad data (Section 2.3) were assigned.

---

[9] A system has been in place since 2002 for the daily coding and processing of the "OTHER, Specify" variables for drugs and tobacco brands. This system now encompasses the daily coding and processing of *all* "OTHER, Specify" variables from the survey that underwent assignment of numeric codes, except for codes pertaining to the industry in which respondents were employed and their current or former occupations. Coding of industry and occupation data was handled by the U.S. Census Bureau.

The remainder of the discussion in this section focuses on issues related to coding of the "OTHER, Specify" data according to the type of other information that was requested from respondents:

- other drugs (Section 2.3.1.1),
- other tobacco brands that respondents used most commonly in the past 30 days (Section 2.3.1.2),
- other race or ethnicity (Section 2.3.1.3), and
- additional "OTHER, Specify" data in noncore sections (Section 2.3.1.4).[10]

Except for mentions of other drugs (Section 2.3.1.1), "OTHER, Specify" data in the 2014 NSDUH typically were intended to capture a single "other" response (e.g., most important other reason for not receiving mental health treatment in the past 12 months). If respondents typed in responses for which multiple codes could apply (e.g., multiple reasons for not receiving mental health treatment instead of the single most important reason), the standard procedure was to assign a code to the first response that could be coded. Relevant sections elsewhere in this report indicate when exceptions were made to this more general approach (e.g., coding of "OTHER, Specify" data for youth mental health service utilization, described in Section 7.4.16).

2.3.1.1 Other Drugs

In the core modules for hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives, respondents could report lifetime use or nonmedical use of drugs other than those that were specifically asked about in the respective modules. Respondents also could report lifetime injection of drugs other than heroin, cocaine, methamphetamine, or other stimulants in the noncore special drugs module. In the noncore substance treatment module, respondents could report current or most recent treatment or the perceived need for treatment in the past 12 months for drugs other than those for which they reported lifetime use in the core modules.

In the "OTHER, Specify" drug questions for both core and noncore modules, respondents could enter up to five responses (in five data entry fields) for other drugs. For the most part, respondents specified a total of only one drug in these questions or they specified only one drug in each field. Neither of these situations required any special handling of the data.

However, additional procedures were required to handle the following situations in the "OTHER, Specify" drug data.

1. Respondents specified more than one drug within a single data entry field but they specified a total of five or fewer drugs across the five available fields.
2. Respondents specified a total of more than five substances across the five available data entry fields (i.e., by definition, they specified multiple drugs within one or more fields).

[10] See Section 1.1 and Tables 1.1 and 1.2 in Chapter 1 for details about the core and noncore structure of the NSDUH questionnaire and contents of the core and noncore sections.

In the first situation, codes for the additional drugs were moved to the next unused "slot(s)" (i.e., variable[s]). For example, if a total of three drugs were specified, with two of them being specified in the first field and the third being specified in the second field, the additional code from the first field was moved to the third "OTHER, Specify" variable. Consequently, the responses in the individual "OTHER, Specify" drug variables did not always correspond to the order in which respondents reported use of these drugs.

When respondents specified more than five substances in the available fields, duplicate mentions of drugs were identified and removed. Duplicate mentions could include the same drug being mentioned more than once in the "OTHER, Specify" data or a drug being reported in the gate questions for a given module and also being reported in the "OTHER, Specify" data (e.g., if question LS01a for LSD was answered as "yes" and LSD also was reported in the "OTHER, Specify" data for hallucinogens). If more than five unique mentions of drugs remained after any redundant mentions had been eliminated, further priority in retaining responses in the final drug codes was given to (1) mentions of illegal or prescription-type drugs, as opposed to "over-the-counter" (OTC) drugs that are legally available without a prescription and (2) mentions that were relevant to the category of interest (e.g., in the hallucinogens section, giving priority to mentions of hallucinogens over mentions of drugs that were not classified as hallucinogens).

In rare instances, the procedures described previously still did not yield five or fewer unique drugs in the "OTHER, Specify" data for a given module. In these situations, the highest priority was given to retaining codes for drugs that had no approved medical use in the United States or were deemed to have a greater potential for dependence or abuse, and the lowest priority was given to retaining codes for drugs that were not classified as controlled substances.[11] In the extremely rare situations where five or more unique drugs remained after all of these measures had been applied, the first five remaining "OTHER, Specify" codes were retained and mentions of any additional drugs were dropped from the coded data that were available for further editing or analysis.

2.3.1.2 Other Tobacco Brands

The CAI instrument included questions to identify the specific brands of tobacco that were used most commonly by respondents who reported use in the past 30 days of cigarettes, chewing tobacco, snuff, or cigars. For these four types of tobacco products, respondents who reported use in the past 30 days could choose from a list of brands shown on the computer screen[12] or they could indicate use of "a brand not on this list." Respondents who gave the latter answer were asked to type in the name of the specific other brand that they used.

The basic coding scheme for the "OTHER, Specify" tobacco brand categories was as follows.

---

[11] The drug scheduling classifications, Schedules I through V, under the Controlled Substances Act were used in making these determinations. See www.deadiversion.usdoj.gov/21cfr/21usc/index.html and www.deadiversion.usdoj.gov/schedules/index.html for details.

[12] For cigarettes, the listing of brands was split between two different computer screens.

- Codes of 101-199 and 1001-1999 were reserved for cigarette brands.
- Codes of 201-299 and 2001-2999 were reserved for chewing tobacco brands.
- Codes of 301-399 and 3001-3999 were reserved for snuff brands.
- Codes of 401-499 and 4001-4999 were reserved for cigar brands.
- Codes of 501-599 and 5001-5999 were reserved for pipe tobacco brands.
- Codes in the 601-699, 701-799, and 801-899 series were reserved for miscellaneous tobacco and nontobacco responses.

Codes were assigned to the "OTHER, Specify" tobacco brand data according to these categories, regardless of whether the response came from the section for cigarettes, snuff, chewing tobacco, or cigars within the tobacco module. This coding scheme was particularly relevant for the smokeless tobacco data for snuff and chewing tobacco, where snuff brands could be reported as the brand of chewing tobacco that was used most often in the past 30 days, or vice versa. Thus, if a respondent specified a brand of snuff in the chewing tobacco section, the "OTHER, Specify" response for the chewing tobacco brand was assigned a code in the 300 or 3000 series for snuff brands. Similarly, if a respondent specified that the brand of other cigarettes that he or she smoked most often in the past 30 days was actually a brand of little cigars, the "OTHER, Specify" response for cigarette brands was assigned a code in the 400 or 4000 series for cigars.

Note that the coding for a particular tobacco brand did not capture further details, because the main aim in the coding was to capture information about any use within a particular brand label. For example, the code for a particular brand did not capture details such as length (e.g., for cigarettes), size or shape (e.g., for cigars), or flavor varieties.

2.3.1.3 Other Race or Ethnicity

In the interviewer-administered core demographics module (see Chapter 4), NSDUH respondents were asked about their Hispanic/Latino origin and race information in QD04 and QD05, respectively. If respondents reported in QD05 that they were Asian, they were asked in QD05ASIA to report which Asian group best described them. However, they could consider that the categories presented to them for Hispanic/Latino origin, race, or Asian ethnicity did not apply to them. In these situations, respondents reported their "other" Hispanicity, race, or Asian ethnicity to the interviewers, who then typed in the respondents' answers.

The computer-assisted coding procedures and use of data dictionaries that were described in Section 2.3.1 applied to these "OTHER, Specify" data for race and ethnicity as well. In most cases, new responses were just new misspellings of an already established category, such as a response of "Porto Rican" instead of "Puerto Rican."

Regardless of the source (QD04, QD05, QD05ASIA) for these "OTHER, Specify" data, the write-in responses were used in subsequent editing of Hispanicity and race to determine the respondents' final Hispanicity and race (see Sections 4.2.7, 4.3.2, 4.3.3, and 4.3.4). Thus, in coding the "OTHER, Specify" data, each write-in was assigned two codes, one for race and the other for Hispanicity. If an interviewer entered both a geographical entity and a group within a particular race in the "OTHER, Specify" response, such as "Black Cape Verdean," the geographical entity was ignored in the race code and the respondent was coded as "Black/African American" for the race code. The geographic information Cape Verde was captured in the Hispanic/Latino code for "Cape Verde."

2.3.1.4 Additional Noncore "OTHER, Specify" Data

There were three types of "OTHER, Specify" questions in the noncore self-administered or interviewer-administered sections:

- those where the "OTHER, Specify" item was a follow-up to a lead question that typically was answered as "yes" or "no";[13]
- those where the "OTHER, Specify" item was a follow-up to a response category for "other" in an "enter all that apply" question; and
- those where respondents did not get the opportunity to choose the "other" response (and specify something) if they already chose another category from the list.

Coding of these noncore "OTHER, Specify" variables was performed according to the general principles described in Section 2.3.1. Otherwise, minimal additional decision making was involved in assigning codes to the responses for these variables. Descriptions of these variables are included here for completeness.

An example of the first type of "OTHER, Specify" question is question TX42JSP in the noncore self-administered substance treatment module (i.e., specify other source that paid/will pay for the last or current substance treatment). Only those respondents who reported in question TX42J that some other source paid for their last substance treatment or counseling or that some other source would pay for their current treatment (TX42J = 1) were routed to TX42JSP and asked to report the other payment source. Respondents could report other sources of payment for their treatment in questions TX42A to TX42I and also report some other payment source in TX42J.

Question ADMT15 in the noncore self-administered adult mental health service utilization module (i.e., specify the other location where outpatient mental health treatment was received in the past 12 months) is an example of the second type of "OTHER, Specify" question. Adult respondents aged 18 or older could report in question ADMT14 that they received mental health treatment or counseling in the past 12 months in up to six different types of outpatient settings, including "some other place." Adults were routed to ADMT15 if they reported receiving outpatient treatment in some other place in ADMT14. Because ADMT14 was an "enter all that apply" question, respondents could choose any of the specific locations that were listed in the question (e.g., an outpatient mental health clinic or center) and also report receipt of mental health treatment in some other place.

The third type of "OTHER, Specify" question is represented by the following three questions:

---

[13] Depending on the nature of the lead question, either an affirmative or a negative response to the lead question could govern whether respondents were asked to specify something.

- SD16SP, which was associated with question SD16 (how respondents obtained their last needle for injecting drugs) in the self-administered special drugs module;
- TX25SP, which was associated with question TX25 (main place where the respondent received or was currently receiving substance abuse treatment in the past 12 months) in the self-administered substance treatment module;[14] and
- QD24SP, which was associated with question QD24 (reason for leaving school without getting a high school diploma) in the interviewer-administered noncore education module.

Respondents first were presented with a list of options in the "lead" question (i.e., SD16, TX25, or QD24), including an option for "other" (e.g., other reason in QD24). If respondents chose any response from the list of options in the lead question except for "other," they were *not* routed to the "OTHER, Specify" question. Rather, data from the lead question and the specify question were combined into a single final variable.[15] When respondents chose the other category in the lead question, but they specified something that was coded as a missing value (i.e., don't know, refused, bad data, or blank), then the final edited variable (e.g., GNNDGET) retained a code corresponding to other, as opposed to being assigned a missing value.

The edits applied to GNNDGET, TXLTYMN, and LFSCHWHY (including similar types of questions that could be added in the future) were designed to provide analysts with a *standardized* way to readily identify when it could be logically inferred that respondents should have chosen a given response option from the preceding question (i.e., rather than "other"). For GNNDGET as an example, codes of 1 to 4 applied to answers that respondents gave directly from question SD16 (e.g., 1 = Bought the needle from a pharmacy). Category 5 in SD16 was "Got the needle some other way." Although the coding sequence for "OTHER, Specify" responses in question SD16SP could have resumed at number 6 for responses corresponding to "Bought the needle from a pharmacy," assigning a code of 6 for "OTHER, Specify" responses that corresponded to category 1 in question SD16 would not enable analysts to readily see the logical connection between the "OTHER, Specify" response and the available response choice in the question that preceded it.

In this example, an alternative to enable analysts to see the logical connection between reports of buying the needle from a pharmacy in SD16 and corresponding reports in SD16SP would be to assign a code of 11 to responses in SD16SP that corresponded to category 1 in SD16. However, this coding scheme cannot be applied to question QD24, because QD24 listed 15 possible reasons for leaving school, not including "other reason." A code of 11 could not be used in identifying responses in QD24SP that corresponded to category 1 in QD24 ("School was boring or I didn't want to be there"), because that code was reserved in the edited variable LFSCHWHY for responses of 11 in QD24 (i.e., Moved here from another country and didn't enroll [or dropped out] because of language or other problems).

[14] Although question TX04ISP (specify other location where the respondent received treatment for alcohol or other drug use in the past 12 months) was preceded by a "yes/no" question, TX04ISP used the same codes as TX25SP to allow the same data dictionary to be used for processing both of these "OTHER, Specify" variables (see Section 2.3.1).

[15] The final edited variables in 2014 were GNNDGET (based on data from SD16 and SD16SP), TXLTYMN (based on data from TX25 and TX25SP), and LFSCHWHY (based on data from QD24 and QD24SP).

In consideration of this issue, responses in these "OTHER, Specify" variables that corresponded to existing response categories in their respective lead questions were coded starting with the number 21, with the coding proceeding in the order of the existing response categories. This procedure prevented overlap of "OTHER, Specify" codes with available responses when lead questions had 10 or more available choices, as was the case with QD24, and offered analysts a standardized way to interpret these values. For example, if analysts wanted to treat codes of 21 to 24 in GNNDGET that came from "OTHER, Specify" data as being equivalent in an analysis to the corresponding codes of 1 to 4, the analyst could simply subtract 20 from any codes with values of 21 to 24 to recode these values to the corresponding codes of 1 to 4; the same procedure could be applied if analysts wanted to treat codes of 21 to 35 in LFSCHWHY as being equivalent to codes of 1 to 15.

For similar reasons, the coding sequence for responses in these "OTHER, Specify" variables that did not correspond to responses from the list in the lead question resumed at number 41 (e.g., for GNNDGET, 41 = Given by/stolen from friend/acquaintance of friend/nonrelative). Although codes for responses that did not correspond to available choices from a lead question could have resumed at a lower number—especially for SD16SP and TX25SP—this approach standardized the assignment of codes across these types of "OTHER, Specify" variables (i.e., new codes always started at 41), minimized the risk of overlap between codes for these two types of responses, and allowed flexibility if new response choices were added to a lead question in a future year.

2.3.2 Investigation of Response Patterns in Usable Records

Although conversion to CAI reduced or eliminated some data quality problems that could occur in a PAPI format, it also was recognized that the audio computer-assisted self-interviewing (ACASI) environment could encourage some respondents to use the computer keyboard to enter nonsensical patterns of answers if they were not paying attention to questions or were not taking the interview seriously for other reasons. Thus, even if a respondent had sufficient data to meet the usable case criteria described in Section 2.2.3, certain patterns of answers could call into question the overall validity of the respondent's data.

In response, a data diagnostics program was developed to screen for the following patterns of responses that might raise questions about the validity of the interview as a whole:

- high numbers of "yes" responses to lifetime use of specific hallucinogens, inhalants, or psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, or sedatives), which might indicate that respondents were indiscriminately keying data without paying attention to what they were entering;
- alternating "yes/no" responses to questions about lifetime use of specific hallucinogens, inhalants, or psychotherapeutics (or alternating patterns of "response entered/not entered" in the psychotherapeutics sections), which might indicate some type of pattern-making;
- high numbers of illegal drugs that respondents reported using every day or just about every day in the past year or past month (where applicable), in which case one might

question either the validity of the answers or the respondent's competence to complete the interview;

- high numbers of substances that respondents reported first using at age 1 or 2, which might indicate indiscriminate keying of 1s or 2s, especially given that the age-at-first-use questions followed gate questions where a response of 1 denoted "yes" and a response of 2 denoted "no"; and
- consistent keying of the same code (either 1 or 2) throughout one or more modules (i.e., after lifetime use has been reported), which would suggest a pattern of indiscriminate answering.[16]

These patterns of responses were examined on a case-by-case basis to determine whether a case should be retained as a final respondent or dropped.

Fewer than 10 cases in 2014 met the usable case criteria but were treated as nonrespondents because their responses were of questionable validity, based on one or more of the patterns described above. In addition, fewer than 10 cases in 2014 were retained as respondents, whose original responses to questions in one or more core drug modules were replaced with bad data codes. This process included setting responses to bad data that indicated they were lifetime users of a given drug. For example, data for some respondents were set to bad data because the respondents keyed values of 1 or 2 to every question that was asked in a module, beginning with the age-at-first-use question. Data for the lifetime variables for these cases were set to bad data as part of the edits for the lifetime drug use variables (Section 6.2.1). For example, if a case was identified that had "bad" stimulants data, the lifetime stimulants data corresponding to responses in questions ST01 through ST05 were set to bad data as part of the lifetime drug use edits, and a flag was set to indicate that data subsequently needed to be set to bad data for related variables pertaining to nonmedical use of prescription-type stimulants (both core and noncore variables pertaining to stimulants).

2.3.3 Edits to "Date-Dependent" Variables

The CAI instrument used the interview date information that was stored by the computer to create "date fills" during the interview that indicated the starting dates for questions pertaining to the "past 30 days" and "past 12 months." Specifically, the starting date for the past 30-day (or 12-month) period was calculated as exactly 30 days (or exactly 12 months) prior to the stored

[16] An important change to the CAI instrument since 2001 is that response categories for certain consistency checks use 4=yes/6=no instead of the format of 1=yes/2=no that was used prior to 2001. For example, if a respondent reported first using marijuana at age 1 or 2, the respondent could not use a response of "1" or "2" in question MJCC05 to verify that this age at first use was correct. This change to the CAI instrument was designed to stop respondents if they had been engaged in a pattern of keying responses of "1" or "2."

interview date.[17] Thus, in the recency-of-use questions that asked respondents when they last used the drug of interest, the response category "within the past 30 days" included a date fill to remind respondents when the past 30-day reference period began for them. Similarly, introductions to specific questions about frequency of use of a particular drug in the past 12 months and past 30 days included date fills to remind respondents of the period they should be thinking about when answering these questions.

Data that could be affected by questionable interview dates were edited as needed once the edited interview date variable INTDATE had been created for all final respondents (Section 4.2.1). As part of the procedures for creating INTDATE, an indicator variable (EIIDATE) was created that specified how the final interview date was assigned. For example, EIIDATE could indicate that the interview date that had been stored by the CAI program during the interview had an incorrect year or was outside of the quarter in which the interview had been completed.

With the exception of 2008, however, no cases since 2004 have had an originally entered interview date that was sufficiently problematic to call into question the respondent's answers to date-dependent questions in self-administered sections of the interview. If this problem were to occur, subsequent data in the self-administered modules that were dependent on the interview date would be considered problematic. For example, if the CAI program calculated a 30-day reference period based on an incorrect interview date, answers that the respondent gave on the number of days that he or she used different drugs in the past 30 days could reflect use of these drugs in a period other than the intended 30-day reference period.

This interview date issue did not present a problem if respondents never used a particular drug (or had never engaged in other behaviors). Moreover, if the respondent reported never having engaged in a particular behavior, the CAI program skipped that person out of questions where the interview date would be important for establishing reference periods. For these reasons, some cases where there was some question about the interview date could still be retained as final respondents.

Instead of cases being dropped, the edits that are in place would set the following types of self-administered questions to bad data if respondents were routed to them:

- questions pertaining to behaviors in the past 30 days;
- questions pertaining to behaviors in the past 12 months;
- questions pertaining to the most recent time that an event occurred (e.g., when a respondent last used a drug of interest); and
- questions pertaining to the respondent's age when an event occurred (e.g., the age when the respondent first used a drug of interest).

---

[17] If a respondent broke off and then resumed the interview at a later date, the "fills" in the interview that defined the past 30-day and past 12-month reference periods were updated for remaining questions when the interview resumed. This was done because 30 days and 12 months from the date when an interview resumed might be later than the 30-day and 12-month periods when a respondent had previously been asked questions. Thus, the "fills" that defined these reference periods during the interview were dynamic rather than static. If a respondent resumed the interview and went back to a section that he or she had completed prior to the breakoff, the date "fill" would be reset based on the new date when the interview resumed. In practice, however, the tendency is to proceed with the parts of the interview that have not been completed rather than to go back to earlier sections.

Self-administered questions about age at first use and ages when other events occurred were not related directly to the interview date but were related indirectly via respondents' ages; that is, respondents' ages were calculated by comparing the date of birth with the interview date. In turn, age-at-first-use and other age data in the self-administered modules were compared for consistency with the respondents' ages.

For the types of questions that were described previously, respondents' answers would be set to bad data before any further editing was done. Setting the responses to bad data prior to any subsequent editing allowed analysts during logical editing to distinguish between situations where the data were deemed to be bad prior to any further editing being done and situations where a variable might be set to bad data in subsequent editing steps because of inconsistencies with other data in the respondent's record (Section 2.4.3).

This edit would not apply to gate questions that asked whether a respondent had ever engaged in a behavior (e.g., "Have you ever smoked part or all of a cigarette?"). As noted previously, whether the respondent had ever engaged in a particular behavior prior to being interviewed is not dependent on the value stored for the interview date. This edit also does not apply to questions within a module that asked whether a more detailed behavior of interest was *ever* true for this respondent (e.g., "Have you smoked at least 100 cigarettes in your entire life?"). The decision also was made that this edit would not apply to sections of the interview that were administered directly by the field interviewers (FIs).

2.4 General Issues and Principles for Editing NSDUH Data

Previous sections discussed broader data processing and coding procedures in the data as a whole or in large sections, such as the core section of the interview. In contrast, this section focuses on the general principles of data processing and editing across and within modules, once the initial processing of transmitted interview data had been completed (i.e., the procedures described in Section 2.2) and after cases with patterned responses or questionable interview date information had been identified (Section 2.3).

The following specific issues are addressed in this section:

- editing across modules (Section 2.4.1);
- assignment of relevant "not applicable" codes (Section 2.4.2);
- additional assignment of NSDUH missing data codes (Section 2.4.3);
- assignment of codes to edited "enter all that apply" variables (Section 2.4.4);
- "back-editing" based on "OTHER, Specify" data (Section 2.4.5); and
- general principles for the "flag and impute" procedures for core drug variables (Section 2.4.6).

2.4.1 Editing across Modules

An important principle that was followed in editing the data was that responses from one module (e.g., hallucinogens) generally were not used to edit variables in another module (e.g., inhalants).[18] In particular, the noncore self-administered data (special drugs module through the consumption of alcohol module) were generally *not* used to edit related variables in the core self-administered modules (tobacco through sedatives).[19] For example, if a respondent reported in the core heroin module that he or she last used heroin more than 12 months ago, but then reported in the noncore special drugs module that he or she last used heroin with a needle more than 30 days ago but within the past 12 months, then the core heroin recency variable HERREC was *not* edited to take into account this more recent indication of heroin use from the noncore module. Rather, HERREC retained the respondent's initial response that he or she last used heroin more than 12 months ago. Consequently, the documentation for the noncore variables in NSDUH codebooks includes a footnote to indicate that these variables may be inconsistent with data in other sections of the interview.

This principle of not using noncore data to edit core data was important for maintaining consistent data to assess trends in substance use. If variables in core modules were allowed to be edited based on respondents' answers in the noncore modules, key drug use estimates could change across years as noncore questions or modules were added or deleted.

One exception to this principle involved situations in which responses in one or more modules governed whether respondents were asked questions in a later module. For example, the substance treatment module was relevant only for respondents who reported some lifetime use of alcohol or other drugs, excluding tobacco products. Respondents who reported in the core modules that they had never used alcohol, illicit drugs, or prescription-type psychotherapeutics for nonmedical reasons (i.e., pain relievers, tranquilizers, stimulants, or sedatives) were not asked corresponding follow-up questions in the substance treatment module. In such cases, during the editing process, blank values in the substance treatment variables were replaced with codes to indicate that respondents were not asked the follow-up questions in this module because they reported never having used any of the relevant core drugs.

2.4.2 Assignment of Relevant "Not Applicable" Codes

Because the CAI logic controlled whether respondents were skipped out of some questions based on their answers to previous questions, an important aspect of editing the NSDUH data involved replacing missing data codes in the unedited data with appropriate codes to indicate that the questions had been skipped because they did not apply. The following codes were assigned when respondents were skipped out of a given question *and it could be determined unambiguously* that the respondent had legitimately been skipped out of the question, based on the answer(s) to one or more previous questions:

- 91 (or 991, or 9991, etc.) = NEVER USED [DRUG(s) OF INTEREST];
- 93 (or 993, or 9993, etc.) = USED [DRUG] BUT NOT IN THE PERIOD OF INTEREST; and
- 99 (or 999, or 9999, etc.) = LEGITIMATE SKIP.

---

[18] See Section 1.1 in Chapter 1 for a definition of modules.

[19] An exception to this principle that is discussed in Chapter 6 involved the editing of core data on use of methamphetamine to incorporate questions about methamphetamine use that have been included in the noncore special drugs module since 2006.

Strictly speaking, codes of 91 and 93 in the CAI data could be considered variants of the more generic legitimate skip code. Their use was designed to provide analysts with more information about the reason that respondents were skipped out of a particular question.

Codes of 91 and 93 were used most often in the core drug sections of the interview. For example, a code of 91 (or 991, etc.) in the marijuana section denotes the pattern where respondents were skipped out of all remaining marijuana questions because they answered "no" to the lifetime marijuana question MJ01. Similarly, a code of 93 in the marijuana section denotes situations where respondents were lifetime marijuana users but were definitely not users in the past 30 days or past 12 months or both.

Codes of 91 and 93 also were used to a limited extent in noncore sections of the interview because the CAI logic took into account respondents' prior answers to core drug use questions to determine whether particular noncore questions applied. For example, questions about cocaine in the substance dependence and abuse module were relevant to respondents who used cocaine in the 12 months prior to the interview. Thus, if a respondent last used cocaine more than 12 months prior to the interview, a code of 93 in the substance dependence and abuse variables pertaining to cocaine would signify to an analyst why the CAI program skipped the respondent out of these questions. Similarly, a code of 91 in the substance dependence and abuse variables for cocaine signified to an analyst that the respondent was skipped out of these questions because he or she had never used cocaine.

A legitimate skip code of 99 was used most often in the noncore self-administered sections of the interview or in interviewer-administered sections. For example, the youth experiences module was intended to be administered only to respondents aged 12 to 17. If a respondent was 18 or older, a code of 99 was assigned in the editing process to the skipped youth experiences variables. Similarly, if a respondent had used alcohol or some other drug at least once in his or her lifetime, but answered the lifetime substance treatment question TX01 as "no," then the CAI program skipped the respondent out of all remaining substance treatment questions. A code of 99 was assigned to the skipped substance treatment variables in this situation to signify that the respondent had used alcohol or drugs at least once but had never received substance abuse treatment.

The following codes also were assigned through editing:

- 81 (or 981, or 9981, etc.) = NEVER USED [DRUG(s)] Logically assigned;
- 83 (or 983, or 9983, etc.) = USED [DRUG(s)] BUT NOT IN THE PERIOD OF INTEREST Logically assigned; and
- 89 (or 989, or 9989, etc.) = LEGITIMATE SKIP Logically assigned.

These codes were given values in the 80s to signify that existing values were to be overwritten during the editing process. For example, if a respondent was somehow routed into the youth experiences module, but that respondent was subsequently classified as being 18 or older (Section 4.2.2), then any answers that the respondent gave in the youth experiences module were overwritten with a code of 89 (or 989, etc.). These codes signify that this adult respondent logically was not eligible to be asked the youth experiences questions.

However, the codes described in this section were assigned *only* in situations where there was total certainty that a respondent should have skipped a question. For example, if a respondent reported in question MJ01 that he or she had never used marijuana, it was absolutely clear that subsequent questions about marijuana use (e.g., age at first use of marijuana) did not apply. See Section 3.1 for a description of imputation indicators, including a brief discussion of the codes assigned to respondents whose imputed values meant that they legitimately skipped out of the corresponding questions.

The CAI skip logic often treated responses of "don't know" or "refused" to gate questions as equivalent to a negative response. For example, if a respondent was uncertain whether he or she had ever used marijuana (and answered question MJ01 as "don't know"), the CAI program skipped the respondent out of all remaining questions about marijuana use, as though the respondent had never used it. From the standpoint of respondent burden, this makes sense. There is little value in asking a respondent who did not know whether he or she had ever used marijuana, "How old were you the first time you used marijuana or hashish?" Implicit in this question is that respondents have used marijuana at least once in their lives.

Although the CAI program skipped respondents out of questions if they answered a gate question as "don't know" or "refused" (or gave similar answers on follow-up if they initially refused to answer a gate question), these types of responses to a gate question are ambiguous and do not provide an analyst with conclusive information one way or the other about the behavior or event of interest. Consequently, such responses could be thought of as *potentially* affirmative responses, as opposed to inferring that they are negative responses. In particular, as noted previously, respondents who initially refused to answer a question about their lifetime use or nonuse of a drug were given a second opportunity to answer the question as "yes" or "no." Similarly, if a respondent who initially did not know whether he or she had ever used a drug had thought about the issue further, the respondent may have recalled a time when he or she in fact had used it—and more detailed questions about use of the drug would have been relevant for this respondent. Alternatively, if the respondent gave more thought to the issue and decided that he or she really should answer the lifetime drug use question as "no," an analyst would have a solid basis for determining that subsequent questions did not apply.

Further, the procedures for statistically imputing missing data did not automatically infer lifetime nonuse when respondents provided ambiguous information about whether they had ever used a given drug. Rather, such respondents were eligible to be statistically imputed to be lifetime users or nonusers. For these reasons, variables retained missing values in the editing procedures when questions were skipped due to respondents answering a lead question as "don't know" or "refused" (or answering in a similar manner in response to a follow-up probe).

2.4.3 Additional Assignment of NSDUH Missing Data Codes as Part of Editing within a Module

Previous sections described the procedures for assigning missing data codes in the interview data as a whole or in large sections, such as the self-administered core and noncore sections, when the interview date was questionable (Section 2.3.3). This section discusses additional principles for assigning (or retaining) missing data codes as part of the editing procedures for a given module.

In particular, if respondents refused a single or multiple gate question or questions that governed a skip pattern in a module, refusal codes were assigned to all of the subsequently skipped items in the module as part of the editing procedures (i.e., the lead refusal was propagated); that is, it was logically inferred that the respondents were globally refusing to answer *any* questions on that topic.

This propagation did not occur when respondents answered a gate question or questions as "don't know." Rather, values of "blank" (no answer) were retained in the skipped questions. Unlike the situation for responses of "refused," it does not follow logically that a response of "don't know" to a gate question would imply that the respondent would answer "don't know" to all subsequent questions on that topic. For example, if a respondent answered the lifetime marijuana question as "don't know," assigning a "don't know" code to the age-at-first-use variable (corresponding to question MJ02) would imply that the respondent was a lifetime user but did not recall the age when he or she first used.

In addition, data sometimes were identified that were inconsistent with other data in a respondent's record. For example, if a respondent reported first using a drug at an age greater than his or her current age, the CAI program indicated to the respondent that this age at first use was inconsistent. The respondent was prompted to revise the age at first use, his or her current age, or both, to make the data consistent. As noted in Chapter 1, however, respondents did not always resolve these types of inconsistencies. If the age at first use was still inconsistent with the respondent's age, despite the opportunity that the respondent had to resolve the inconsistency, then a bad data code was assigned to the age-at-first-use variable to indicate that the data were inconsistent with other data. As was the case with prior survey rounds, the following codes were assigned to denote "bad data" (i.e., usually inconsistent with other data): 85 (or 985, or 9985, etc.) = BAD DATA Logically assigned.

Other situations where bad data codes were assigned are discussed below in connection with specific steps in the editing process.

2.4.4 Assignment of Codes to Edited "Enter All That Apply" Variables

As noted in Section 2.2.2, the initial creation of separate variables for "enter all that apply" questions involved assignment of the relevant code that was shown to respondents during the interview for that question, or a code of 98 (blank) if respondents did not choose that item from the list; these procedures were implemented for "enter all that apply" questions anywhere in the entire interview. The additional edits described in this section were implemented within modules that had "enter all that apply" variables. In 2014, most variables were coded as 1 if respondents chose that item from the list, and values of 98 were recoded with a value of 6. Documentation for these edited values for "enter all that apply" variables was as follows:

- 1 = Response entered, and
- 6 = Response not entered.

For example, if a respondent reported lifetime nonmedical use of codeine, the code of 4 that was assigned to the variable in the initial processing was reassigned a code of 1 as part of the editing procedures. If the respondent did not choose codeine from the list of drugs in question

PR04A, but reported nonmedical use of another pain reliever from the list, then the code of 98 in the variable for codeine (which had been assigned during the remapping of "enter all that apply" variables, as described in Section 2.2.2) was replaced with a code of 6 during the editing process.

Use of the code of 6 was intended to indicate to analysts that not choosing an available response from the list was not exactly the same as an answer of "no" in questions that required respondents explicitly to answer "yes" or "no" about a behavior of interest. In other words, a response of "no" in a "yes/no" type of question can be thought of as an *active* indication that the behavior or characteristic of interest did not apply, whereas not choosing a response in an "enter all that apply" list can be thought of as a *passive* indication that a particular behavior did not apply. In practice, however, not choosing a response from a list often was treated as being *equivalent* to a response of "no" in subsequent editing steps. For example, see the discussion about editing for the most recent nonmedical use of the pain reliever OxyContin® in Section 6.2.2 in Chapter 6.

However, this procedure was modified for "enter all that apply" questions that were added to the interview beginning in 2013. These new variables were coded as 1 if respondents chose that item from the list, and values of 98 were recoded with a value of 2. Documentation for these values was as follows:

- 1 = Yes, and
- 2 = No (not entered).

For example, questions QD10A, QD10B, and QD10C were added in 2013 for respondents who had ever served on active duty in the United States Armed Forces or Reserve components. Respondents who reported in question QD10A that they had ever served on active duty were asked in question QD10B to indicate the period when they served on active duty, with question QD10B being an "enter all that apply" question. If a respondent reported serving on active duty between May 1975 and July 1990, the code of 3 that was assigned to the variable in the initial processing was reassigned a code of 1 as part of the editing procedures. If the respondent did not report being on active duty during this period but reported being on active duty during another period, then the code of 98 in the variable for this period of active duty military service was replaced with a code of 2 during the editing process. As for the code of 6 that was described previously, the documentation of "No (not entered)" for the code of 2 was intended to indicate to analysts that not choosing an available response from the list was not exactly the same as an answer of "no" in questions that required respondents explicitly to answer "yes" or "no."

2.4.5 "Back-Editing" Based on "OTHER, Specify" Data

"Back-editing" refers to situations in which answers to a given question can be used to make inferences about how a preceding question *within the same module* should have been answered. Specifically, the principles and procedures that are discussed in this section refer to use of "OTHER, Specify" data to go "back" and edit an earlier variable according to what respondents specified in a later series of questions for that module.

In particular, a type of inconsistency that could commonly occur in the NSDUH data involved situations in which respondents did not answer a question affirmatively (e.g., in question LS01A, whether they ever used LSD). In the same module, however, they reported something in "OTHER, Specify" items that indicated that the other question should have been answered as "yes" (e.g., specifying use of LSD as some "other" hallucinogen that the respondent had ever used). When respondents specified something that corresponded to an item they had been asked about previously, but they had not answered that previous item as "yes," then the editing procedures assigned a value of "yes" to the relevant question. The following code typically was used when a response of "yes" was logically inferred: 3 = Yes LOGICALLY ASSIGNED.

If there was a lead to the "OTHER, Specify" question that was in the form of a "yes/no" question (e.g., "Have you ever, even once, used any other hallucinogens besides the ones that have been listed?"), the affirmative answer was retained in the lead to the "OTHER, Specify" question (i.e., having ever used any other hallucinogens), in addition to the inference being made that the answer to another question logically was "yes" (e.g., "Have you ever, even once, used LSD, also called 'acid'?"). The redundant specify code also was retained in the "OTHER, Specify" variable(s) to indicate to analysts the source of the logically inferred "yes" value.

This principle also applied to the editing of variables in "enter all that apply" questions based on answers in "OTHER, Specify" data. The following code typically was used when a response was logically inferred in a variable in an "enter all that apply" question: 3 = Response entered LOGICALLY ASSIGNED.[20] For example, if a respondent did not choose codeine from the list in question PR04A, but specified codeine as another pain reliever that he or she misused, then it was inferred that codeine logically should have been chosen from the list. The individual edited variable corresponding to lifetime nonmedical use of codeine was assigned a code of 3.

2.4.6 "Flag and Impute" Principles for Drug Use Data

The editing procedures for establishing when respondents last used a substance are critical for creating final published estimates from NSDUH of the prevalence of substance use in the United States. In addition, data from core drug use modules on most recent use are important for establishing whether skipped questions in noncore modules truly were not applicable or if there might be some question about whether these skipped questions might have applied to the respondent. For example, respondents who reported that they used cocaine in their lifetime but that they last used it more than 12 months ago were not asked questions about cocaine dependence or abuse. However, if any of these respondents also reported that they first used cocaine at their current age, these reports of recent initiation would suggest that they may have used cocaine in the past 12 months, in which case they should have answered the dependence or abuse questions for cocaine.

[20] No "OTHER, Specify" questions were associated with new "enter all that apply" questions that were added to the interview for 2014.

Under the deterministic 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 or she had never used a substance and indicated use of that substance in another question on the answer sheet, 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 or 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. A drawback of this approach was that decisions to infer more recent use could have an appreciable impact on estimates of use in the past 30 days for less commonly used substances, such as cocaine and heroin.

Conversion of the instrument to a CAI format in 1999 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 the same kinds of edits for the CAI recency variables as were done for PAPI.

Four possible ways of editing the data for most recent use of a drug were examined as part of the methodological research for data processing and estimation procedures using the CAI data (Kroutil & Myers, 2002). The flag and impute rule that was adopted for editing the CAI recency-of-use variables 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 inconsistencies for recency-of-use and related data to be resolved through statistical imputation.

For example, if a respondent originally reported last using a drug more than 12 months ago but also reported first using it at his or her current age, this procedure inferred that the respondent was *at least* a lifetime user. In the imputation procedures, this case's data for most recent use was imputed to indicate most recent use in any period (i.e., in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago). Also, the data on past year initiation that were inconsistent with most recent use of the drug more than 12 months ago were set to bad data (Section 2.4.3), which treated the initiation data as missing. Consequently, if the respondent was imputed to have last used the drug at some point in the past 12 months (including use in the past 30 days), then the respondent could be imputed to have initiated use of the drug within the past 12 months or more than 12 months ago based on data from a donor whose reports of first use and most recent use were consistent. However, if the respondent was imputed to have last used the drug more than 12 months ago, then initiation data from the donor respondent also would be consistent with initiation more than 12 months prior to the interview date.

The beauty of this edit rule lies in its simplicity: If a respondent gives an answer within a substance's module that conflicts with the original answer to the recency-of-use question, then the recency variable is statistically imputed using data from a suitable donor record without these inconsistent data (see Chapter 3). 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, nor does it automatically infer that the respondent last used a substance more recently than he or she originally reported (as was the case with editing procedures for the former PAPI data). However, if final assignment occurred to indicate use in a more recent period than the respondent originally reported, this decision typically was made through statistical imputation rather than deterministic editing.[21]

[21] Limited exceptions that involved deterministic editing of recency variables are discussed in Chapter 6.

3. Imputation and the Predictive Mean Neighborhood Methodology

3.1 Introduction

As with most large-scale sample surveys, the respondent datasets for the 2014 National Survey on Drug Use and Health (NSDUH) contained missing responses for some items, inconsistent or invalid responses, and violations of skip patterns. Although the survey instrument was designed to enforce skip patterns and to perform some consistency checks as data were collected, invalid and inconsistent responses still occur. These response errors are a source of bias in the analysis of NSDUH data (Cox & Cohen, 1985).

Deterministic editing to correct erroneous and inconsistent responses and to replace missing values is appropriate when a unique association exists between predictor variables and the variable to be predicted (Cox & Cohen, 1985). For instance, gender often can be inferred from the respondent's relationship to the head of a household (e.g., son, daughter). However, even when good predictor variables are present, an unambiguous prediction may not be possible for every record having missing or faulty data (e.g., "cousin" does not clarify the gender of a respondent). In such cases, the remaining faulty or missing data often are replaced with statistically imputed data.

"Imputation" is the term used to describe the replacement of missing data with plausible values. Most commonly, imputation is used when a respondent answers some questions on a survey but not others. This is a condition known as "item nonresponse." By contrast, when a selected individual does not respond to any question on the survey at all, or does not respond to enough key questions for the case to be useful for research purposes, this is a condition referred to as "unit nonresponse." In such cases, weighting adjustments are normally employed to account for these missing data. As an initial step, prior to any processing of the data, unit nonrespondents were discarded, and only unit respondents (i.e., item respondents and item nonrespondents for any given questionnaire item) were included in the subsequent editing, imputation, and analysis of NSDUH data. Throughout the remainder of this report, the terms "nonresponse" and "nonrespondent" are used to denote item nonresponse and item nonrespondent, respectively.

Once processed, imputed values cannot be distinguished from nonmissing values for a given variable in the final dataset. Therefore, observations with imputed data must be identified with a concomitant indicator variable. The vast majority of imputation-revised variables for the 2014 NSDUH have the prefix "IR" attached to their names.[22] Although no missing data were possible for gender because a response to this item was required before the interview could proceed, the "IR" prefix for IRSEX was maintained for continuity with past years. Each imputed variable has an associated indicator variable, identified by the prefix "II" that can be used to identify which values were imputed and which were not. For some imputation-revised variables,

[22] Exceptions to this rule included the imputation-revised employment status variables EMPSTAT4 and EMPSTATY and the core-plus-noncore methamphetamine and stimulant variables CPNMTHFG, CPNMTHYR, CPNMTHMN, CPNSTMFG, CPNSTMYR, and CPNSTMMN.

additional imputation indicators were created with the prefix "II2." These indicators gave more details about the source of the imputed or logically assigned value. The levels of a typical "II" imputation indicator are as follows:

- 1 = From questionnaire
- 2 = Logically assigned
- 3 = Statistically imputed
- 9 = Legitimate skip

Assignment of a code of 9 to an imputation indicator reflected the prior assignment of a legitimate skip code as part of the editing process (Section 2.4.2). Also, if a question had missing or ambiguous data and this question governed the skip logic for subsequent questions, respondents typically were skipped out of the subsequent questions. Respondents requiring imputation for the variables that governed a skip pattern typically receive a value of 3 (statistically imputed) for the imputation indicators associated with both the governing variables and the imputed variables that were nested within the skip pattern. For example, if a respondent had missing data for whether he or she had ever used a particular substance and was imputed to have never used it, the imputation indicators for recency of use of the substance, age at first use, frequency of use in the past 12 months (if applicable), and frequency of use in the past 30 days (if applicable) all were assigned a code of 3.

3.2 Development of the Predictive Mean Neighborhood Methodology

Various methods of imputation have been used since the NSDUH was first administered in the early 1970s.[23] With the expansion of the NSDUH sample size in 1999, the predictive mean neighborhood (PMN) method for imputation was implemented and is currently used for most variables. PMN is designed to incorporate the complex interrelationships among items in the current NSDUH, thus maintaining data consistency within individual respondent records. Table 3.1 provides a summary of the types of imputation procedures used for each of the variables imputed in the NSDUH samples from 1999 through 2014.

Table 3.1 Summary of Item Imputation Procedure Used, by Variable and NSDUH Year

Variable	1999[1]	2000	2001	2002-2003	2004-2014
Interview Date	Random[2]	Random	None	None	None
Age	None[3]	None	None	None	None
Birth Date	None	Random	Random	Random	Random
Gender	None	None	None	None	None
Race	USHD[4]	PMN	PMN	PMN	PMN
Hispanic or Latino Origin Indicator	USHD	PMN	PMN	PMN	PMN
Marital Status	USHD	PMN	PMN	PMN	PMN
Hispanic or Latino Origin Group	USHD	PMN	PMN	PMN	PMN
Education	USHD	USHD	PMN	PMN	PMN

[23] Prior to the 2002 survey year, when it was renamed, the National Survey on Drug Use and Health (NSDUH) was originally known as the National Household Survey on Drug Abuse (NHSDA).

Table 3.1 Summary of Item Imputation Procedure Used, by Variable and NSDUH Year (continued)

Variable	1999[1]	2000	2001	2002-2003	2004-2014
Employment Status	USHD	USHD	PMN	PMN	PMN
Immigrant	Not imputed	Not imputed	Not imputed	WSHD[5]	PMN
Health Insurance	PMN	PMN	PMN	PMN[6]	PMN
Lifetime Drug Usage	PMN	PMN	PMN	PMN	PMN
Recency and Frequency of Use[7]	PMN	PMN	PMN	PMN	PMN
Age at First Use	PMN	PMN	PMN	PMN	PMN
Age at First Daily Cigarette Use	PMN	PMN	PMN	PMN	PMN
Personal and Family Income (Binary)	PMN	PMN	PMN	PMN	PMN
Personal and Family Income (Finer Categories)	PMN	PMN	PMN	PMN	PMN
Nicotine Dependence	Not imputed	Not imputed	Regression	Regression	Regression
Household Size (Roster-Derived)	PMN	PMN	PMN	PMN	PMN
Other Household Composition (Roster-Derived)	PMN	PMN	PMN	PMN	PMN

[1] The 1999 survey year also included a paper-and-pencil interviewing sample. The procedures listed here are from the computer-assisted interviewing sample.
[2] "Random" refers to a random assignment within a quarter for the interview date and a random assignment using age and interview date for the birth date.
[3] "None" means that no missing values were encountered after editing, and thus no imputation was necessary. For gender (from the 2002 survey onward) and age, missing values were precluded by design (see Chapter 4).
[4] "USHD" refers to the unweighted sequential hot-deck method of item imputation described in this report (see Section 3.2.1.1).
[5] "WSHD" refers to the weighted sequential hot-deck method of item imputation described in this report (see Section 3.2.1.2).
[6] Although PMN was the method used for health insurance in all years since the 1999 survey, imputation also was applied to more detailed health insurance variables in the surveys from 2002 onward.
[7] "Recency and Frequency of Use" included variables measuring recency of use, 12-month frequency of use, 30-day frequency of use, and binge drinking frequency in past 30 days. "Binge drinking" was defined as having five or more drinks on the same occasion on a given day.

3.2.1 Previously Used Hot-Deck Imputation Methods

With any method of imputation, missing responses for a particular variable (hereafter, termed "base" variable) are replaced by values from similar respondents with respect to a number of characteristics (hereafter, "auxiliary variables"). If "similarity" is defined in terms of a single predicted value from a model, these auxiliary variables can be represented by that value. The respondent with the missing value for the base variable is called the "recipient," and the respondent from whom values are borrowed to replace the recipient's missing value is called the "donor." Donors and recipients are distinguished by the completeness of their records with regard to the variable(s) of interest (i.e., the donor has complete data for that variable, and the recipient does not). The term "hot deck" is used to refer to imputations made on recipient base variables using donor values from the same dataset. The PMN methodology utilized on the NSDUH is a specialized hot-deck method and is described in greater detail later in this chapter. For more information on the general hot-deck method of item imputation, see Little and Rubin (1987, pp. 62-67).

For the 2014 NSDUH, the only imputations that did not incorporate the PMN method were those used for the birth date, date of first use, and nicotine dependence variables, described in Section 4.2.5, Section 6.3.3.4, and Chapter 7, respectively. Two other hot-deck methods—unweighted sequential hot deck (USHD) and weighted sequential hot deck (WSHD) (Cox, 1980, pp. 721-725; Iannacchione, 1982)—were used in past surveys.[24] In the sections that follow, the features and limitations of USHD, WSHD, and the random nearest neighbor hot deck (NNHD) are discussed as background for the development of the PMN methodology.

3.2.1.1 Unweighted Sequential Hot Deck

In a sequential hot-deck procedure, data are first ordered using specific criteria, and the last reported value in the sequence is substituted for each missing value as the data are processed. In USHD, the selection of a response for imputation purposes is independent of the sampling weight associated with the data record from which the response is taken and the data record to which a response is being imputed. USHD imputation is, therefore, based upon the tacit assumption that nonrespondents would answer in a manner similar to that of respondents immediately adjacent to them in an appropriately sorted data file and hence that the data associated with the nearest neighbor are appropriate for the imputation of missing values (Cox, 1980, p.721).

Implementation of the USHD method (and of hot-deck methods, in general) involves three basic steps:

1. **Construct imputation classes**. When there is a strong logical association between the base variable and certain auxiliary variables, the dataset is partitioned by these auxiliary variables, and imputation procedures are implemented independently within the resulting imputation classes defined by the cross of these auxiliary variables.
2. **Sort the analytic file.** Within each imputation class, the file is sorted by auxiliary variables relevant to the item being imputed. The sort order of the auxiliary variables is chosen to reflect the degree of importance of the auxiliary variables in their relation to the base variable being imputed (i.e., those auxiliary variables that are better predictors for the item being imputed are used as the first sorting variables). In general, two types of sorting procedures—a straight sort and a serpentine sort[25]—were used in previous surveys to sort the files prior to imputation.
3. **Replace missing values with imputed values.** The sorted file is read sequentially. Each time an item respondent is encountered (i.e., the base variable is nonmissing),

[24] The USHD method was used exclusively for the 1991-1998 surveys, for the paper-and-pencil interviewing sample from the 1999 survey, and for all demographic variables in the computer-assisted interviewing sample from the 1999 survey. In the 2002-2003 surveys, missing values in the immigrant variables required WSHD imputation. Note, however, that the USHD and WSHD methods have not been used on the NSDUH since the 2000 and 2003 survey years, respectively.

[25] Under a straight sort, a set of variables is sorted in ascending order by the first variable specified. Then, within each level of the first variable, the file is sorted in ascending order by the second variable specified, and so forth. In a serpentine sort, a set of variables is sorted so that the direction of the sort (ascending or descending) for subsequent variables changes each time the value of the preceding variable changes. The serpentine sort has the advantage of minimizing the change in the entire set of auxiliary variables every time any one of the variables changes its value. For an example of each, see Appendix A of the 2009 imputation report (Ault et al., 2011).

the base variable response is stored, updating the donor response. Any subsequent nonrespondent in the file receives the stored donor response, which in turn results in a statistically imputed response. Because the file is sorted by relevant auxiliary variables, the preceding item respondent (donor) closely matches the neighboring item nonrespondent (recipient) with respect to the auxiliary variables.

For any particular item being imputed under USHD, there is the risk of several nonrespondents appearing next to one another on the sorted file; in this situation, each would receive imputed values from the same donor. To detect this problem on the NSDUH in the survey years prior to 2001, the imputation donor was identified for every item being imputed, and frequencies by donor were examined. If several nonrespondents were aligned next to one another after sorting, sort variables were added or eliminated, or the ordering of the sort variables was modified, to ensure that multiple nonrespondents did not comprise adjacent records on the resulting file.

3.2.1.2 Weighted Sequential Hot Deck

WSHD improves upon USHD by incorporating the sampling weights when replacing missing values among recipient records in the final hot-deck assignment step. The earlier steps taken to impute for missing values under the WSHD method are the same as those for the USHD method; as in USHD, WSHD requires the formation of imputation classes and appropriate sorting (straight or serpentine) of the analytical file.

The WSHD procedure used in surveys prior to 2004 followed directly from Cox (1980). Specifically, once the imputation classes were formed, the data were divided into two datasets: one for respondents and one for nonrespondents. Scaled weights v_j were then derived for all nonrespondents using the following formula:

$$v_j = w_j s_+ / w_+ ; \; j = 1, 2, \ldots, n,$$

where n is the number of nonrespondents, w_j is the sample weight for the j^{th} nonrespondent, w_+ is the sum of the sample weights for all the nonrespondents, and s_+ is the sum of the sample weights for all the respondents. The respondent data file was partitioned into zones of width v_j, where the imputed value for the j^{th} nonrespondent was selected from a respondent in the corresponding zone of the respondent data file. This selection algorithm is an adaptation of Chromy's (1979) sequential sample selection method.

WSHD controls the number of times a donor can be selected and allows each respondent the chance to be a donor because a respondent is selected within each v_j. Consequently, the most important benefit of the weighted sequential hot-deck method is the elimination of bias in the estimates of means and totals, particularly when the response rate is low or when the covariates explain only a small amount of variation in the specified variable. In addition, many surveys sample subpopulations at different rates, and using the sample weights allows the imputed data for the nonrespondents to have the same mean (for the specified variables) as the respondents. In other words, the weighted hot deck preserves the respondent's weighted distribution in the imputed data (Cox, 1980).

3.2.1.3 Unweighted Random Nearest Neighbor Hot Deck

Another commonly used imputation method—one not directly used on the NSDUH, but related to the PMN method—is random nearest neighbor hot deck (NNHD) (Little & Rubin, 1987, p. 65). With this method, a donor set or neighborhood deemed "close to" the recipient, with respect to a number of covariates, is used to select a donor at random. The distance between the values of the recipient and potential donors for each of the auxiliary variables is calculated, and then the donors for the neighborhood are chosen such that the maximum of these distances is less than a certain threshold value, referred to as "delta." This neighborhood is restricted, using imputation classes described previously, so that the potential donors' values of the base variable are consistent with the recipient's preexisting nonmissing values of related variables.

Because a distance function is used to define "closeness" between the recipient and a donor under NNHD, there is less of a problem of sparseness of the donor class when imputing for continuous variables. It should be noted, however, that the distance function involving categorical or nominal variables is typically ad hoc and often hard to justify.

3.2.2 Advantages of the Predictive Mean Neighborhood Methodology

The PMN methodology developed for and implemented on the 1999 NSDUH was an attempt to address the shortcomings, while retaining the positive characteristics, of the hot-deck imputation methods discussed above. It is a combination of two commonly used imputation methods: non-model-based NNHD (Little & Rubin, 1987, p. 65) and a modification of Rubin's model-assisted predictive mean matching (PMM) method (Rubin, 1986), which matches a missing value to the observed value with the closest predicted mean. The PMN method enhances Rubin's PMM method, in that PMN can be applied to both discrete and continuous variables, either individually or jointly. PMN also enhances the NNHD method for discrete variables so that the distance function used to find neighbors is no longer in terms of the original predictor variables and therefore does not require arbitrary scaling.

In addition, the PMN methodology offers the following advantages over the imputation methods employed on earlier NSDUHs:

- **A greater number of auxiliary variables may be used to determine donors.** Using a model-based hot-deck technique like PMN allows auxiliary variables to be incorporated in two ways: first, as covariates in models, and second, in likeness constraints[26] applied to potential donors. Under USHD and WSHD, the number of auxiliary variables is limited in part by the problem of sparse neighborhoods; donors must match recipients for all variables used to form imputation classes. If too many variables are used to form imputation classes, some classes may be very small and contain few or no item respondents to serve as donors. By contrast, under PMN, the donors need only be "close" to the recipients with respect to the predicted values determined by the models, even when the models include numerous covariates. Moreover, PMN ensures that a sufficient number of potential donors comprise the donor neighborhood, so that likeness constraints may be applied on the donor set as needed.

[26] Likeness constraints are flexible constraints that govern the similarity between donors and recipients. See Section 3.3.1.3 for details.

- **Relative importance of auxiliary variables is determined by standard estimating equation techniques.** Under USHD and WSHD, as implemented, the selection of classing and sorting variables was sometimes ad hoc, and in the former instance, weights were not utilized. In PMN, by contrast, objective criteria based on a more rigorous methodology (i.e., regression) quantify the relationship between a given covariate and the response variable in the presence of other covariates, so that the response variable itself is indirectly used to determine donors. Further, the sampling weights can be incorporated in PMN regression models without difficulty.
- **Internal consistency of the post-imputation record is guaranteed.** In PMN, the donor pool can be restricted to those making the post-imputation record logically consistent. For example, if a recipient must receive a cocaine past year frequency of use between 30 and 50, the donors can be restricted to ensure that the recipient receives such a value. In USHD and WSHD, the classing and sorting variables cannot easily be used to guarantee this; there may not be a donor in the imputation class who will create a consistent record.
- **Correlations across response variables are accounted for by making the imputation multivariate.** In comparison with other model-based methods, discrete and continuous variables can be handled jointly and relatively easily in PMN by using the idea of sequential univariate modeling. Further, differential weighting factors can be objectively assigned to different elements of the predictive mean vector depending on the variability of predicted means in the dataset.

3.3 Implementation for the Predictive Mean Neighborhood Methodology

The implementation of PMN on the NSDUH involves three basic steps: response propensity (RP) adjustment, prediction (PRD) modeling, and hot-deck imputation. At the most basic level, the RP adjustment reallocates the weights of the item nonrespondents to item respondents; the prediction model calculates predicted means for both; and the hot-deck step assigns final values to the item nonrespondents based on a distance function derived from these predicted means. These steps are described in more detail in the following sections and are combined in three different ways, called PMN "types" (Section 3.4), to complete imputation procedures.

3.3.1 Step 1: Response Propensity Adjustment

Response propensity is defined as the probability of response, whether at the unit level or item level. The purpose of response propensity is to adjust the sampling weights for item nonresponse so that the item respondent weights that are used only during the imputation process are representative of the entire domain of interest. In the response propensity step of PMN, the item response propensity is modeled as a function of a predetermined set of covariates. The model can be thought of as a special case of the generalized exponential model (GEM)[27]

[27] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures and is described in detail in Appendix A of the person-level sampling weight calibration report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2016a).

developed for weighting procedures, in that imputations that are done at the item level are similar in nature to the weight adjustments made for entire units.

There are four key inputs to the item response propensity modeling step:

1. **Analysis weights.** For all imputation procedures, the best available analysis weight is used as an input to the NSDUH imputation procedures. Because of the timing of 12-month processing and, in particular, the coordination between the weighting and imputation tasks in each NSDUH year, most variables that undergo imputation utilize the preliminary analysis weight, PANALWT. PANALWT was developed by the NSDUH weighting team for imputation purposes. The dwelling unit-level design weights were adjusted for nonresponse using a weighting class method by state. The nonresponse-adjusted dwelling unit weights and the person-level design weights were further adjusted for nonresponse at the person level using a weighting class method by state and age groups (12-17, 18-25, 26-34, 35-49, and 50+). PANALWT is the product of all design weights and the two nonresponse adjustment factors. For those variables that are processed later in the annual cycle, the final analysis weight ANALWT may be used instead, if it is available at the time of imputation processing for that variable. The pair variables described in Chapter 11 utilize yet another weight, PRANALWT. See the person-level sampling weight calibration report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2016a) for full descriptions of the final analysis weights. See the questionnaire dwelling unit-level and person pair-level sampling weight calibration report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2016b) for full descriptions of the pair weights.
2. **Domain indicator.** In this report, a "domain" is defined as the set of respondents who are included in models and for whom predicted means are calculated. For many NSDUH variables that undergo imputation, the domain includes all unit respondents. For others, the domain is a subset of unit respondents. For example, the marital status question is only asked of respondents aged 15 or older. In this case, the domain indicator is set to zero for respondents aged 12 to 14 and one for respondents aged 15 or older. The domain indicator is an important input to the tables in Appendix A, where item response rates are reported for each variable that undergoes imputation on the NSDUH.
3. **Item response indicator.** The item response indicator is set to zero for item nonrespondents and one for item respondents. GEM uses this indicator to identify the item nonrespondents and item respondents when reallocating the weights appropriately. The item response indicator is an important input to the tables in Appendix A, where item response rates are reported for each variable that undergoes imputation on the NSDUH.
4. **Covariates.** GEM uses a predetermined list of covariates to allocate the weights from item nonrespondents to item respondents. The covariates tend to be variables that are correlated with (1) the propensity to respond, (2) the variable that is undergoing imputation, or (3) both. The goal is to avoid bias in the prediction models by allocating the weights of the item nonrespondents to similar item respondents, without too greatly inflating the variance of the estimates that utilize these weights (Center for Behavioral Health Statistics and Quality, 2016a). Appendix D lists the starting and final covariate lists for each response propensity model fit in NSDUH imputation procedures.

3.3.2 Step 2: Prediction Modeling

Utilizing the response propensity-adjusted weights that were derived in the previous step, the prediction model calculates predicted means, which are used in the hot-deck step(s) to create neighborhoods and select donors. The dependent variable in the model is usually the variable, or some transformation of that variable, that is undergoing imputation. Each model is built using only those cases within the domain with complete responses for that item. Predicted means are then calculated for all of the domain members, whether or not they were item respondents, using the values for the covariates and the estimates for the regression coefficients.

For categorical outcome variables, logistic regression models are used for the prediction models. For continuous variables, linear regression models are fit. For count variables, Poisson regression models are used. For response variables that are proportions (e.g., months on welfare; see Chapter 9), a logit transformation is applied to the proportion, and a linear regression model is utilized. The variable sets in which some transformations of the response variables were implemented include the noncore demographics (Chapter 5), drugs (Chapter 6), and income (Chapter 9).

The goal of the prediction model is to protect against bias for as many analysis domains as possible, so these models tend to start with long lists of covariates (Allison, 2008).[28] Appendix D lists the starting and final covariate lists for each prediction model fit in NSDUH imputation procedures. In contrast to explanatory (association) models where model parsimony is a relevant metric of a model's appropriateness, the focus in a prediction setting is on the predicted values only.

The SUDAAN software package is used to fit nearly all the prediction models used in the NSDUH.[29] All covariates from the applicable starter list are utilized unless SUDAAN produces warning messages, which indicate nonconvergence or model instability. Generally, in these cases, covariates whose coefficients have high p-values are dropped until SUDAAN no longer produces these warning messages. The primary advantage of using SUDAAN to fit prediction models is that the standard errors associated with the regression coefficients properly account for the complex survey design. The predicted means are the same using SUDAAN as they are using, for example, the analogous base SAS procedure (given the same set of covariates), but the decision on which covariates to drop in the event of model instability or nonconvergence is more informed under SUDAAN because the standard errors, and the corresponding test statistics and p-values for model coefficients, account for stratification and clustering.

In the particular case of some of the logistic regression models, the warning messages produced by SUDAAN may be triggered when a cross-classification of the outcome variable and a covariate has empty or nearly empty cells. Covariates of this type are highly correlated with the outcome variable but cannot be used in the prediction model. However, they are often used in the hot-deck step to identify suitable donors.

---

[28] Allison (2008) writes that "… imputation models should be relatively 'rich' so that they may be congenial with lots of different models that could be of interest."

[29] The only exception is the finer income categories, described in detail in Chapter 9.

For the types of regression-based prediction models used for each variable that underwent imputation using PMN, see Table 3.2.

Table 3.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood

Variable	Domain	Type of Regression Model	SAS/SUDAAN Procedure[1,2]
Demographics			
Marital Status	15 years or older	Multinomial Logistic	MULTILOG
Race	All respondents	Multinomial Logistic	MULTILOG
Hispanic or Latino Indicator	All respondents	Binomial Logistic	RLOGIST
Hispanic or Latino Group	Hispanics	Multinomial Logistic	MULTILOG
Education Level	All respondents	Multinomial Logistic	MULTILOG
Employment Status	15 years or older	Multinomial Logistic	MULTILOG
Immigrant Status: Born-in-U.S. Indicator	All respondents	Binomial Logistic	RLOGIST
Immigrant Status: Age of Entry	Not born in U.S.	Simple Linear	REGRESS
Drugs			
Lifetime Drug Use	All respondents	Binomial Logistic	RLOGIST
Recency of Drug Use, "Hierarchical" Drugs	All lifetime users for past year vs. not past year; all past year users for past month vs. not past month	Binomial Logistic	RLOGIST
Recency of Drug Use, Pipes	All lifetime users	Binomial Logistic	RLOGIST
Recency of Drug Use, All Other Drugs	All lifetime users	Multinomial Logistic	MULTILOG
12-Month Frequency of Drug Use	All past year users	Simple Linear	REGRESS
Daily Drug Use over Past 30 Days, Cigarettes, Chewing Tobacco, and Snuff	All past month users	Binomial Logistic	RLOGIST
30-Day Frequency of Drug Use, Cigarettes, Chewing Tobacco, and Snuff	All past month users except those who used daily over the past 30 days	Simple Linear	REGRESS
30-Day Frequency of Drug Use, All Other Drugs	All past month users	Simple Linear	REGRESS
Age at First Drug Use	All lifetime users	Simple Linear	REGRESS
Household Composition			
Total Number of Rostered People	All respondents	Poisson	LOGLINK
Total Number of Children Younger than 18	All respondents	Poisson	LOGLINK
Total Number of People Aged 65 or Older	All respondents	Poisson	LOGLINK
Indicator of Whether the Respondent Has Family Members in Household	All respondents	Binomial Logistic	RLOGIST
Total Number of Respondent's Family Members in the Household (Excludes Foster Relationships)	All respondents	Poisson	LOGLINK
Total Number of Respondent's Family Members in the Household Younger than 18 (Excludes Foster Relationships)	All respondents	Poisson	LOGLINK

Table 3.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood (continued)

Variable	Domain	Type of Regression Model	SAS/SUDAAN Procedure[1,2]
Total Number of Respondent's Family Members in the Household (Includes Foster Relationships)	All respondents	Poisson	LOGLINK
Total Number of Respondent's Family Members in the Household Younger than 18 (Includes Foster Relationships)	All respondents	Poisson	LOGLINK
Income			
Source of Income	All respondents	Binomial Logistic	RLOGIST
Months on Welfare	All respondents who received welfare payments or welfare services in the past year	Simple Linear	REGRESS
Total Income (Binary)	All respondents	Binomial Logistic	RLOGIST
Finer Category Income	All respondents	Time-to-Event (Survival)	LIFEREG
Health Insurance			
Specific Types of Health Insurance Coverage	All respondents	Binomial Logistic	RLOGIST
Any Other Health Insurance	All respondents with none of the specific types of health insurance	Binomial Logistic	RLOGIST

[1] SAS® software is a registered trademark of SAS Institute Inc. SUDAAN® is a registered trademark of Research Triangle Institute.

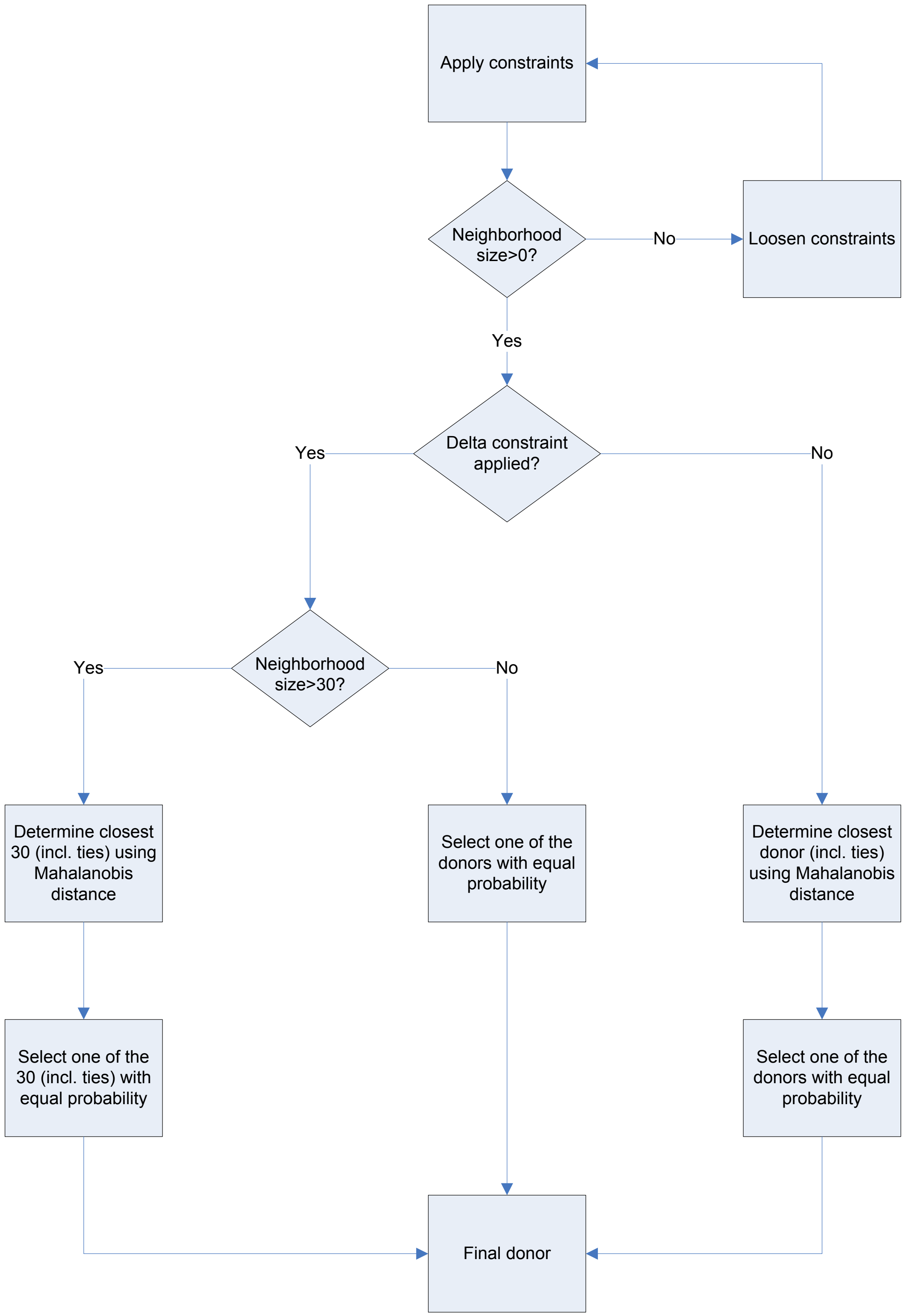
[2] See RTI International (2013) for more information on all SAS-callable SUDAAN procedures in this table except LIFEREG and LOGISTIC. See SAS Institute Inc. (2014) for more information on the LIFEREG and LOGISTIC procedures.

3.3.3 Step 3: Hot-Deck Imputation

After sampling weights have been appropriately adjusted in the response propensity step and predicted means have been calculated in the prediction step, the hot-deck step[30] of PMN is applied to select a donor for each item nonrespondent. The algorithm used to select donors is graphically displayed in the flowchart in Figure 3.1. Briefly, likeness constraints are loosened in an iterative fashion until PMN yields a nonempty donor neighborhood. Mahalanobis distance is then used to rank donors by closeness to the item nonrespondent, and a final donor is selected at random for a given recipient. Many of the hot-deck components used in PMN are described below and appear explicitly in the tables of Appendix E.

[30] There is one situation on the NSDUH in which the imputation is not a hot-deck step, but is a stochastic imputation based solely on the predicted mean(s) of the recipient, of the type described in Section 5.1 of the evaluation of imputation methods for the NSDUH report (Center for Behavioral Health Statistics and Quality, in press). These ideas have their origin in Singh, Grau, and Folsom (2004), where Centered PMN is discussed as an alternative to PMN. This one exception is the provisional hot-deck step for imputation set 2 for the health insurance variables (Section 10.3.2.3).

Figure 3.1 Donor Selection Algorithm



3.3.3.1 Logical and Likeness Constraints

Logical constraints and likeness constraints are restrictions placed on the set of donors to make imputed values consistent with preexisting, nonmissing values of the item nonrespondents (recipients) and to make candidate donors as much like the recipients as possible. Logical constraints are fixed constraints that prevent logical inconsistencies between variables, and likeness constraints are flexible constraints that govern the similarity between donors and recipients.

The logical constraints are never removed, because to do so would risk the selection of a donor that produces an inconsistent post-imputation record. For example, for the employment status variable, if the item nonrespondent is known to be employed, but full-time vs. part-time status is unknown, the imputed value must come from a donor who is employed as well.

Likeness constraints are placed on the pool of donors to make the attributes of the neighborhood as close as possible to those of the recipient. For example, age and employment status are correlated. A likeness constraint exploits this correlation by requiring the donor's age to be within 5 years of the item nonrespondent's age, but likeness constraints may be loosened if they happen to force the donor pool to be empty.

One likeness constraint that is used in all hot-deck steps, regardless of the variable being imputed, is the delta constraint. This particular likeness constraint requires the donor's predicted mean to be within 5 percent (delta) of the item nonrespondent's predicted mean for each element of the predictive mean vector. If the predicted means are probabilities, the values of delta vary depending upon the value of the predicted mean.

Each delta is defined as 5 percent of the predicted probability if the probability were less than 0.5 and is defined as 5 percent of 1 minus the predicted probability if the probability were greater than 0.5. This allows for a looser delta for predicted probabilities close to 0.5 and a tighter delta for predicted probabilities close to 0 or 1. The range of values for delta across various predicted probabilities is shown in Table 3.3.

Table 3.3 Values of Delta for Various Predicted Probabilities

Predicted Probability (p)	Delta
$p \leq 0.5$	$0.05p$
$p > 0.5$	$0.05(1 - p)$

NOTE: In very rare cases, due to transformations of the response variable for the prediction model, p was slightly less than 0 or slightly greater than 1. This occurred only for 30-day frequency and 12-month frequency. If p was less than 0, delta was set to $0.05|p|$. If p was greater than 1, delta was set to $0.05(p - 1)$. See Section 6.3.2.6 for a detailed description of these transformations and their associated back-transformations.

Logical constraints and likeness constraints, including the order in which likeness constraints are loosened for some variables, are presented in the tables in Appendix E.

3.3.3.2 Predictive Mean Vector

The predicted means from the prediction step play a central role in the donor selection algorithm depicted in Figure 3.1, through the construct of the predictive mean vector. The predictive mean vector is essentially a list of predicted means from the prediction modeling step. In simple cases, the predictive mean vector contains only one element, such as the predicted age at which a respondent began using a drug. In complex cases, the predictive mean vector includes several elements from several different prediction models, such as the predicted recency and predicted frequency of use for a given drug.

When the prediction model is a logistic regression model, predicted means are calculated for each level of the outcome variable. For example, the employment status variable that undergoes imputation has four levels: employed full time, employed part time, unemployed, or other. Therefore, a single prediction model is fit using a four-level outcome variable, yielding predicted probabilities for each level, as follows:

- E1: P(respondent is employed full time)
- E2: P(respondent is employed part time)
- E3: P(respondent is unemployed)
- E4: P(other)

Note, however, that the predictive mean vector for the employment status variable contains only three elements. It does not include the predicted probability for the reference cell, which in this case is the "other" level, since that level is implicitly defined by the presence of the other three predicted means.

Occasionally, the predicted means are adjusted so that they are made conditional on what is known for a given respondent. Continuing the example above, some respondents report that they have a job but are unclear about the number of hours they usually work in a week. Because the NSDUH definition of part-time versus full-time employment status was based on working a minimum of 35 hours in a usual week, the predictive mean vector is made conditional on employment of any sort for these respondents. Therefore, the single predicted mean used for these respondents is equal to E1 ÷ (E1 + E2), P(respondent is employed full time | respondent is employed). Conditional probabilities are also used in the binary income hot-deck step and the drug recency/frequency hot-deck steps.

Predictive mean vectors are presented in the tables in Appendix E.

3.3.3.3 Univariate vs. Multivariate Matching and Assignment

If the predictive mean vector consists of only one element, *univariate* matching is used to select a donor. If the predictive mean vector consists of more than one element, *multivariate* matching is used to select a donor. The donor may also give values to the item nonrespondent for more than one variable, a situation known as multivariate assignment. Similarly, if the donor provides values for only one variable, the hot-deck step uses univariate assignment. Table 3.4

shows examples of NSDUH variables that were imputed using each of the four combinations of univariate/multivariate matching and assignment.

Table 3.4 Examples of Variables Imputed Using Each of the Four Combinations of Univariate/Multivariate Matching and Assignment

	Variables Imputed One at a Time (Univariate Assignment)	**Variables Imputed in a Set (Multivariate Assignment)**
Predictive Mean Vector Has One Element (Univariate Matching)	Hispanic/Latino Origin (Section 4.3.3)	Finer Category Income (Section 9.3.2)
Predictive Mean Vector Has More Than One Element (Multivariate Matching)	Marital Status (Section 4.3.1)	Lifetime Drug Use (Section 6.3.1)

Whether the hot-deck step employs univariate or multivariate matching, Mahalanobis distance is used to rank the donors by closeness to the item nonrespondent. The Mahalanobis distance is used instead of Euclidean distance in order to standardize the distance in terms of the population variances and covariances of vector components.[31] It is given by

$$\sqrt{(\boldsymbol{\mu}_R - \boldsymbol{\mu}_{NR})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_R - \boldsymbol{\mu}_{NR})}\,,$$

where $\boldsymbol{\mu}_R$ refers to the predictive mean vector for a given item respondent, and $\boldsymbol{\mu}_{NR}$ is the predictive mean vector for a given item nonrespondent. The matrix $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the predictive mean vector, using the set of item respondents that comprise that domain. Because the square of the Mahalanobis distance is a monotone function of the distance itself, and only the ranking of the donors (instead of the absolute distance measure) is used in the algorithm, the additional step of taking the square root of the squared distance is not performed in practice.

3.3.3.4 Missingness Patterns

For many variables imputed on the NSDUH, item nonrespondents were segregated into patterns of nonresponse called missingness patterns. Missingness patterns arise in two ways. First, for sets of variables that underwent multivariate assignment, item nonrespondents were segregated into missingness patterns based on which variables were missing. Second, a new missingness pattern could emerge when logical editing restricted an item nonrespondent to only a subset of the variable's possible values. The example for employment status discussed above applies here as well: respondents whose employment status was completely unknown had a different missingness pattern than did those who were known to be employed. Often, different predictive mean vectors were used, and different constraints were applied, for different missingness patterns. Many of the tables in Appendix E are segregated by missingness pattern for this reason.

[31] When univariate matching is employed, the ranking of donors by Mahalanobis distance is equivalent to the ranking of donors by Euclidean distance. This is not necessarily true when multivariate matching is employed.

3.3.3.5 Final Assignment of Donor Values

Logical and likeness constraints are used to form a neighborhood of potential donors from the pool of item respondents within each missingness pattern. Logical constraints are always imposed to maintain internal consistency, whereas likeness constraints are removed or relaxed in a predetermined order until this donor neighborhood is nonempty. Once a nonempty neighborhood is found, the rest of the PMN donor selection algorithm depends on whether or not the delta constraint was applied.

If the delta constraint was applied, all the members of the neighborhood are similar to the recipient with respect to the predictive mean vector. The final donor is then randomly selected with equal probability from among the "closest" (in terms of Mahalanobis distance) 30 members of the neighborhood; potential donors whose Mahalanobis distance from the recipient are equal ("ties") are accounted for in the donor selection algorithm depicted in Figure 3.1. If, on the other hand, the delta constraint was not applied, to ensure that the final donor is as close to the item nonrespondent as possible with respect to the predicted means, the donor with the smallest Mahalanobis distance is selected as the final donor. If there is more than one "closest" donor (i.e., there are ties), the final donor is randomly selected with equal probability from among the closest donors. At the conclusion of the hot-deck step, the item nonrespondent receives values from the selected donor for a single variable (in the univariate assignment case) or for a set of variables (multivariate assignment).

3.4 Predictive Mean Neighborhood "Types"

There are three types of PMN as applied on the NSDUH: Type 1, *single response propensity (RP)/single prediction (PRD)* (Section 3.4.1); Type 2, *multiple RP/multiple PRD* (Section 3.4.2); and Type 3, *single RP/multiple PRD* (Section 3.4.3). Each of the three PMN types is a coordinated application of the three basic steps of PMN discussed in Section 3.3.

In PMN, an imputation "set" is a set of variables for which a single donor is used in the final hot-deck step.[32] Sets are formed based on the extent of correlation among variables and the level of missingness in the data. Variables with few missing values and no strong relationships with other variables tend to be processed in an imputation set by themselves. Closely related variables tend to be processed together in the same set to preserve, as much as possible, correlations between variables in the data. However, the more variables that are included in a multivariate set, the less likely it is that a nonempty neighborhood can be found using the delta constraint. Even though there are many advantages to using a multivariate imputation set, one disadvantage in several instances is not being able to apply the delta constraint.

Table 3.5 lists the imputation sets for each variable group discussed in this report and the PMN type used to process each set.

[32] Section 3.4.2 defines and discusses the differences between *provisional* and *final* hot-deck steps in the context of PMN.

Table 3.5 PMN Types Applied to Each Variable Group and Imputation Set

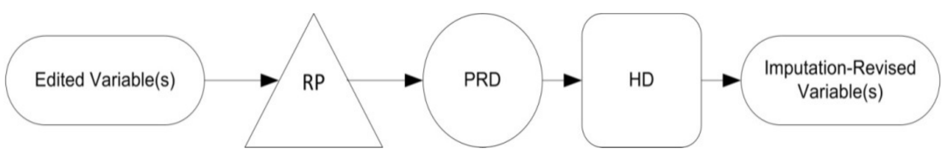
Variable Group	Imputation Set	PMN Type
Core Demographics	All (5 sets)	1 (Single RP/Single PRD)
Noncore Demographics	All (3)	1
Drugs	Lifetime	3 (Single RP/Multiple PRD)
	Recency of Pipe Use	1
	Recency/Frequency, Other Drugs (13)	2 (Multiple RP/Multiple PRD)
	Cigarette Ever Daily Used	1
	Age at First Use (14)	1
Roster	All (8)	1
Income	Binary	3
	Finer Categories	1
Health Insurance	Types of Health Insurance	3
	Any Other Health Insurance	1
Roster Pair	Pair Relationship	1
	Multiplicities (6)	1
	Household Counts, Sibling-Sibling and Spouse-Spouse (4)	1
	Household Counts, Parent-Child	2

PRD = prediction; PMN = predictive mean neighborhood; RP = response propensity.

3.4.1 Type 1: Single Response Propensity/Single Prediction

PMN Type 1, the single RP/single PRD type, involves a single iteration of the three basic steps described in Section 3.3: response propensity, prediction, and hot-deck imputation. Many variables that undergo imputation in the standard processing cycle use this type, including all the demographics and roster variables and the age-at-first-use drug variables. Figure 3.2 illustrates the single RP/single PRD type of PMN imputation.

Figure 3.2 PMN Type 1: Single Response Propensity/Single Prediction



Usually the single RP/single PRD type involves univariate assignment in the hot-deck step,[33] but it may involve univariate or multivariate matching, depending on the prediction model. If the prediction model is a dichotomous logistic regression, linear regression, or Poisson regression model, univariate matching is used because the model produces only one predicted mean. If, on the other hand, the prediction model is a polytomous logistic regression model, multivariate matching is used because the model produces more than one predicted mean

[33] Finer income categories is an example of an imputation set that uses the single RP/single PRD type, but its hot-deck step utilizes multivariate assignment. If the item nonrespondent is missing the finer income category at both the personal and family level, the donor will provide values for both variables in a single hot-deck step. The prediction model is fit using the family-level finer income category.

(i.e., the predicted probability associated with each level of the response variable). In either implementation, there is only one prediction model.

In the single RP/single PRD type, for the univariate assignment case, the item response indicator is based on the single variable that is being assigned in the hot-deck step. If the single variable is missing, the case is an item nonrespondent; otherwise, the case is an item respondent. In the multivariate assignment case, the case is an item respondent if all variables that are assigned in the hot-deck step are nonmissing.

3.4.2 Type 2: Multiple Response Propensity/Multiple Prediction

PMN Type 2, multiple RP/multiple PRD, involves multiple iterations of the single RP/single PRD type. However, for all iterations except the last, the hot-deck step is provisional instead of final and involves univariate matching and univariate assignment.[34] These provisional hot-deck steps tend to be straightforward with respect to constraints and predictive mean vectors, because their only purpose is to fill in missing values so that variables earlier in the sequence can be used as covariates in the RP and PRD models for variables later in the sequence.[35] In the last iteration, a final hot-deck step is completed, where final imputed values are assigned for all variables involved in the models. This final hot-deck step always involves multivariate matching and multivariate assignment. The predicted means from all PRD models are used in this final hot-deck step, and a single record is used to fill all the missing values, thus preserving the relationships among the variables of interest. This is the most refined type of PMN. The recency and frequency variables (within each drug family) follow this type. Figure 3.3 illustrates the multiple RP/multiple PRD type of PMN imputation.

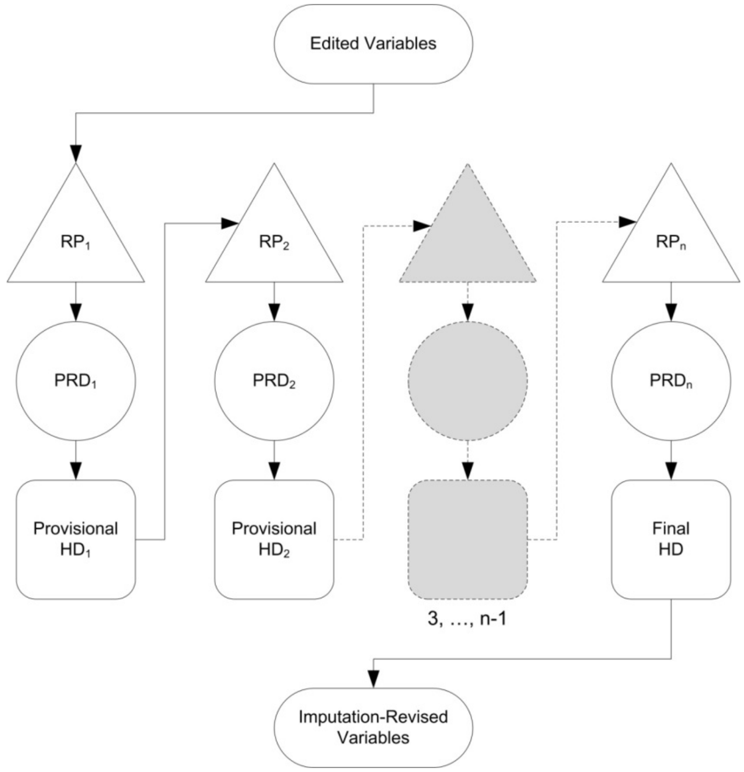
In the multiple RP/multiple PRD type, multiple univariate prediction models are used. The standard approach to multivariate modeling, with a given set of outcome variables (including both discrete and continuous), is likely to be computationally intensive due to the volume of model parameters and the 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, because of the obvious problem of specifying an accurate probability distribution underlying survey data, none of the existing solutions take the survey design into account. In the multiple RP/multiple PRD type, a multivariate model is fitted by a series of univariate parametric models (including the polytomous case), such that variables modeled earlier in the sequence have a chance to be included in the covariate set for subsequent models in the sequence.

---

[34] There is one exception to the rule that provisional hot-deck steps involve univariate matching and univariate assignment. The provisional hot-deck step for cocaine and crack lifetime use utilizes multivariate assignment, since both variables are used in the subsequent PRD model for heroin. The delta constraint refers to both predicted means, but in the calculation of Mahalanobis distance, only the cocaine predicted mean is used. Therefore, with respect to matching, this is not strictly univariate or multivariate; it is a little of both. See Section 6.3.1.5.

[35] There are exceptions. In a few imputation sets that use PMN Type 2 or PMN Type 3 (single RP/multiple PRD), provisional hot-deck steps are not completed because the variables earlier in the sequence are not used as covariates for variables later in the sequence. This occurs for some of the imputation sets for health insurance (Chapter 10) and roster pairs (Chapter 11).

Figure 3.3 PMN Type 2: Multiple Response Propensity/Multiple Prediction



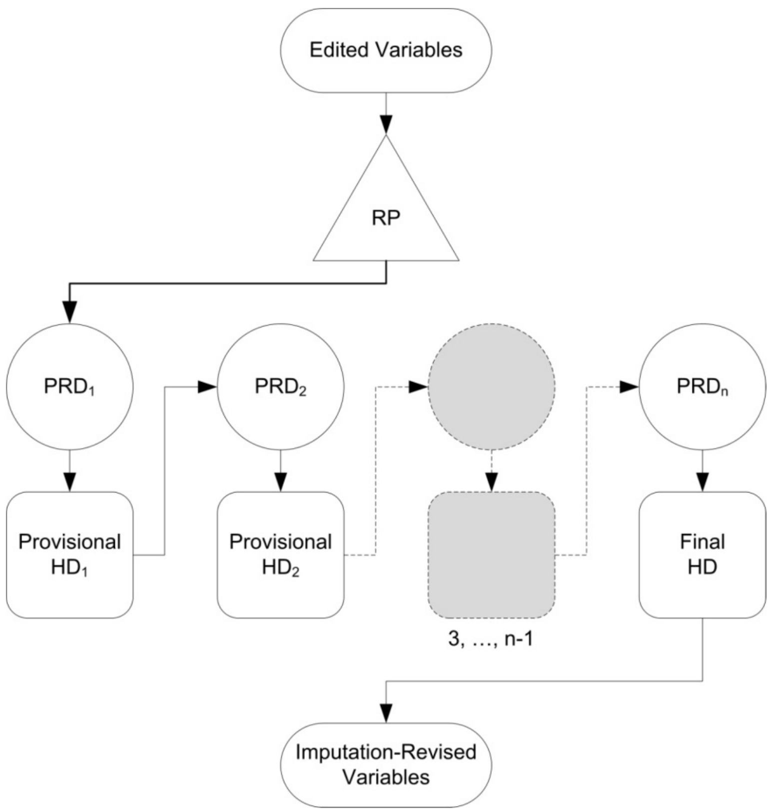
For variables imputed by PMN Type 2 and PMN Type 3 (single RP/multiple PRD), the order in which variables were modeled is of some importance because variables early in the sequence have the potential to be part of the set of covariates for variables later in the sequence, but variables late in the sequence cannot be used for modeling for the earlier variables because of missing values. Note that usually not all variables in the sequence were missing for a particular incomplete record. Nevertheless, models were developed for all the variables in a univariate fashion for reasons mentioned earlier. For the drugs, the sequence of imputation was determined by considering such factors as the level of prejudice and discrimination associated with the drugs, the level of "missingness" in the data (Appendix A), and the degree to which one set of drugs could be used as predictors for other drugs. The decisions on sequencing for other imputation sets were made using similar criteria. For some respondents, some but not all of the variables in the imputation set are missing. This gives rise to missingness patterns (Section 3.3.3.4). Typically, in the final hot-deck step, only the predictive mean vector elements corresponding to missing variables are used to match donors with item nonrespondents. However, likeness constraints (and sometimes logical constraints) are often used to preserve relationships between the missing and nonmissing variables. 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 using these constraints.

The multiple RP/multiple PRD type works well for closely related variables that have different domains and different nonresponse patterns, because the separate RP steps account for these. The recency and frequency variables provide a good example: the domain of the recency models consists of all lifetime users; the domain of the 12-month frequency model (if applicable) consists of all past year users; and the domain of the 30-day frequency model (if applicable) consists of all past month users. The provisional imputation-revised values may be used as covariates in later models, or even may be used to define the domains of later RP models.

3.4.3 Type 3: Single Response Propensity/Multiple Prediction

In PMN Type 3, the single RP/multiple PRD type, a single RP model is applied to all the variables modeled in the PRD steps. This is a less refined version of the preceding type, because it involves the fitting of only one RP model and is not as sensitive to different domains and response patterns among the outcome variables. The same weights are used for all PRD models. The lifetime drug use variables and source-of-income variables are examples of imputation sets that follow this type. Figure 3.4 illustrates the single RP/multiple PRD type of PMN imputation.

Figure 3.4 PMN Type 3: Single Response Propensity/Multiple Prediction



3.5 Special Auxiliary Variables: Age Group and State Rank

The age group and state of residency auxiliary variables apply to several of the imputation sets described in Chapters 4 through 11. Across variable groups, most imputation sets are processed separately by age group, regardless of the type of PMN that was used. The state of residence is used to construct a state-rank variable, which is then used in imputation for the drug variables (Chapter 6) and the income variables (Chapter 9).

3.5.1 Age Groups

The variables related to drug use, household composition, income, and health insurance are highly correlated with age. This, along with the desire to use parallel processing to expedite the time it takes to impute all the variables, led to the decision to separate the imputation procedures for these variables into distinct age groups. Therefore, the drug use variables were imputed within each of three age groups: 12 to 17, 18 to 25, and 26 or older. The household composition (roster-derived), income, and health insurance variables were imputed within the following four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older.[36] The roster pair variables (i.e., the variables related to the relationship between two respondents from the same sampled dwelling unit) were often divided into age groups depending on the ages of both pair members.

In the hot-deck step, the age group restriction could be considered a likeness constraint. However, the models also were built separately within the age groups, so this restriction was not loosened unless no other options were available. Although the demographic variables did not always show a high correlation with age, the imputation of missing values in the demographic variables also was performed within age groups. This was done to maintain consistency with how the other variables were imputed and to facilitate parallel processing. The same three age groups that were used for drugs were also used for demographics. Occasionally, small sample sizes necessitated the aggregation of age groups at the modeling stage. In particular, the models for education level (highest grade completed) were fit within the 12-17 and 18-or-older age groups. In the employment status models, the 15-17 and 18-25 age groups were aggregated and models were fit within the 15-25 and 26-or-older age groups. Finally, all age groups were aggregated for the Hispanic/Latino group, marital status, and immigrant age-of-entry models.

3.5.2 State Rank

Because state-level estimates are an important product of the NSDUH, there has been interest in requiring the donor to be from the same state as the recipient. However, this could not always be implemented because of insufficient pools of donors.[37] In such cases, information about the state of residence of each respondent was incorporated into the modeling and hot-deck steps of the PMN procedure by grouping respondents into three categories based on the ranking of their state of residence. For lifetime drug use, the states were ranked by the weighted proportion of lifetime users of the drug of interest. For recency and frequency of drug use, the states were ranked by the weighted proportion of past month users of the drug of interest. For income, the states were ranked by the weighted proportion of respondents whose personal incomes during the prior calendar year were greater than or equal to \$20,000. These state-rank variables were used as covariates in the RP and PRD steps and sometimes in likeness constraints in the hot-deck step.

---

[36] Age groups were sometimes aggregated for the health insurance procedures. See Chapter 10 for details.

[37] In the hot-deck step for some of the demographic variables, a likeness constraint required the donor to be from the same segment as the item nonrespondent. Segments never cross state lines, so this can be viewed as a refined use of the state of residence. In practice, this constraint often had to be removed because many segments included only a handful of unit respondents.

This page intentionally left blank

4. Editing and Imputation for the NSDUH Core Demographics Variables

4.1 Introduction

As noted in Chapter 1, the NSDUH questionnaire includes both "core" and "noncore" modules. Questions about demographics are included in both of these types of modules. Core demographic questions include age, birth date, gender, marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level (highest grade completed). These questions remain in the survey each year and are important for trend estimates.

This chapter discusses procedures for editing and imputing the core variables for demographic characteristics. Although the interview date was not classified as a core demographic variable, it is used along with a respondent's birth date (when provided by the respondent) as the starting point for determining the respondent's current age. Therefore, the editing procedures for the interview date also are included in this chapter. The noncore demographic variables (including but not limited to immigrant status and employment status) are discussed in Chapter 5.

Prior to imputation, editing was performed on all of these core demographics variables. This editing could range from simply assigning legitimate skip codes (Section 2.4.2), as was the case for marital status, to coding "OTHER, Specify" responses (Section 2.3.1.3) and resolving inconsistencies, as was the case for race and ethnicity.

After editing, the variables were processed in one of four ways:

- No imputation required: interview date, age, gender. These are described in Sections 4.2.1, 4.2.2, and 4.2.4, respectively. The edited data contained no missing values.
- No imputation performed: current state residency reported by interviewers,[38] number of times married, military status,[39] perceived health status. These are described in Sections 4.2.3, 4.2.6, 4.2.8, and 4.2.10, respectively.
- Random assignment: birth date. This procedure is summarized in Section 4.2.5 because it is straightforward and does not involve the predictive mean neighborhood (PMN) method, which is described in Chapter 3.
- PMN: marital status, race, Hispanic/Latino origin, Hispanic/Latino group, education level. These are described in Sections 4.2.6, 4.2.7, and 4.2.9, respectively.

---

[38] The state residency that was reported by interviewers differs from the variable STATE, which contains the Federal information processing standards (FIPS) codes for states and is created at the sampling stage. The variable STATE is recommended for use in state-level analyses.

[39] A final imputation-revised variable is not created for military service. However, a limited non-PMN imputation is performed for military service as part of the imputation procedures for health insurance, as described in Section 10.3.2.4.

Overall, the core demographics variables that are discussed in this chapter either did not have missing data by design or tended to have high item response rates. Except for race, the item response rates when nonresponse was allowed (i.e., "don't know" or "refused") were more than 99 percent.[40]

4.2 Editing Selected Core Demographics Variables

In this section, editing procedures are described for the following:

- interview date (Section 4.2.1);
- age (Section 4.2.2);
- state residency (Section 4.2.3);
- gender (Section 4.2.4);
- birth date (Section 4.2.5);
- marital status and number of times married (if respondents had ever been married) (Section 4.2.6);
- race, Hispanic/Latino origin, and Hispanic/Latino group (Section 4.2.7);
- U.S. military service history and current military status (if respondents had ever been in the U.S. military) (Section 4.2.8);
- education level (Section 4.2.9); and
- perceived health status (Section 4.2.10).

4.2.1 Creating the Edited Interview Date Variable (INTDATE)

The program for the computer-assisted interviewing (CAI) captured information about the time and date at the start of the interview, at the start of each module to which respondents were routed, and at the completion of the interview. This information was recorded in separate variables that were included in the interview data that were transmitted from the field to the data processing staff. These variables containing the start and end dates and times of the interview and modules are referred to as time and date "stamps." For example, the time and date stamps for the beginning of the interview were captured in the variable TBEGINTR, the time and date stamps for the end of the interview were captured in the variable TENDINR, and the time and date stamps for the alcohol questionnaire module were captured in the variable TBEG_ALC. These time and date stamps were sequentially arranged in the dataset in the order that the questionnaire modules appeared. In most cases, the time and date stamps were compared and edited to obtain the final edited interview date (INTDATE). The editing indicator associated with the interview date variable (EIIDATE) specified the module date stamp that was used to create the edited interview date.

[40] When given the opportunity to enter a race, many respondents entered "Hispanic" or some Hispanic/Latino group such as "Mexican," resulting in a comparatively larger amount of missing data for the race question. As a result, the item response rate for the race variables is still high but tends to be about 96 to 97 percent.

In some situations a "breakoff" occurred (i.e., the respondent did not finish the interview in one sitting and resumed the interview on another date), and the respondent's birthday occurred between the beginning and the end of the interview. In these situations, the interview date was set to the end-of-interview date stamp, which was consistent with the first date stamp after the respondent's birthday. (This date stamp was indicated in the CAI.)

A date stamp was not used to set the interview date if any of the following conditions were true:

1. The date stamp was more than 14 days outside the quarter in which the interview was supposed to take place.
2. The date stamp was later in time than a subsequent date stamp.
3. The date stamp occurred before a birthday, which in turn occurred before the end of the interview.

If none of the date stamps were usable, then the interview date was imputed to be sometime in the quarter in which the interview was assumed to have taken place.

4.2.2 Creating the Edited Age Variable (AGE)

The age that respondents reported at the beginning of the interview (CALCAGE) was determined in one of two ways:

1. The respondent reported his or her birth date to the interviewer. This information was captured in the variable AGE1. The interviewer confirmed with the respondent that the birth date information had been recorded correctly and could not proceed further with the interview until the respondent verified the birth date. CALCAGE was determined from AGE1 and the date that was recorded at the start of the interview. The interviewer then confirmed with the respondent that this calculated age was correct. If the respondent indicated that the calculated age was not correct, then the interviewer could not proceed further until any additional corrections to the birth date had been made and the respondent verified both the date of birth as well as the calculated age.
2. If the respondent did not know or refused to report a birth date, the respondent was asked to provide his or her correct age, which was captured in the variable DKREFAGE. In this situation, CALCAGE was set to the age that had been recorded in DKREFAGE.

Interviews were terminated if respondents did not report their age in one of these two ways. Interviews also were terminated if CALCAGE indicated that respondents were less than 12 years old.

Even after CALCAGE has been determined for respondents at the beginning of the interview, respondents could change their age in response to consistency checks in self-administered modules pertaining to their substance use (e.g., if they reported first using a drug at an age that was greater than their current age). Any changes that respondents made to their age during the questionnaire were captured in the temporary variable CURNTAGE. Therefore, it was possible for the age that was recorded for the respondent at the beginning of the interview (CALCAGE) to be different from the value of CURNTAGE that was captured at the end of the interview and stored in the variable NEWAGE. As discussed before, NEWAGE is the final value stored in the variable CURNTAGE at the end of the interview.

The final age variable, AGE, was determined using CALCAGE and NEWAGE and three other sources: the age calculated from the final edited interview date (INTDATE) and the reported birth date (AGE1), the age corresponding to the "self" in the questionnaire household roster (if it existed) (Section 8.2.3), and the pre-interview age that was reported during screening of the dwelling unit (DU).[41] In most situations when the final edited continuous age was determined, priority was given to CALCAGE, NEWAGE, and the age that was calculated from AGE1 and INTDATE. There were occasions, however, where the age corresponding to the "self" in the household roster was used, even if it did not agree with CALCAGE and NEWAGE. If the final age (AGE) did not agree with the date of birth that had originally been entered (AGE1), the birth date also was edited. An intermediate value for age was determined in the following manner:

Intermediate value for age =

- NEWAGE, if nonmissing and exactly equal to CALCAGE, where TBEG_TUT (the interview date time stamp at the beginning of the tutorial) = INTDATE (the edited interview date) (age indicator = 1); else
- NEWAGE, if nonmissing, TBEG_TUT and INTDATE were not equal, but NEWAGE was exactly equal to CALCAGE (adjusted by Blaise[42] to a changed interview date if the interview date was changed within the questionnaire), and the respondent's birthday did not fall between the dates corresponding to TBEG_TUT and INTDATE (age indicator = 1); else
- NEWAGE, if nonmissing, TBEG_TUT and INTDATE were not equal, the respondent's birthday fell between the dates corresponding to TBEG_TUT and INTDATE, the given value of CALCAGE agreed with what it should be based on INTDATE and the given birth date (i.e., EIIDATE not equal to 6), and NEWAGE and CALCAGE were exactly equal (age indicator = 1); else
- age calculated from INTDATE and the reported birth date, if the birth date was nonmissing, TBEG_TUT and INTDATE were not equal, the respondent's birthday fell between the dates corresponding to TBEG_TUT and INTDATE, and the given value of CALCAGE did not agree with what it should be based on INTDATE and the given birth date (EIIDATE = 6), where the newly calculated age based on INTDATE

[41] When contacting the DU, the field interviewer (FI) asked to speak with an adult resident of the household aged 18 or older who could serve as the screening respondent. Using a handheld computer, the FI completed a 5-minute procedure with the screening respondent that involved listing all household members along with their basic demographic data (including age). The computer used the demographic data in a preprogrammed selection algorithm to select zero, one, or two sample people, depending on the composition of the household.

[42] Blaise is the computer program within the CAI instrument that was used to direct the respondent and interviewer through the questionnaire.

was exactly equal to the screener age and/or the roster age (if it existed) (age indicator = 2); else

- NEWAGE, if NEWAGE differed from CALCAGE and NEWAGE = screener age and NEWAGE = roster age (if it existed), and the interview date at the beginning of the interview (TBEGINTR) was within the appropriate quarter (age indicator = 3); else
- CALCAGE, if CALCAGE differed from NEWAGE and CALCAGE = screener age and CALCAGE = roster age (if it existed), and the interview date at the beginning of the interview (TBEGINTR) was within the appropriate quarter (age indicator = 4); else
- age calculated from reported birth date and INTDATE, if EIIDATE = 5 and NEWAGE = CALCAGE (but neither was equal to the correct age) (age indicator = 5); else
- NEWAGE, if NEWAGE differed from CALCAGE, but NEWAGE = roster age, provided roster age existed (age indicator = 6); else
- CALCAGE, if CALCAGE differed from NEWAGE, but CALCAGE = roster age, provided roster age existed (age indicator = 7); else
- NEWAGE, if NEWAGE differed from age calculated from reported birth date and INTDATE, but NEWAGE = CALCAGE, screener age, and roster age (if it existed) (age indicator = 8); else
- CALCAGE, if CALCAGE differed from NEWAGE, but CALCAGE = age calculated from INTDATE and the reported birth date, and CALCAGE was within 1 year of screener age and roster age (age indicator = 9).

After these rules had been applied, this intermediate age value was compared with the age corresponding to the "self" in the household roster. For most respondents, the final edited value for the age variable (AGE) was set to this intermediate age value. There were exceptions, however, as detailed below.

By the time the interviewer reached the roster part of the questionnaire, there had been multiple opportunities to change the value of age in response to consistency checks. This value of age was called CURNTAGE by the Blaise program. One of the consistency checks in the questionnaire household roster was to verify the value of the respondent's own entry for age in the household roster (the "self" entry) against the value of CURNTAGE. If the self age differed from CURNTAGE, then the interviewer could either change the respondent's age that was entered in the roster or override the consistency check and provide an explanation for why the roster age did not match CURNTAGE. If the consistency check for age was overridden, then the value for age corresponding to the "self" in the roster may not match the intermediate age value that was described previously. However, if sufficient explanation was given for overriding the age consistency check, other evidence pointed to the veracity of the roster age, and the difference between CURNTAGE and the roster age for self was less than 2 years, then AGE was set to the roster age, even if it disagreed with both NEWAGE and CALCAGE. In particular, all of the following conditions had to be met for this to occur:

1. The interviewer specifically indicated that the roster age was the correct one.
2. The pre-interview screener age matched the roster age.
3. The other household member's roster supported the roster age value, if another member of the household completed the interview.

It also was possible for interviewers to jump back from the household roster or elsewhere in interviewer-administered sections (i.e., after respondents had completed the audio computer-assisted self-interviewing [ACASI] section of the interview) to change the respondent's date of birth. In this situation, the values of CALCAGE and NEWAGE could be consistent, but these changes to the respondent's age between the beginning and end of the interview would not be readily apparent. Rather, this pattern would become evident in reviewing ACASI data. For example, if respondents were aged 21 or older at the beginning of the interview and they did not change their age during the ACASI portion of the interview, questions were skipped in the noncore consumption of alcohol module about alcohol use by people aged 12 to 20. However, if their final age indicated they were aged 12 to 20, then they would have been eligible to be asked questions about alcohol use by people aged 12 to 20 (see Sections 7.3 and 7.4.17).

Three age category variables were created from the final age: CATAGE with four levels (12 to 17, 18 to 25, 26 to 34, and 35 or older); CATAG2 with three levels (12 to 17, 18 to 25, and 26 or older); and CATAG3 with five levels (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). These variables were used instead of the continuous age variables in some subsequent imputations and analyses.

4.2.3 Creating the Edited Current State Residency Variable (STATELOC)

Field interviewers (FIs) were instructed to use the FI "checkpoint" FIPE4 at the beginning of the interview to report the state where the DU was located. The term "checkpoint" refers to an item that was completed by the FI about the location of the DU or characteristics of the sample within the DU. These checkpoints were not administered to the respondents but were used to customize the wording of questions in subsequent sections of the interview, such as state-specific names for Temporary Assistance for Needy Families (TANF) programs (Section 9.2.3).

The state that interviewers entered in FIPE4 sometimes failed to match the state residence information that was used to sample a given case. These mismatches were investigated by field staff during data collection. Some of these mismatches existed for a valid reason, such as if a respondent had been selected in a DU in one state but had moved to another state. In these situations, if FIPE4 reflected the state where the respondent was currently living, the edited variable STATELOC (corresponding to FIPE4) retained the value from FIPE4. Otherwise, if the state information in FIPE4 was entered incorrectly, STATELOC was assigned a code for bad data (Section 2.4.3). In turn, the setting of STATELOC to bad data affected the editing of other variables that were dependent on the state where the respondent was reported to be a resident.

4.2.4 Creating the Edited Gender Variable (IRSEX)

As with previous surveys since 2002, FIs in 2014 were required to enter the respondent's gender in question QD01. As a result, it was not possible to have missing values for this question. To maintain continuity with surveys prior to 2002 in which missing data could exist for gender, the variable name IRSEX was used to describe gender in the 2014 survey. However, it was not necessary to create an imputation indicator, because IRSEX and QD01 were equivalent.

As for the situation described previously for age, interviewers also could jump back to the beginning of the interview and change the respondent's gender based on roster information. This type of change to QD01 between the beginning and end of the interview would not be readily apparent. However, this pattern would become evident in reviewing ACASI data. For example, if respondents were defined as being male at the beginning of the interview, questions in the noncore health care module about pregnancy were skipped. However, if their final gender and age indicated that they were female and aged 12 to 44, then they would have been eligible to be asked questions about pregnancy (see Sections 7.3 and 7.4.9).

4.2.5 Creating the Edited Birth Date Variable (BRTHDATE)

To proceed with the interview, respondents were required to provide their date of birth or current age (if they did not provide their date of birth) at the beginning of the interview. Each completed case respondent possessed a current age, although a number of cases had missing birth dates. If the birth date was nonmissing but was inconsistent with AGE and INTDATE (either in the unedited data or as a result of editing age and/or interview date), then the reported birth month and day were preserved, and the birth year was logically edited according to the interview date and age.

In cases with missing birth dates, a birth date was randomly selected from all possible birth dates, given the final age and interview date. Each date in this period (365 or 366 days, depending on whether the period includes February 29 in a leap year) had an equal probability of selection.

4.2.6 Creating the Edited Marital Status Variables (MARITAL, EDMARIT, and NOMARR)

In the 2014 questionnaire, a single core question (QD07) asked about the respondent's marital status if respondents were aged 15 or older. If respondents indicated that they were currently married, widowed, or divorced or separated, they also were asked to report the number of times they had been married (QD08). The exact wording of the questions was as follows:

QD07: Are you now married, widowed, divorced or separated, or have you never married?

1 MARRIED
2 WIDOWED
3 DIVORCED OR SEPARATED
4 HAVE NEVER MARRIED

QD08: [IF QDO7 = 1 OR 2 OR 3] How many times have you been married?

NUMBER OF TIMES: _________ [RANGE: 1 - 9]

MARITAL and NOMARR underwent minimal processing. Legitimate skip codes were assigned to both variables (Section 2.4.2) if respondents were aged 12 to 14. Similarly, legitimate skip codes were assigned to NOMARR if respondents had never been married. The base variable for creating an imputation-revised version of marital status was called EDMARIT. This variable was equivalent to MARITAL, with the exception that all legitimate skip codes were collapsed into a single legitimate skip code (99), and missing values were set to the SAS[43] missing code (.) so that they could be properly handled by the modeling programs.

4.2.7 Creating the Edited Race and Hispanic/Latino (Origin and Group) Variables

In the 2014 questionnaire, two core questions focused on the respondent's ethnicity[44] (QD03 and QD04) and two focused on the respondent's race (QD05 and QD05ASIA). For those questions with multiple categories (QD04, QD05, and QD05ASIA), the respondent had the opportunity to select more than one category. Two more Hispanic/Latino group categories were added to QD04 since the 2004 survey: Dominican (from Dominican Republic) and Spanish (from Spain). These new categories were added to the survey because of the large number of "OTHER, Specify" responses in previous NSDUHs that mapped to these categories.

The questions as they appear in the survey instrument are presented below.

QD03: Are you of Hispanic, Latino, or Spanish origin or descent?

1 YES
2 NO

QD04: (Asked only if QD03 = 1) Which of these Hispanic, Latino, or Spanish groups best describes you?

1 MEXICAN, MEXICAN AMERICAN, MEXICANO, OR CHICANO
2 PUERTO RICAN
3 CENTRAL OR SOUTH AMERICAN
4 CUBAN OR CUBAN AMERICAN
5 DOMINICAN (FROM DOMINICAN REPUBLIC)
6 SPANISH (FROM SPAIN)
7 OTHER (SPECIFY)

QD05: Which of these groups describes you?

1 WHITE
2 BLACK OR AFRICAN AMERICAN
3 AMERICAN INDIAN OR ALASKA NATIVE (AMERICAN INDIAN INCLUDES NORTH AMERICAN, CENTRAL AMERICAN, AND SOUTH AMERICAN INDIANS)
4 NATIVE HAWAIIAN
5 GUAMANIAN OR CHAMORRO

---

[43] SAS® software is a registered trademark of SAS Institute Inc.

[44] The questions about ethnicity were limited to determining whether a respondent was Hispanic/Latino or not, and the specific Hispanic/Latino group to which a Hispanic/Latino respondent belonged.

6 SAMOAN
7 OTHER PACIFIC ISLANDER
8 ASIAN (INCLUDING: ASIAN INDIAN, CHINESE, FILIPINO, JAPANESE, KOREAN, AND VIETNAMESE)
9 OTHER (SPECIFY)

QD05ASIA: (Asked only if QD05=8) Which of these Asian groups describes you?

1 ASIAN INDIAN
2 CHINESE
3 FILIPINO
4 JAPANESE
5 KOREAN
6 VIETNAMESE
7 OTHER (SPECIFY)

As stated in the guidelines from the Office of Management and Budget (OMB),[45] "Hispanic/Latino" was categorized as an ethnicity, not a race. However, when given the opportunity to enter a race, many respondents entered "Hispanic" or some Hispanic/Latino group, resulting in missing data for the race question. Even though the final drug use tables were cross-classified with a variable that combined race and ethnicity, separate variables were initially created for race and ethnicity, and the race/ethnicity variables used in the tables were derived from these separate variables.

Due to the relationship between Hispanicity and race reporting, Hispanicity was used in the editing of race, and vice versa. In the process of editing race, the "OTHER, Specify" response to the Hispanic/Latino group question (QD04) was consulted (if it existed) if no race information was identified in QD05 or QD05ASIA. Similarly, in the process of editing the Hispanic/Latino group, the "OTHER, Specify" responses to the race questions (QD05 and QD05ASIA) were consulted (if they existed) if no Hispanic/Latino group information was identified in QD04. Because of the interdependence of race and Hispanicity, the editing of these variables is discussed together in this section.

The procedures used to edit the race and Hispanicity variables in the surveys since 2008 differed in several ways from the procedures used in previous surveys. One of the major differences was in the handling of race for multiple-race respondents. The first procedural changes were triggered by the elimination of the QD06 question, which appeared in the survey from 1999 to 2002. QD06 asked respondents who selected more than one racial category from QD05 and QD05ASIA combined to choose the race with which they identified the most. Without this question, it was impossible to determine (directly) the single race that a given multiple-race respondent would most closely identify for himself or herself. In the 2003-2007 surveys, QD06 responses were "simulated" based on models built using true QD06 responses

[45] In October 1997, the OMB released a notice, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" (OMB, 1997) that provides new standards for maintaining, collecting, and presenting Federal data on race and ethnicity.

from the 2000-2002 surveys.[46] However, because racial demographics in the United States had changed since the 2000 survey and because recent data that were needed to update these models were not available, this method was not used after 2008 and single races were not assigned for multiple-race respondents. Refer to Section 3.3 of the 2008 imputation report (Ault et al., 2010) for more details.

4.2.7.1 Categories Used in Race and Hispanic/Latino Variables

4.2.7.1.1 Race Categories

For editing purposes, the 7 specific categories in QD05 (white, black or African American, American Indian or Alaska Native, Native Hawaiian, Guamanian or Chamorro, Samoan, and Other Pacific Islander) and the 6 specific categories in QD05ASIA (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese) were combined to produce 13 racial categories. Two other categories also were created: "Other Asian" (where the responses to QD5ASIA did not fit into the above category) and "Asian nonspecific" (where no response was selected to QD05ASIA, even though Asian was selected in QD05). Respondents could choose almost any subset of these categories. The only subsets that were not logically possible were those that included "Asian nonspecific" in combination with one or more specific Asian categories. Combining the information from QD05 and QD05ASIA, as well as QD04 when necessary, allowed for the creation of all the edited and imputation-revised race variables.

4.2.7.1.2 Hispanic/Latino Categories

With the addition of two Hispanic/Latino categories since the 2004 survey, respondents were given the choice of seven categories in QD04 (Mexican/Mexican American/Mexicano/Chicano, Puerto Rican, Central or South American, Cuban/Cuban American, Dominican (from Dominican Republic), Spanish (from Spain), or some other Hispanic/Latino group),[47] and they could choose more than one category. As with QD05, interviewers could manually enter the alternative to the choices given, which would be either coded to some subset of the existing seven categories or set to missing. The "OTHER, Specify" responses to QD05 and/or QD05ASIA, if nonmissing, were consulted if no Hispanic/Latino origin group information was available from QD04. The final imputation-revised Hispanic/Latino group variable, IRHOGRP4, included all seven Hispanic/Latino group levels and a legitimate skip code (99) for respondents who were not Hispanic/Latino.

4.2.7.2 Classification of "OTHER, Specify" Codes

All "OTHER, Specify" responses from QD04, QD05, and QD05ASIA were assigned both a race code and a Hispanic/Latino code. Each of these codes was mapped to at least one of

---

[46] Because of the questionnaire differences between the 1999 survey and the 2000-2002 surveys, the procedure for simulating QD06 responses for the 2003-2007 surveys was made simpler by limiting the QD06 results from the 2000-2002 surveys. During the 2003-2007 surveys, for the purpose of allocating respondents into imputation classes, a model was used to select a single race for respondents who had selected more than one race (IRRACE2). The selection of a single race was based on models that were fit using data from the 2000-2002 surveys. This method is described in Appendix E of the 2007 imputation report (Ault et al., 2009).

[47] When listing the six Hispanic/Latino defined categories in QD04, they shall henceforth be listed in this chapter as Mexican, Puerto Rican, Central or South American, Cuban, Dominican, and Spanish.

the categories described in Section 4.2.7.1 and in this section, or to some other code that was informative in the final imputation described in Sections 4.3.2 and 4.3.3. A summary of categories of "OTHER, Specify" codes and how they were handled is given in the following sections. Appendix C provides the individual "OTHER, Specify" codes and more details about how they were handled.

4.2.7.2.1 Mapping of Race "OTHER, Specify" Codes to Edited Values

This section describes the directly and indirectly mapped race codes. The edits following from either of these types of mapped codes resulted in values that were considered "final" in that no imputation was necessary for them.

The directly mapped codes were mapped to one or more of the categories given in the questionnaire (Section 4.2.7). There were directly mapped racial category codes and directly mapped geographic category codes. Racial category codes were exactly equivalent to one or more categories in QD05 or QD05ASIA, and were mapped directly to those categories regardless of whether the write-in response was in QD05 or QD05ASIA. (Respondents were still considered at least part Asian, even if the write-in response in QD05ASIA was non-Asian. The racial makeup of a respondent who entered a non-Asian racial category in QD05ASIA was determined on a case-by-case basis.) For example, a response such as "Han" mapped directly to a category in QD05ASIA ("Chinese"), and a response such as "mestizo" mapped directly to two categories in QD05, "white" and "Native American."

By contrast, geographic category codes corresponding to a country where census data indicated a racially homogeneous society depended on the corresponding question. For example, an entry of "Polish" in QD05 mapped to white because the Polish census data indicated nearly all Poles were white. On the other hand, an entry of "Polish" in the QD05ASIA "OTHER, Specify" mapped to "Other Asian." Geographic category codes also included ethnic groups where the racial identification was not immediately obvious. For example, a response of "Arab" would be automatically mapped to "white" if the response was a write-in response for QD05. However, as with the "Polish" entry, if the "Arab" response was a write-in response in QD05ASIA, the respondent was considered "Other Asian."

Indirect mapping was used for countries that were racially heterogeneous. A racial category was chosen by generating a random number and allocating the race based on a comparison of the random number with the proportions of races in the country's census.[48] For example, an entry of "Bolivian" would have a 55 percent chance of being allocated to the American Indian/Alaska Native category, because the latest Bolivian census indicated 55 percent of Bolivians were American Indian/Alaska Native. For countries where the census indicated a small proportion of some indistinct category such as "other" and the randomly generated number indicated an allocation to this proportion, the final race was left to imputation (appropriately constrained based upon the indistinct response).

If two or three heterogeneous countries were entered in the "OTHER, Specify" response (e.g., Bolivian and Peruvian), the final race was allocated using the following procedure:

---

[48] See https://www.cia.gov/library/publications/the-world-factbook/fields/2075.html (Ethnicity and Race by Countries) for more information.

(1) randomly assign races based on the proportions for each country mentioned; and (2) combine the results. Exceptions to these rules occurred with the categories Mexicans, Puerto Ricans, Cubans, Dominicans, Central or South Americans (no country listed), and Spanish, which were given codes described under the next heading, with a final value determined using the formal imputation procedures described in Section 4.3.2. Starting with the 2006 survey, the imputation processing of indirectly mapped codes obtained from QD05ASIA has been simplified. In prior survey years, this type of write-in response was mapped to a race through country census information; since the 2006 NSDUH, all census-based write-in responses to the Asian race question were mapped directly to the "Other Asian" racial category.

4.2.7.2.2 Mapping of Race "OTHER, Specify" Codes to Inform Imputation

"OTHER, Specify" responses that could not be mapped definitively to a specific race category resulted in incomplete values requiring imputation. These responses were assigned two types of codes, either informative or noninformative, for the formal imputation procedures for race described in Section 4.3.2.

Responses that provided information were used to limit the final imputation. For example, a response of "mixed" resulted in an imputation among donors with two or more races, and a response of "brown" resulted in an imputation among donors who were not single race white.

A noninformative response (e.g., American) that was not accompanied by a response to one of the precoded categories[49] in QD05 or QD05ASIA (i.e., those other than the "OTHER, Specify" response) resulted in an unrestricted imputation.

4.2.7.2.3 Subsequent Editing of Race "OTHER, Specify" Codes

Subsequent to the initial mapping of the race "OTHER, Specify" codes, edits were sometimes implemented that revised or clarified the initial mapping before final races were allocated. These edits were necessary if multiple sources of information, including "OTHER, Specify" responses, provided conflicting or confusing information. These edits were implemented when (1) the final mapping depended upon the source question (i.e., QD04, QD05, and QD05ASIA); (2) the responses were given to both the "OTHER, Specify" and precoded categories of QD05 or QD05ASIA; or (3) the different "OTHER, Specify" responses were present in at least two of QD04, QD05, and QD05ASIA. In some cases, it was necessary to individually examine the responses to determine the appropriate mapping.

Occasionally, the final mapped value depended upon whether the "OTHER, Specify" code was in QD04, QD05, or QD05ASIA. An example from directly mapped codes is "Indian." This response would be mapped to "American Indian/Alaska Native" if the "OTHER, Specify" response was in QD05, but it would be mapped to "Asian Indian" if the "OTHER, Specify" response was in QD05ASIA. Indirectly mapped codes also could depend upon the source question. The census data from many countries included Asian categories. If the "OTHER,

[49] For example, precoded categories in QD05 in 2014 were 1 = White; 2 = Black or African American; 3 = American Indian or Alaska Native; 4 = Native Hawaiian; 5 = Guamanian or Chamorro; 6 = Samoan; 7 = Other Pacific Islander; and 8 = Asian. Category 9 (Other) led to the "OTHER, Specify" question.

Specify" response was in QD05ASIA, the random imputation to a census category was limited to the Asian categories. "OTHER, Specify" responses that were not specifically Asian sometimes occurred in the "OTHER, Specify" category of QD05ASIA. These were carefully examined, but the "Asian" part of the response was always preserved.

If "OTHER, Specify" responses to QD05 or QD05ASIA accompanied responses to the precoded categories in QD05 and QD05ASIA, it was necessary to reconcile these responses. In some cases, the combination of responses mapped to one of the multiple race categories. For example, if a respondent selected "black/African American" in QD05 and wrote in "black and American Indian," then the respondent would be assigned both racial categories "black/African American" and "American Indian/Alaska Native."

There were instances, however, when the "OTHER, Specify" response was ignored because of responses to the precoded categories in QD05 and QD05ASIA. In particular, the "OTHER, Specify" response was always ignored if a precoded response category was selected, and the "OTHER, Specify" response was a geographic category code.[50] For example, if the interviewer selected the category for "black/African American" for the respondent and also wrote in "Polish," it was assumed that the respondent was a black Pole and, for racial identification purposes, was considered single-race black/African American. This was true even though the Polish census did not identify significant numbers of nonwhite people in the Polish population.

In some instances, it was necessary to reconcile the "OTHER, Specify" responses to QD04, QD05, and QD05ASIA. In these cases, the responses were examined on an individual basis, and sometimes a new code was assigned that more accurately reflected the situation.

4.2.7.2.4 Mapping of Hispanic/Latino "OTHER, Specify" Codes

Certain Hispanic/Latino codes were considered "Definitely Hispanic." If any of these appeared in QD05 or QD05ASIA, the respondent was considered Hispanic/Latino regardless of the response to QD03. Examples included "Hispanic" and "Dominicano" (Spanish for "Dominican"). There was also a code to handle respondents who were definitely not Hispanic/Latino (i.e., the respondent reported "Not Hispanic/Latino"). If this code appeared in QD04, QD05, or QD05ASIA, then the respondent was considered non-Hispanic/Latino regardless of the response to QD03. All other Hispanic/Latino codes either mapped directly to one or more of the seven Hispanic/Latino group categories or provided no new information (e.g., Hispanic).

4.2.7.3 Edited Race Variables

4.2.7.3.1 Individual Race Categories (EDQD051-EDQD0515)

Edited variables were created that correspond to the 15 racial categories described in Section 4.2.7.1.1. These variables were called EDQD05*xx*, where *xx* represented a number between 1 and 15, corresponding to each of the 15 categories.

[50] Actually, this "edit" was not "subsequent" to the initial mapping. Instead, the initial mapping was ignored under the circumstances described.

EDQD05*xx* =

- 1, if the level *xx* was selected by the respondent in QD05 or QD05ASIA; else
- 2, if the level *xx* was indicated by a directly mapped code in QD05 or QD05ASIA; else
- 3, if no EDQD05*xx* variables had values of 1 or 2, and the level *xx* was indicated by a directly mapped code in QD04 (Hispanic/Latino status); else
- 4, if (a) no EDQD05*xx* variables had values of 1, 2, or 3, and (b) the level *xx* was indicated by an indirectly mapped code in QD04, QD05, and/or QD05ASIA; else
- missing.

EDQD0515 (Asian nonspecific) was a little different from the others. In particular, there was no specific level of QD05 or QD05ASIA that corresponded to it. It was used mainly to preserve a response of "Asian" to QD05, even if the respondent selected nothing in QD05ASIA. The value of EDQD0515 was set to 1 if the respondent selected "Asian" in QD05 but mentioned nothing that mapped to a specific Asian category in QD05ASIA. It also could have values of 2, 3, or 4, depending on the "OTHER, Specify" codes.[51]

4.2.7.3.2 Broad Categories of Race (EDRACE)

The EDRACE is a 25-level variable that indicates which of four broad racial categories (white, black/African American, American Indian/Alaska Native, Asian/Other Pacific Islander) were identified in QD04, QD05, and QD05ASIA, and it also has levels to indicate how the imputation should be restricted based on the race of the donor. The first three broad racial categories corresponded to EDQD051, EDQD052, and EDQD053, respectively. "Asian/Other Pacific Islander" was considered to have been identified if any of EDQD054 through EDQD0515 was nonmissing. EDRACE was created using the following rules, under five possible scenarios:

Scenario 1: If only one broad racial category was identified in QD04, QD05, and/or QD05ASIA, EDRACE =

- 1 (white only), if EDQD051 was nonmissing; else
- 2 (black/African American only), if EDQD052 was nonmissing; else
- 3 (American Indian/Alaska Native only), if EDQD053 was nonmissing; else
- 4 (Asian/Other Pacific Islander only), if any of EDQD054 through EDQD0515 were nonmissing.

[51] A value of 2 indicated that the respondent wrote "Asian" in the QD05 other-specify blank. A value of 3 indicated that the response was obtained from the other-specify part of the Hispanic/Latino group question (QD04). Finally, a value of 4 indicated that the respondent gave a country of origin as a response to QD05, and the census for that country had "Asian" as one of its categories.

Scenario 2: If two broad racial categories were identified in QD04, QD05, and/or QD05ASIA, EDRACE =

- 5 (white and black/African American only), if both EDQD051 and EDQD052 were nonmissing; else
- 6 (white and American Indian/Alaska Native only), if both EDQD051 and EDQD053 were nonmissing; else
- 7 (white and Asian/Other Pacific Islander only), if EDQD051 was nonmissing and at least one of EDQD054 through EDQD0515 were nonmissing; else
- 8 (black/African American and American Indian/Alaska Native only), if both EDQD052 and EDQD053 were nonmissing; else
- 9 (black/African American and Asian/Other Pacific Islander only), if EDQD052 was nonmissing and at least one of EDQD054 through EDQD0515 were nonmissing; else
- 10 (American Indian/Alaska Native and Asian/Other Pacific Islander only), if EDQD053 was nonmissing and at least one of EDQD054 through EDQD0515 were nonmissing.

Scenario 3: If three broad racial categories were identified in QD04, QD05, and/or QD05ASIA, EDRACE =

- 11 (white, black/African American, and American Indian/Alaska Native only), if all of EDQD051 through EDQD053 were nonmissing; else
- 12 (white, black/African American, and Asian/Other Pacific Islander only), if both EDQD051 and EDQD052 were nonmissing and at least one of EDQD054 through EDQD0515 were nonmissing; else
- 13 (white, American Indian/Alaska Native, and Asian/Other Pacific Islander only), if both EDQD051 and EDQD053 were nonmissing and at least one of EDQD054 through EDQD0515 were nonmissing; else
- 14 (black/African American, American Indian/Alaska Native, and Asian/Other Pacific Islander only), if both EDQD052 and EDQD053 were nonmissing and at least one of EDQD054 through EDQD0515 were nonmissing.

Scenario 4: If all four broad racial categories were identified in QD04, QD05, and/or QD05ASIA, EDRACE = 15.

Scenario 5: If none of the broad racial categories were identified in QD04, QD05, and/or QD05ASIA, EDRACE =

- 16 (multiple race, no other information), if an "OTHER, Specify" answer such as "biracial" or "mixed" appeared in QD04, QD05, or QD05ASIA; else
- 17 (nonwhite, no other information), if an "OTHER, Specify" answer such as "brown," "tan," or similar answers in Spanish appeared in QD04, QD05, or QD05ASIA; else

- 18 (white, or both white and American Indian/Alaska Native), if the random assignment of a census data code resulted in imputation restricted to donors who were either white, or both white and American Indian/Alaska Native; else
- 19 (not American Indian/Alaska Native, in part or in full), if the random assignment of a census data code resulted in imputation restricted to donors who were not American Indian/Alaska Native, in part or in full; else
- 20 (non-Hispanic Mexican), if "Mexican" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 21 (non-Hispanic Cuban), if "Cuban" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 22 (non-Hispanic Central or South American), if "Central or South American" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 23 (non-Hispanic Dominican), if "Dominican" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 24 (non-Hispanic Spanish), if "Spanish" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 25 (non-Hispanic Puerto Rican), if "Puerto Rican" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- missing.

4.2.7.3.3 Finer Categories of Race (EDNWRACE)

EDNWRACE was a 15-level edited variable used as a base variable for the imputation-revised finer racial category variable IRNWRACE. It also had a 16th level to indicate when the imputation should be restricted to Asian-specific categories. It was created using the following rules, under three possible scenarios:

Scenario 1: If only one of EDQD051 through EDQD0515 was nonmissing,

EDNWRACE =

- 16 (Asian nonspecific only), if EDQD0515 was the nonmissing variable; else
- *xx* (one known racial category only), where EDQD05*xx* was the nonmissing variable out of EDQD051 through EDQD0514.

Scenario 2: If more than one of EDQD051 through EDQD0515 was nonmissing,

EDNWRACE =

- 13 (Native Hawaiian and Other Pacific Islander only where Other Pacific Islander includes Guamanian/Chamorro and Samoan), if both EDQD054 and one or more of EDQD055, EDQD056, or EDQD057 were nonmissing, and all other EDQD05*xx* variables were missing; else

- 14 (Asian multiple category), if all of EDQD051 through EDQD057 were missing (i.e., at least two of the ordinary Asian categories were selected); else
- 15 (more than one race).

Scenario 3: If all of EDQD051 through EDQD0515 were missing,

EDNWRACE =

- 15 (more than one race), if EDRACE = 16; else
- missing.

4.2.7.4 Edited Hispanic/Latino Variables

4.2.7.4.1 Hispanic/Latino Indicator (EDHOIND)

An edited Hispanic/Latino indicator, EDHOIND, was created using responses to QD03 and, in rare cases, the "OTHER, Specify" responses to QD04, QD05, and/or QD05ASIA. This indicator variable was created as follows:

EDHOIND =

- 1 (Hispanic/Latino), if QD03 = 1 and no "OTHER, Specify" response stated that the respondent was definitely not Hispanic/Latino, or if the "OTHER, Specify" response to QD05 or QD05ASIA indicated that the respondent was definitely Hispanic/Latino; else
- 2 (not Hispanic/Latino), if QD03 = 2 and no "OTHER, Specify" response stated that the respondent was definitely Hispanic/Latino, or if the "OTHER, Specify" response to QD04, QD05, and/or QD05ASIA indicated that the respondent was definitely not Hispanic/Latino; else
- missing.

The race "OTHER, Specify" responses, which were considered "definitely Hispanic/Latino," and the single Hispanic/Latino "OTHER, Specify" response, which was considered "definitely not Hispanic/Latino," are listed in Appendix C.

4.2.7.4.2 Individual Hispanic/Latino Group Categories (EDQD041-EDQD047)

The edited variables EDQD041 through EDQD047 were created to match the seven Hispanic/Latino group categories described in Section 4.2.7.1.2: Mexican, Puerto Rican, Central or South American, Cuban, Dominican, Spanish, and Other Hispanic/Latino.

EDQD04*xx* =

- 1, if the level *xx* was selected by the respondent in QD04; else
- 2, if the "OTHER, Specify" response from QD04 mapped directly to level *xx*; else

- 3, if no EDQD04*xx* variables had values of 1 or 2, and the "OTHER, Specify" response from QD05 or QD05ASIA mapped directly to level *xx*; else
- missing.

4.2.7.4.3 Edited Hispanic/Latino Group (EDHOGRP)

The edited variable EDHOGRP was the base variable for creating an imputation-revised Hispanic/Latino group variable. It had seven levels to match the seven Hispanic/Latino group categories described in Section 4.2.7.1.2, plus several other more general Hispanic/Latino levels that could be used in a restricted imputation. Those respondents with EDHOIND = 2 were assigned EDHOGRP = 99. It was created using the following rules, under four possible scenarios:

Scenario 1: If EDHOIND = 2,

EDHOGRP = 99.

Scenario 2: If EDHOIND = 1 or missing and only one of EDQD041 through EDQD047 was nonmissing,

EDHOGRP = *xx*, where EDQD04*xx* was the nonmissing one.

Scenario 3: If EDHOIND = 1 or missing and more than one of EDQD041 through EDQD047 was nonmissing,

EDHOGRP =

- 1 (Mexican), if EDQD041 was nonmissing; else
- 2 (Puerto Rican), if EDQD042 was nonmissing; else
- 3 (Central or South American), if EDQD043 was nonmissing; else
- 4 (Cuban), if EDQD044 was nonmissing; else
- 5 (Dominican), if EDQD045 was nonmissing; else
- 6 (Spanish), if EDQD046 was nonmissing; else
- 7 (Other), if EDQD047 was nonmissing.

For the multiple Hispanic/Latino group respondents, a priority rule similar to the one used in the surveys prior to 2004 was applied in determining a single Hispanic/Latino group. The only difference is the addition of two more Hispanic/Latino group categories since the 2004 survey, resulting in the following order: Mexican, Cuban, Puerto Rican, Central or South American, Dominican, Spanish, and Other Hispanic/Latino.

Scenario 4: If EDHOIND = 1 or missing and all of EDQD041 through EDQD047 were missing, EDHOGRP =

- EDRACE + 7 (imputation restricted by race), if $1 \leq$ EDRACE ≤ 14; else
- missing.

4.2.8 Creating the Edited Military Service History and Military Veteran Status Variables

4.2.8.1 Current Service Variables (SERVICE and MILSTAT)

SERVICE and MILSTAT were created from QD09 and QD10, respectively, which asked respondents aged 17 or older about their current and past service in the U.S. military. The questions as they appear in the survey instrument are presented below.

QD09: Have you ever been in the United States Armed Forces?

1 YES
2 NO

QD10: Are you **currently** on **active** duty in the United States Armed Forces, in a Reserve component, or now separated or retired from either reserves or active duty?

1 ON ACTIVE DUTY IN THE ARMED FORCES
2 IN A RESERVES COMPONENT
3 NOW SEPARATED OR RETIRED FROM EITHER RESERVES OR ACTIVE DUTY

Respondents who were currently on active duty in the U.S. military were not eligible to be interviewed for NSDUH. If respondents reported in QD10 that they were currently on active military duty, the interviewers were asked to confirm this answer with the respondents. The interview was terminated if respondents confirmed that they were on active duty in the U.S. military. Consequently, there were no final respondents in the final NSDUH data who reported that they currently were on active military duty. However, some final respondents could be civilians who were currently in the military reserves or were separated or retired from the military. In addition, the industry and occupation variables in the noncore employment section could include military-related codes for some respondents (Section 5.2.4)

Unlike the situation in most places in the interview (Section 2.4.2), responses of "don't know" or "refused" to the question about lifetime military service were treated as potential indications of military service. Thus, respondents who did not know or refused to report whether they had ever been in the U.S. military also were asked QD10 in order to determine their eligibility status for the interview.

Legitimate skip codes were assigned to SERVICE and MILSTAT if respondents were aged 12 to 16. In addition, legitimate skip codes were assigned to MILSTAT if respondents were aged 17 or older but they reported in QD09 that they had never been in the U.S. armed forces.

Two questions about respondents' current service in a military reserve component have been included in the core demographics module since 2013, as shown below.

V2A Are you a member of a Reserve component currently serving full-time in an active duty status?

1 YES
2 NO
DK/REF

V2B Are you currently serving full-time in a Reserve component? Full-time service does not include annual training for the Reserves or National Guard.

1 YES
2 NO
DK/REF

Question V2A was administered if respondents reported in question QD10 that they were currently on active duty in the U.S. armed forces. Because respondents who reported in question QD10 and confirmed on follow-up that they were currently on active duty were not eligible to be interviewed for NSDUH, no final NSDUH respondents had data for question V2A. Consequently, an edited variable was not created that corresponded to this question.

Question V2B was administered if respondents reported in question QD10 that they were serving in a reserves component. The edited variable MILRSRV corresponded to this question. As for SERVICE and MILSTAT, legitimate skip codes were assigned to MILRSRV if respondents were aged 12 to 16. Legitimate skip codes also were assigned to MILRSRV if respondents were aged 17 or older but they reported in QD09 that they had never been in the U.S. armed forces or they reported in QD10 that they were now separated or retired from the reserves or active duty.

4.2.8.2 Military Veteran Status Variables

Questions also have been included in the core demographics module since 2013 for respondents who had ever served in the U.S. armed forces regarding whether they ever served on active duty, the period(s) when they served, and whether they served in a military combat zone where they drew imminent danger pay or hostile fire pay. These questions are shown below.

QD10A Have you ever served on **active duty** in the United States Armed Forces or Reserve components?

Active duty does not include training for the Reserves or National Guard, but **does** include activation, for example, for a national emergency or military conflict.

1 YES
2 NO
DK/REF

QD10B When did you serve on active duty in the United States Armed Forces or Reserve components? Just give me the number or numbers from the card.

TO SELECT MORE THAN ONE CATEGORY, PRESS THE SPACE BAR BETWEEN EACH CATEGORY YOU SELECT.

1 September 2001 or later
2 August 1990 to August 2001 (including Persian Gulf War)
3 May 1975 to July 1990
4 Vietnam era (March 1961 to April 1975)
5 February 1955 to February 1961
6 Korean War (July 1950 to January 1955)
7 January 1947 to June 1950
8 World War II (December 1941 to December 1946)
9 November 1941 or earlier

QD10C Did you ever serve on active duty in the United States Armed Forces or Reserve components in a military combat zone or an area where you drew imminent danger pay or hostile fire pay?

1 Yes
2 No
DK/REF

In addition, some response categories in question QD10B were revised for 2014. In the 2013 survey, response category 4 for the Vietnam era indicated that this period was from August 1964 to April 1975, and category 5 indicated that the period between the Korean and Vietnam conflicts was from February 1955 to July 1964. The category in 2014 for the Vietnam era was revised to start in March 1961 (instead of August 1964), and the end date for the category for the period between the Korean and Vietnam conflicts was revised to February 1961. Despite this change, the edited variable for service during the Vietnam era retained the name ACTDVIET that had been created for 2013. However, the variable in 2014 for service between the Korean and Vietnam conflicts was given the name ACTDB4VN (where B4VN = before Vietnam) because the previous name for the variable in 2013 (ACTD5564) defined a period of military service (i.e., from 1955 to 1964) that was no longer applicable for 2014.[52]

The edited variables corresponding to questions QD10A through QD10C were assigned legitimate skip codes if respondents were aged 12 to 16 or respondents indicated in question QD09 (SERVICE) that they had never served in the U.S. armed forces. Similarly, the edited variables for QD10B (ACTD2001 through ACTD1941) and QD10C (COMBATPY) were assigned legitimate skip codes if question QD10A (edited variable ACTDEVER) was answered as "no."

As noted in Section 2.4.4, the response categories for the individual response categories in the "enter all that apply" question QD10B differed from those that were used for similar variables prior to 2013. Documentation for these values was as follows:

---

[52] The 2013 variable ACT5564 also was subsequently renamed to ACTDB4VN so that data for this variable from 2013 could be included on data files that also contained the corresponding 2014 variable.

- 1 = Yes, and
- 2 = No (not entered).

In addition, the editing procedures included identifying inconsistencies between respondents' ages and reports of being on active duty in a particular period based on respondents' birth dates and the particular period when they reported being on active duty. Variables for the specific period of active duty were assigned codes for bad data (Section 2.4.3) if respondents' answers would lead to the conclusion that the respondents would have been younger than 16 or older than 65 during the entire period when they reported serving on active duty. These edits were performed based on calculated upper and lower bounds for the possible ages when respondents could have been in the military during a given period of military service, based on their birthdates and the boundaries of the service periods in the questionnaire.

No editing was done if respondents would have been aged 16 to 65 at some point during the period for which they reported service. These edits resulted in data on periods of active military duty being set to bad data for a small number of respondents (fewer than 10) in 2014, including some respondents who reported serving on active duty in a period before they were born.

4.2.9 Creating the Edited Highest Grade Completed Variables (EDUC and EDEDUC)

EDUC and EDEDUC were created using the responses to the core education question QD11, which asked about the highest grade in school that the respondent had completed. The question from the survey instrument appears below:

QD11: What is the highest grade or year of school you have **completed**?

0	NEVER ATTENDED SCHOOL
1	1^{ST} GRADE COMPLETED
2	2^{ND} GRADE COMPLETED
3	3^{RD} GRADE COMPLETED
4	4^{TH} GRADE COMPLETED
5	5^{TH} GRADE COMPLETED
6	6^{TH} GRADE COMPLETED
7	7^{TH} GRADE COMPLETED
8	8^{TH} GRADE COMPLETED
9	9^{TH} GRADE COMPLETED
10	10^{TH} GRADE COMPLETED
11	11^{TH} GRADE COMPLETED
12	12^{TH} GRADE COMPLETED
13	COLLEGE OR UNIVERSITY / 1^{ST} YEAR COMPLETED
14	COLLEGE OR UNIVERSITY / 2^{ND} YEAR COMPLETED
15	COLLEGE OR UNIVERSITY / 3^{RD} YEAR COMPLETED
16	COLLEGE OR UNIVERSITY / 4^{TH} YEAR COMPLETED
17	COLLEGE OR UNIVERSITY / 5^{TH} OR HIGHER YEAR COMPLETED

No editing was performed to create EDUC based on other questionnaire information. In particular, EDUC was not edited with respect to education variables in the noncore demographics section (e.g., current grade), nor was it edited with respect to the respondent's age. Consequently, the core education variable would not be affected by changes that might occur in the content of noncore education variables in subsequent years. A second variable, EDTEDUC, was created as part of the noncore demographics processing (Section 5.2.3.3), but EDTEDUC was not used in any published estimates involving educational attainment.

The base variable for creating an imputation-revised version of education was called EDEDUC. It was equivalent to EDUC, except that missing values that described the type of item nonresponse (i.e., "don't know" or "refused") were set to the SAS missing code (.) so that they were properly handled by the imputation-related modeling programs.

4.2.10 Creating the Perceived Health Status Variable (HEALTH)

HEALTH was created from core question QD12, which asked the respondent to rate his or her health. The question from the survey instrument appears below:

QD12: This question is about your overall health. Would you say your health in general is excellent, very good, good, fair, or poor?

1 EXCELLENT
2 VERY GOOD
3 GOOD
4 FAIR
5 POOR

No processing of this variable was done beyond that of assigning the edited variable name HEALTH instead of QD12 (the variable name obtained from the CAI).

4.3 Imputation for Selected Core Demographics Variables

In this section, the imputation procedures applied to the marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level are described. These variables comprised five independent manifestations of the single response propensity (RP)/single prediction (PRD) type of PMN described in Chapter 3. Each imputation set is discussed in a separate section below.

4.3.1 Marital Status Variable (Imputation Set 1)

The first core demographic variable that underwent imputation was the marital status variable. The four substantive levels of the imputation-revised variable IRMARIT matched the four answer categories of QD07 (i.e., married, widowed, divorced or separated, or never married). Respondents aged 12 to 14 were automatically assigned an IRMARIT value of 99, a "legitimate skip" code. Since this is the first variable to undergo imputation in each cycle, there were no imputation-revised variables to use as auxiliary variables. This tended to make the imputation process simple and straightforward.

In marital status imputation procedures, only one RP model and only one PRD model was fit; most other NSDUH imputation procedures are run separately within three or four age groups. Single models were used across all age groups to ensure adequate sample size for response categories that would be rare within certain age groups (e.g., the "widowed" category for younger age groups). To account for the correlation between age and marital status, AGE was used in both the RP and PRD model steps and in a likeness constraint in the hot deck step.

The marital status variable has a very high response rate (see Appendix A). There are often fewer than 10 missing values in the entire sample in a given survey year.

4.3.1.1 Response Propensity Step

The response propensity model for imputation set 1 utilized the preliminary analysis weight, PANALWT. The marital status question QD07 was only asked of respondents aged 15 or older. Therefore, the domain contained unit respondents with AGE $\geq$ 15. The creation of the AGE variable is described in Section 4.2.2. Unit respondents in the domain with nonmissing EDMARIT values were considered item respondents. The EDMARIT variable is described in Section 4.2.6. See Table D.1 in Appendix D for details of the covariates used in the RP model for this variable.

4.3.1.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the marital status variable was modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The outcome variable had four levels, which mapped to the four answer categories of QD07. The four predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent selected each of the four answer categories of QD07.

4.3.1.3 Hot-Deck Step

The predicted means from the PRD step play a central role in the donor selection algorithm applied in the hot-deck step, but unlike the RP and PRD steps, the hot-deck steps for marital status were run separately within three age groups: 15 to 17, 18 to 25, and 26 or older. This was done to allow parallel processing, which decreases the time required for implementation. No logical constraints were used, and the only likeness constraint other than the delta constraint involved the continuous AGE variable. The few unit respondents requiring imputation for this variable are usually handled in the first attempt to find a donor, due to the mild set of constraints and large domain. The only imputation-revised variable created in the hot-deck step was IRMARIT.

4.3.2 Race Variables (Imputation Set 2)

As discussed in Section 4.2.7, race and Hispanicity were closely related. Therefore, race was used in the imputation of Hispanic/Latino origin and Hispanic/Latino group, and Hispanicity was used in the imputation of race. Since race underwent imputation first, imputation-revised versions of the Hispanic/Latino indicator and the Hispanic/Latino group were not available. This precluded their usage in race models. However, they were used extensively in constraints in the hot-deck step. The RP, PRD, and hot-deck steps were all run separately within three age groups: 12 to 17, 18 to 25, and 26 or older.

The race questions had low response rates relative to other questions in the NSDUH, due to high item nonresponse among Hispanic/Latino respondents. Nearly all of the race nonrespondents reported being of Hispanic/Latino origin (Table 4.1). The likeness constraints involving Hispanic/Latino group strongly influenced the final imputed values.

4.3.2.1 Response Propensity Step

The response propensity models for imputation set 2 utilized the preliminary analysis weight, PANALWT. The domain for the RP models included all unit respondents. Item respondents were those with EDRACE values from 1 to 15 and EDNWRACE values from 1 to 15. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

4.3.2.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the race variables were modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The outcome variable was the five-level variable EDRACEFORMODEL, which had the following levels:

1. White Only
2. Black/African American Only
3. American Indian/Alaska Native Only
4. Asian/Other Pacific Islander Only
5. Multiple Race

In survey years prior to 2008, multiple-race respondents were assigned to one of the first four categories above. An edited variable that did not include a category for more than one race was useful in the past because (1) the multiple race cell contained a small number of respondents, making imputation models difficult to fit, and (2) it was necessary to be used as a base variable for the final imputation-revised variable that did not include a category for more than one race (between 2003 and 2007, called IRRACE2). On the first point, the multiple race category has become less sparse over time (refer to Section 3.3 of the 2008 imputation report [Ault et al., 2010] for more details). On the second point, because multiple-race respondents were classified as a separate category starting in 2008, a decision was made to cease to create IRRACE2, where multiple-race respondents were assigned a single race as shown in the first four categories above. It was replaced in most cases with the variable RACE4. The variable RACE4 is described in Section 4.3.3.4.

EDRACEFORMODEL is a recode of the variable EDRACE, described in Section 4.2.7.3.2:

EDRACEFORMODEL =

- EDRACE, if $1 \leq$ EDRACE ≤ 4; else
- 5, if $5 \leq$ EDRACE ≤ 16; else
- missing.

The five predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent had each value of EDRACEFORMODEL.

4.3.2.3 Hot-Deck Step

Each item nonrespondent in the hot-deck step was assigned one of 11 missingness patterns. Ten of the missingness patterns, all rare, were set up to handle cases where something was known about the race categories such as "known to be Asian." The 11th missingness pattern, by far the most common, handled cases where nothing was known about the race categories. For a description of these missingness patterns, see Table E.5. Logical constraints applied to the cases where something was known about the race categories. Otherwise, for the cases where nothing was known, only likeness constraints were used. Sometimes, what was "known" about the race categories came from a random assignment for indirectly mapped codes, as described in Section 4.2.7.2.1.

Besides the segment and delta likeness constraints, the likeness constraints based on Hispanic/Latino group were important determinants of the final imputed value, because the vast majority of the item nonrespondents for race were Hispanic/Latino. In 2014, 17.7 percent of the overall respondent pool were of Hispanic/Latino origin, but 98.38 percent of the item nonrespondents for race were of Hispanic/Latino origin. Table 4.1 reports the distribution of Hispanic/Latino group among race item nonrespondents in 2014. Almost all are Hispanic/Latino and most (more than two thirds) of the Hispanic/Latino nonrespondents are Mexican only.

Table 4.1 Hispanic/Latino Status of Item Nonrespondents for Race

	Item Nonrespondents for Race	
Hispanic/Latino Status	**Number**	**Percentage**
Not Hispanic/Latino or Missing Hispanic/Latino Indicator	39	1.62
Hispanic/Latino	2,363	98.38
Mexican Only	1,658	69.03
Puerto Rican Only	250	10.41
Central/South American Only	203	8.45
Dominican Only	131	5.45
Other/Unknown	121	5.04
Total	2,402	100.00

Depending on the missingness pattern, the item nonrespondent received values from the selected donor for some subset of EDRACE, EDNWRACE, EDQD051-EDQD054, and collapsed versions of EDQD055-EDQD057 and EDQD058-EDQD0515. The collapsed versions of EDQD055-EDQD057 and EDQD058-EDQD0515 are simply the minimum of these variables. The first is an indicator of whether the respondent was Other Pacific Islander (including

Guamanian/Chamorro and Samoan), and the second is an indicator of whether the respondent was Asian. Most receive values for all variables. Item nonrespondents in missingness pattern 2 (known to be Asian but missing an Asian finer category) received values for only EDNWRACE, and item nonrespondents in missingness pattern 3 (known to be multiple race, but no other information) received values for everything except EDNWRACE. The imputation-revised versions of these variables are shown in Table 4.2.

Table 4.2 Edited Race Variables and Their Imputation-Revised Counterparts

Edited Race Variable	Imputation-Revised Race Variable
EDQD051	IRRACEWH
EDQD052	IRRACEBK
EDQD053	IRRACENA
EDQD054	IRRACENH
EDQD055-EDQD057 (collapsed)	IRRACEPI
EDQD058-EDQD0515 (collapsed)	IRRACEAS
EDRACE	IRDETAILEDRACE
EDNWRACE	IRNWRACE

IRDETAILEDRACE is not included in the final data files because the information it contains is covered by the other imputation-revised race variables. It is used in a likeness constraint for the Hispanic/Latino group variable discussed in Section 4.3.4.

Due to the strict constraints, the delta constraint had to be dropped sometimes. However, the likeness constraints related to Hispanic/Latino group were never dropped.

4.3.3 Hispanic/Latino Origin Variable (Imputation Set 3)

For the Hispanic/Latino origin indicator, the RP, PRD, and hot-deck steps were all run separately within three age groups: 12 to 17, 18 to 25, and 26 or older. Details on the procedures are given in the next four sections. The base variable for imputation, EDHOIND, is described in Section 4.2.7.4.1. The item response rate for this variable was much higher than for race.

4.3.3.1 Response Propensity Step

The response propensity models for imputation set 3 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all unit respondents. Item respondents were those with a nonmissing value for EDHOIND. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

4.3.3.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the Hispanic/Latino origin indicator was modeled using dichotomous logistic regression as implemented by the LOGISTIC procedure in SUDAAN.[53] The outcome variable was EDHOIND. The single predicted mean was the predicted probability that the respondent was of Hispanic/Latino origin.

4.3.3.3 Hot-Deck Step

The hot-deck step for the Hispanic/Latino origin indicator included a single predicted mean from the PRD step, no logical constraints, and only the segment and delta likeness constraints. EDHOIND is the base variable for imputation, and the imputation-revised version is called IRHOIND. Details on the hot-deck step, including the likeness constraints, are available in Tables E.6 and E.7 in Appendix E.

4.3.3.4 Recodes for Additional Race/Ethnicity Variables

The imputation-revised race (IRNWRACE) and imputation-revised Hispanic/Latino indicator (IRHOIND) variables were used to create several additional combined race/ethnicity variables. One of these (RACE4) was used in the subsequent processing of imputation-revised variables and had four levels: non-Hispanic/Latino white, non-Hispanic/Latino black/African American, Hispanic/Latino, and non-Hispanic/Latino other/multiple race. The NEWRACE1 and NEWRACE2 variables also were created from IRNWRACE and IRHOIND and were used extensively in the production of the 2014 detailed tables (Center for Behavioral Health Statistics and Quality, 2015a).

4.3.4 Hispanic/Latino Group Variable (Imputation Set 4)

The edited variable EDHOGRP, described in Section 4.2.7.4.3, categorized Hispanic/Latino respondents into Hispanic/Latino groups. These categories were directly mapped to the same categories in the imputation-revised variable, IRHOGRP4, which had eight possible values: Puerto Rican, Mexican, Cuban, Central or South American, Dominican, Spanish, Other Hispanic/Latino, and not Hispanic/Latino. The closely-related imputation-revised variable IRHOGRPM was also created to identify respondents who selected more than one Hispanic/Latino group; recall that a priority rule is used to assign a single group to multiple-group respondents in the creation of EDHOGRP (and therefore IRHOGRP4).

Imputations were not conducted separately within age groups, as was the case for marital status. The Hispanic/Latino group variables were created only for respondents of Hispanic/Latino origin as defined by IRHOIND. This results in a small domain. The models were likely to be better when age groups were combined because (1) none of the response categories were sparsely populated and (2) sufficiently large donor pools were ensured in the hot-deck step.

4.3.4.1 Response Propensity Step

The response propensity models for imputation set 4 utilized the preliminary analysis weight, PANALWT. The domain indicator included all respondents of Hispanic/Latino origin as defined by IRHOIND. Item respondents were those with a nonmissing value for EDHOGRP who

[53] In SAS-callable SUDAAN, this is the RLOGIST procedure to avoid confusion with SAS's own LOGISTIC procedure.

selected only a single Hispanic/Latino group. The multiple-group respondents whose EDHOGRP was assigned by the priority rule (Scenario 3, described in Section 4.2.7.4.3) were not used to fit the PRD model in the next step. See Table D.1 in Appendix D for details of the covariates used in the RP models for these variables.

4.3.4.2 Prediction Step

Because the model would have been much more difficult to fit if all seven levels were used, the EDHOGRP variable was collapsed into a four-level categorical variable (EDHOGRP2). Table 4.3 shows the mapping of EDHOGRP levels to EDHOGRP2 levels. Using the adjusted weights that are outputs of the RP step, EDHOGRP2 was then modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The four predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent had each value of EDHOGRP2.

Table 4.3 Mapping of EDHOGRP Levels to EDHOGRP2 Levels

EDHOGRP (Base Variable)	EDHOGRP2 (Modeled Variable)
Mexican	Mexican
Puerto Rican	Puerto Rican
Central or South American	Other Hispanic/Latino
Cuban	Cuban
Dominican	Other Hispanic/Latino
Spanish	Other Hispanic/Latino
Other Hispanic/Latino	Other Hispanic/Latino
Not Hispanic	Not Hispanic

4.3.4.3 Hot-Deck Step

The hot-deck step for the Hispanic/Latino group variables was straightforward. Besides the segment and delta likeness constraints, the most notable feature was a likeness constraint involving race. A five-level race variable was used as a covariate in the RP and PRD models with the following levels: White Only, Black/African American Only, American Indian/Alaska Native Only, Asian Only, and Multiple Race. To further exploit the relationship between race and Hispanic/Latino group, a likeness constraint required the donor's IRDETAILEDRACE variable to match a subset of the racial categories mentioned by the recipient. The constraint did not apply if the recipient was an item nonrespondent for race.

IRHOGRP4 was the imputation-revised version of EDHOGRP. The other imputation-revised variable IRHOGRPM was set equal to 8 (more than one Hispanic/Latino group) if either the respondent reported membership in more than one group, or the donor for a particular item nonrespondent reported membership in more than one group. Otherwise, IRHOGRPM was set equal to IRHOGRP4.

The Hispanic/Latino group variables generally have low imputation rates. The number of cases with missing data is usually fewer than 100 in each survey year.

4.3.4.4 Recodes for Additional Analyses

Among the recoded variables that were created from IRHOGRP4, the variable HISPGRP2 was used in subsequent processing and was created by collapsing the levels of IRHOGRP4 into four levels: Puerto Rican, Mexican, Other Hispanic/Latino (includes Cuban, Central or South American, Dominican, Spanish, and Other Hispanic/Latino), and not Hispanic/Latino.

4.3.5 Education Level Variable (Imputation Set 5)

The imputation-revised education level variable was similar to the imputation-revised Hispanic/Latino group variable in that it was categorical with numerous levels, and as with the Hispanic/Latino group, the response variable for the PRD model was collapsed into fewer levels for ease of modeling. There were generally very few missing cases for this variable—for some years, fewer than 10—so the application of the method tended to be straightforward. Two age groups were used for RP and PRD modeling: 12 to 17 and 18 or older. However, the hot-deck step was implemented separately for three age groups: 12 to 17, 18 to 25, and 26 or older.

4.3.5.1 Response Propensity Step

The response propensity models for imputation set 5 utilized the preliminary analysis weight, PANALWT. The domain indicator for each of the two RP models included all unit respondents. Item respondents were those with a nonmissing value for EDEDUC. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

4.3.5.2 Prediction Step

EDEDUC was collapsed into fewer levels for modeling. The response variables were different for the two PRD models: the response variable for the 12-17 age group had five levels, and the one for the 18-or-older age group had four. The mapping of EDEDUC to the response variable RESPEDUC is shown in Table 4.4.

Table 4.4 Mapping of EDEDUC Levels to RESPEDUC Levels

<table>
<tr><th rowspan="2">EDEDUC (Base Variable)</th><th colspan="2">RESPEDUC (Modeled Variable)</th></tr>
<tr><th>12-17</th><th>18+</th></tr>
<tr><td>Never attended school</td><td rowspan="6">Less than elementary school</td><td rowspan="12">Less than high school</td></tr>
<tr><td>1st grade completed</td></tr>
<tr><td>2nd grade completed</td></tr>
<tr><td>3rd grade completed</td></tr>
<tr><td>4th grade completed</td></tr>
<tr><td>5th grade completed</td></tr>
<tr><td>6th grade completed</td><td rowspan="2">Elementary school</td></tr>
<tr><td>7th grade completed</td></tr>
<tr><td>8th grade completed</td><td rowspan="2">Middle school</td></tr>
<tr><td>9th grade completed</td></tr>
<tr><td>10th grade completed</td><td rowspan="2">Some high school</td></tr>
<tr><td>11th grade completed</td></tr>
<tr><td>12th grade completed</td><td>High school</td><td>High school</td></tr>
</table>

Table 4.4 Mapping of EDEDUC Levels to RESPEDUC Levels (continued)

EDEDUC (Base Variable)	RESPEDUC (Modeled Variable)	
	12-17	**18+**
College or university/1st year completed		Some college
College or university/2nd year completed		
College or university/3rd year completed		
College or university/4th year completed		College or higher
College or university/5th or higher year completed		

Using the adjusted weights that are outputs of the RP step, both PRD models were fit using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The predicted means matched the levels of the response variable, so there were five predicted means for the 12-17 hot-deck step and four for the 18-25 and 26-or-older hot-deck steps.

4.3.5.3 Hot-Deck Step

The hot-deck step for the education level variable was straightforward and implemented separately for three age groups: 12 to 17, 18 to 25, and 26 or older. The only base variable was EDEDUC, and the imputation-revised version was called IREDUC. Both variables are based on the detailed 18-level variable, as compared with the simplified RESPEDUC variable used in the RP and PRD steps. No logical constraints were required. In addition to the segment and delta likeness constraints, the third likeness constraint required the donor to be the same age as the recipient. This was an especially important constraint for the 12-17 age group, because the age covariate often had to be dropped from the PRD model due to near-empty cells when the variables were cross-tabulated, causing instability in the estimates.

4.3.5.4 Recode for Additional Education Variable

EDUCCAT2, a recoded education variable, was created using the imputation-revised highest grade completed variable (IREDUC). EDUCCAT2 had five levels (less than high school and aged 18 or older, high school graduate and aged 18 or older, some college and aged 18 or older, college graduate and aged 18 or older, or 12 to 17 years old). If a respondent reported completing the 12th grade, then he or she was presumed to be a high school graduate. Similarly, if a respondent reported completing 4 or more years of college, then he or she was presumed to be a college graduate. EDUCCAT2 was often used as a covariate in later imputation models.

This page intentionally left blank

5. Editing and Imputation for the NSDUH Noncore Demographics Variables

5.1 Introduction

This chapter discusses procedures for editing and imputing the demographic variables from the "noncore" section of the interview. As noted in Chapter 1, demographic variables in noncore modules could be subject to change. The core demographics variables that were edited and imputed in the 2014 National Survey on Drug Use and Health (NSDUH) were discussed in Chapter 4.

This chapter discusses editing procedures for the following noncore demographics sections:

- moves in the past 12 months and state residency in the past 2 years,
- immigrant status,
- noncore education (i.e., education-related questions other than the highest grade attained),
- employment and workplace, and
- field interviewer (FI) debriefing questions (completed by the FI after the conclusion of the interview).

This chapter also discusses imputation procedures for the variables pertaining to immigrant status and current employment status. Other edited variables for noncore demographics that are discussed in this chapter were not imputed.

5.2 Editing the Noncore Demographics Variables

This section documents the editing procedures for the noncore demographics sections that were listed in Section 5.1. As noted in Section 1.1, the noncore demographics sections of the interview were administered by the interviewers using computer-assisted personal interviewing (CAPI).

5.2.1 Moves in the Past 12 Months and State Residency in the Past 2 Years

This section covers issues related to changes of residence. Question QD13 (edited variable MOVEPYRR in 2014) asked respondents to report the number of times that they had moved in the past 12 months. If a respondent moved at least once in the past 12 months, the respondent was asked to report the state where he or she was living a year prior to his or her interview date (question QD13A; edited variable LIVSTYAR in 2014). If a respondent reported in question SEN04 (for adults aged 18 older) or YE04 (for youths aged 12 to 17) that he or she moved at least once in the past 5 years, the respondent was asked to report the state where he or she was living 2 years prior to the interview date (question QD13B; edited variable LIVST2YR in 2014). Respondents who moved from another state to their current one (edited variable STATELOC; see Section 4.2.3) within the past 2 years were asked to report the month and year when they moved to their current state (question QD13C; edited variables MOVSTMOR and MOVSTYRR in 2014).

The skip logic for question QD13 changed for the 2013 NSDUH. Prior to the 2013 survey year, all respondents were asked QD13. Starting with the 2013 survey, respondents were asked QD13 only if they reported at least one move in the past 5 years in question YE04 in the Youth Experiences module (for adolescents aged 12 to 17) or in question SEN04 in the Social Environment module (for adults aged 18 or older). Thus, QD13 was skipped in the 2013 NSDUH if respondents reported zero moves in YE04 or SEN04 or if they answered these questions as "don't know" or "refused."

Consequently, the unweighted response distribution for QD13 in the 2013 NSDUH data changed relative to the distribution in 2012. Documentation of the effect of this change is included in the editing and imputation report in the 2013 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2015b).

This change for QD13 also affected the distributions for question QD13A (state residence 1 year ago), QD13B (state residence 2 years ago), and QD13C (month and year when respondents moved to their current state). Respondents who were not asked question QD13 because they answered YE04 or SEN04 as "don't know" or "refused" also had values of blank for QD13A, QD13B, and QD13C because their eligibility for being asked these follow-up questions was unknown.

In addition, the unweighted distributions changed for the "editing indicator" variable EIMOVPYR (which indicated when values were logically assigned through editing) and the flag variable MOVYRFLG (which indicated when the number of reported moves in the past year was greater than the number of moves that was previously reported for the past 5 years). For MOVYRFLG, the change to the skip logic did reduce the occurrence of inconsistencies between data in QD13 and in YE04 or SEN04.

Because of these changes in the distributions between the 2012 and 2013 surveys, the decision was made to change the names of the edited variables for moves in the past 12 months and state residency for the 2013 NSDUH. Table 5.1 compares the names for these edited variables in the 2012 and 2013 surveys. The names for 2013 continued to be used in 2014.

Table 5.1 Comparison of Names for Edited Variables for Number of Moves in the Past 12 Months and State Residency in the 2012 and 2013 NSDUH

Question	2012 Name	2013 Name	Explanation
QD13	MOVESPYR	MOVEPYRR	Number of moves in the past 12 months.
	EIMOVPYR	EIMOVYRR	Editing indicator for logical assignment of values for number of moves in the past year.
	MOVYRFLG	MOVYRFGR	Flag indicator when number of moves in the past year is greater than the previous number in the past 5 years.

Table 5.1 Comparison of Names for Edited Variables for Number of Moves in the Past 12 Months and State Residency in the 2012 and 2013 NSDUH (continued)

Question	2012 Name	2013 Name	Explanation
QD13A	LIVSTYRA	LIVSTYAR	State where respondent lived 1 year ago.
QD13B	LIVST2YA	LIVST2YR	State where respondent lived 2 years ago.
QD13C	MOVSTTMO	MOVSTMOR	Month when respondent moved to the current state.
QD13C	MOVSTTYR	MOVSTYRR	Year when respondent moved to the current state.

As noted in Section 4.2.3, the state that interviewers reported for the location of the dwelling unit sometimes did not match the state residence information that was used to sample a given case. If STATELOC was set to bad data because of incorrect information for the current state residency, MOVSTYRR and MOVSTMOR also were set to bad data; respondents were asked question QD13C only if they previously reported in QD13A or QD13B that they had moved from a different state. Therefore, when STATELOC was set to bad data, this called into question whether respondents should have been asked to report the month and year when they moved to their current state.

As noted previously, information from question QD13C was captured in two edited variables pertaining to the month (MOVSTMOR) and year (MOVSTYRR) when respondents moved to their current state. Assignment of values to these two variables took into account the logic that the computer-assisted interviewing (CAI) program used to assign specific months and years to the response categories in QD13C, which depended on the interview date and the answers in questions QD13A and QD13B. Specifically, if QD13A had been answered (i.e., for the state where the respondent was living a year ago) and the state where the respondent lived a year ago did not equal the respondent's current state residence, then the respondent saw the response categories that were filled in QD13C as follows: the first response option (QD13C = 1) was filled with the month and year that occurred 12 months ago (i.e., same month as the interview month but in the year 2013), the next response option was filled with the month that occurred 11 months ago, and so on, up through the last response option (QD13C = 13), which was the current month in which the respondent was interviewed.

Suppose, for example (not necessarily actual data), that a respondent was interviewed in May 2014, QD13A was answered, and QD13A was different from the state where the respondent currently was living. The first response option in QD13C that the respondent saw would have been "May 2013," corresponding to the same month as the interview month but in the prior calendar year. If a value of 4 was keyed in QD13C, that response corresponded to "August 2013" based on this interview month. In this situation, the edited variable MOVSTYRR was coded as 2013 and MOVSTMOR was coded as 8.

Otherwise, the month that the CAI logic filled in the response options for QD13C still began with the interview month. However, the year that was filled in the response options began with 2012. Response option 13 in QD13C was filled with the interview month in 2013. Suppose, for example, that QD13A was blank because the respondent did not move in the past 12 months (QD13 = 0) or the respondent did not move in the past 5 years (SEN04 = 0 or YE04 = 0), but the state that the respondent lived in 2 years ago (from QD13B) did not equal the state where the sample dwelling unit was located. If the interview was conducted in April 2014, then the first response option in QD13C would be filled with April 2012. If a value of 2 was keyed in QD13C, that response would correspond to the respondent moving to the current state in May 2012. In this example, MOVSTYRR was coded as 2012 and MOVSTMOR was coded as 5.

Key aspects of processing the variables MOVEPYRR, LIVSTYAR, LIVST2YR, MOVSTMOR, and MOVSTYRR for the 2014 NSDUH involved logical inference of zero moves in MOVEPYRR and assignment of legitimate skip codes based on the CAI logic (Section 2.4.2). These edits occurred as follows:

- If respondents did not move in the past 5 years (i.e., SNMOV5YR = 0, corresponding to SEN04 from the social environment module for respondents aged 18 or older, or YEMOV5YR = 0, corresponding to YE04 from the youth experiences module for respondents aged 12 to 17), then blanks in QD13 were replaced with codes of 0. The variable EIMOVYRR was changed from the default value of 1 (questionnaire data) to a value of 2 (logically assigned data).
- If respondents did not move in the past 12 months (i.e., MOVEPYRR = 0, including situations in which zero moves was logically inferred), then LIVSTYRR was assigned a legitimate skip code. If respondents moved zero times or moved within the same state, then MOVSTYRR and MOVSTMOR were assigned legitimate skip codes. This condition would not hold if respondents reported in either QD13A or QD13B that they moved to their current state from outside of the United States.

In addition, the number of moves in the past 12 months (from MOVEPYRR) sometimes was greater than the number of moves in the past 5 years (from SNMOV5YR or YEMOV5YR). No editing was done to MOVEPYRR, SNMOV5YR, or YEMOV5YR in this situation. However, a "flag" variable (MOVYRFGR) was created that indicated when this inconsistency occurred. The default value was 98 (i.e., blank) when MOVEPYRR was consistent with either SNMOV5YR (for adults) or YEMOV5YR (for adolescents). MOVYRFGR had a value of 1 when MOVEPYRR was inconsistent with SNMOV5YR or YEMOV5YR.

5.2.2 Immigrant Status

Edits described in this section for the immigrant status variables have applied since 2004, when the content of the immigrant status variables changed. Question QD14 asked whether respondents were born in the United States. If they were not born in the United States, question QD15 asked respondents for their country of birth. The former question QD16 (length of time that respondents had lived in the United States, corresponding to edited variable LIVEDUSA) was replaced in 2004 with three variables pertaining to the following:

- whether respondents had lived in the United States for at least 1 year (question QD16A or edited variable LIVUS1YR);
- the number of years that respondents had lived in the United States, if they reported in QD16A that they had lived in the United States for at least 1 year (question QD16B or edited variable LIVUSYRS); and

- the number of months that respondents lived in the United States, if they answered QD16A as "no," indicating that they had not lived in the United States for at least 1 year (question QD16C or edited variable LIVUSMOS).

An important aspect of processing the immigrant status variables involved assigning legitimate skip codes where relevant (Section 2.4.2). For example, if respondents reported that they were born in the United States (i.e., the edited variable BORNINUS was answered as "yes"), the edited variables BORNINOT, LIVUS1YR, LIVUSYRS, and LIVUSMOS were assigned legitimate skip codes. Similarly, if LIVUS1YR = 1 (i.e., "yes"), LIVUSMOS was assigned a legitimate skip code. If LIVUS1YR = 2 (i.e., "no"), LIVUSYRS was assigned a legitimate skip code. When LIVUS1YR was coded as 94 ("don't know") or 97 ("refused"), the appropriate code for "don't know" or "refused" was assigned to the variables LIVUSYRS and LIVUSMOS that had been skipped.

If respondents reported that they were born outside the United States, however, it was possible for them to specify an answer in question QD15 that logically would mean that they were born in the United States. If this inconsistency occurred in the data (i.e., it had not been resolved by the interviewer), then the edited variable BORNINUS was logically inferred to be answered as "yes."[54] The edit procedures also logically inferred that the edited variables BORNINOT, LIVUS1YR, LIVUSYRS, and LIVUSMOS should have been skipped.

It also was possible for respondents to report in question QD16B that they had lived in the United States for a number of years greater than their current age. When this situation occurred, the edited variable LIVUSYRS was assigned a bad data code to indicate that the answer was inconsistent with the respondent's age. No editing was done to LIVUSYRS when LIVUSYRS was equal to the respondent's age because that answer in LIVUSYRS could be interpreted to mean that the respondent came to the United States as a baby.

Because an imputation-revised variable was desired for the age of entry to the United States for respondents who were born elsewhere, the final step in editing these variables was to use LIVUS1YR, LIVUSYRS, and LIVUSMOS to compute the age at which a respondent entered the United States. The age of entry to the United States was computed from continuous forms of the length of time that the respondent was living in the United States and the respondent's age. Because QD16B and QD16C were mutually exclusive, the edited variables LIVUSMOS and LIVUSYRS were used to create a continuous variable (LNGTHLIV) that indicated how many years a respondent had lived in the United States. In most cases, LNGTHLIV had the same value as LIVUSYRS. However, if the respondent had lived in the United States for less than 1 year, his or her LNGTHLIV values were obtained from LIVUSMOS by converting the number of months into fractions of 1 year. LNGTHLIV was set to missing when LIVUSYRS and LIVUSMOS had missing data codes. A continuous age variable, CONTAGE, was defined as CONTAGE = (interview date – birth date + 1) ÷ 365.25. Because the interview date and birth date, as described in Chapter 4, had no missing values, CONTAGE also had no missing values. For respondents who were born in the United States, a legitimate skip code of 999 was assigned to both the LNGTHLIV and CONTAGE variables.

[54] If respondents reported being born in Alaska or Hawaii and were born before 1959 (i.e., before Alaska and Hawaii became states), these respondents still were considered to have been born in the United States.

The variable ENTRYAG2 was the base variable for creating the imputation-revised variable IRENTAG2 and represented the (continuous) age at which an immigrant entered the United States. ENTRYAG2 was defined as ENTRYAG2 = CONTAGE – LNGTHLIV and was set to missing if LNGTHLIV was missing. ENTRYAG2 also had a legitimate skip code (999) for respondents who were born in the United States.

5.2.3 Noncore Education

The noncore education module assessed whether respondents were currently enrolled in school and, if not, whether respondents who were aged 12 to 25 had dropped out of school without having received a high school diploma. Figure 5.1 shows the logic for asking follow-up questions according to whether or not respondents were enrolled in school. Respondents who reported that they were enrolled were asked to report their current grade in school (or the grade they would be in once they returned from school break), whether they were a full-time or part-time student, and if they were full-time students, the number of days that they missed school in the past 30 days because they were sick or because they skipped school (questions QD18 through QD21). Respondents who were aged 25 or younger, had completed the 12th grade or lower (from question QD11), and were not enrolled in school were asked whether they had received a high school diploma (question QD22). Respondents in this age group who reported that they left school without receiving a high school diploma were asked to report whether they had received a general educational development (GED) certificate of high school completion, why they left school before receiving a high school diploma, and their age when they left school (questions QD23 through QD25). Table 5.2 shows the edited variables that corresponded to the questions in the noncore education module.

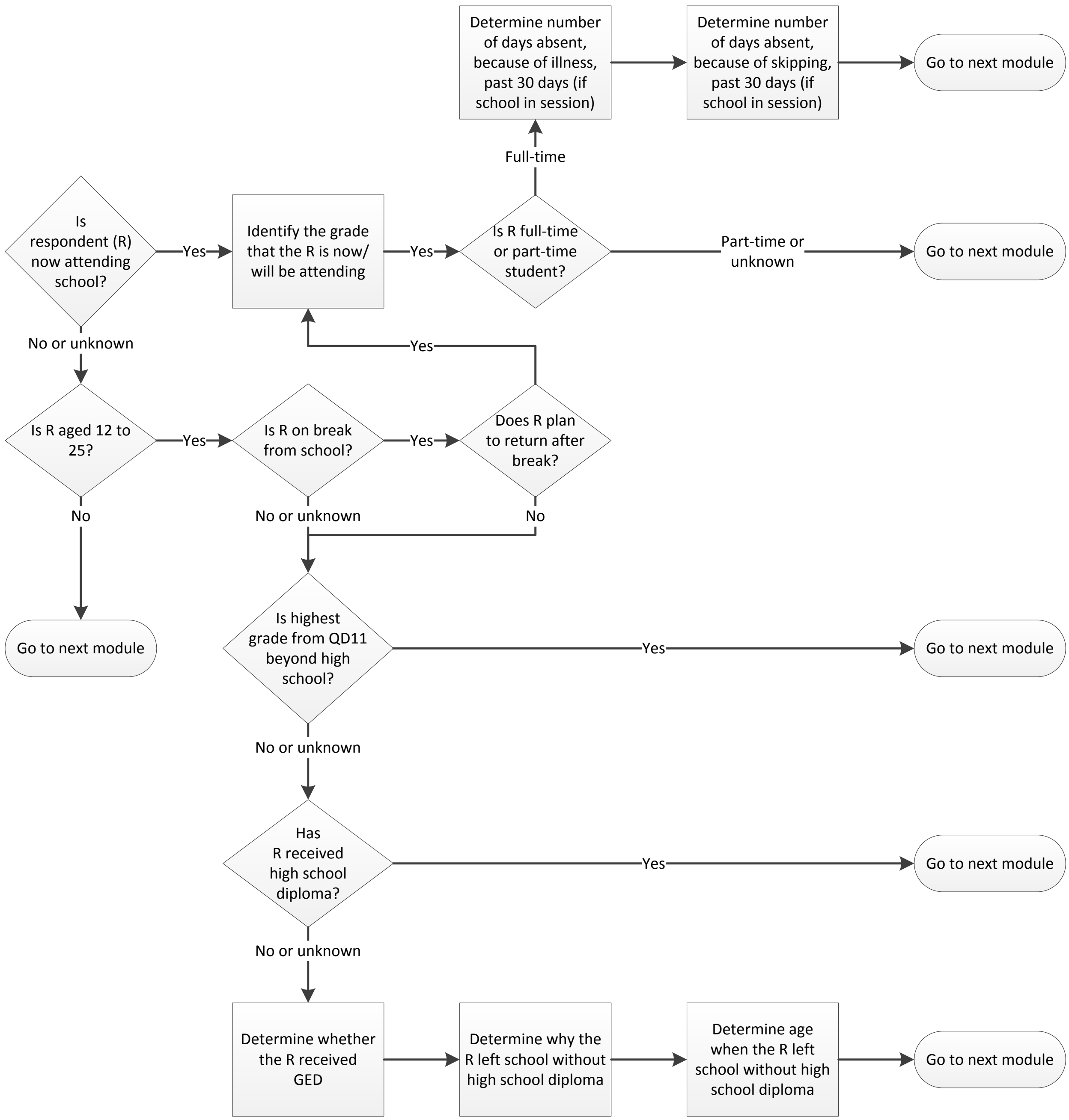
Consistent with this module being in a noncore section of the interview, the content of questions in this module has changed over time. However, these questions have not changed since the start of the new NSDUH baseline in 2002. Documentation of historical changes to the noncore education questions prior to the new baseline in 2002 can be found in the report on editing of interviewer-administered data that was prepared for the 2011 NSDUH methodological resource book (Kroutil, Chien, Handley, & Bradshaw, 2013).

The remainder of this section discusses the following issues that were relevant to editing of the noncore education variables:

- overall routing logic for the noncore education module (Section 5.2.3.1);
- edits involving respondents' school enrollment status (Section 5.2.3.2);
- general edits involving the last and current grade for respondents who were in school (Section 5.2.3.3);
- specific edits involving the last and current grade for respondents aged 12 to 18 who were in school (Section 5.2.3.4);
- specific edits involving the last and current grade for respondents aged 19 or older who were in school (Section 5.2.3.5); and
- miscellaneous edits involving the noncore education variables (Section 5.2.3.6).

Table B.1 in Appendix B describes edits for the noncore education variables pertaining to current school enrollment, last and current grade, and receipt of a high school diploma or GED. Table B.1 also is intended as a companion to the discussion of the editing procedures for the noncore education variables in Sections 5.2.3.2 through 5.2.3.5.

Figure 5.1 Follow-Up Question Logic for Respondents in the Noncore Education Module



GED = general educational development.

Table 5.2 Mapping of Noncore Education Questions to Edited Variables

Question	Edited Variable	Question Text
QD17	SCHENRL	The next questions are about school. Are you now attending or are you currently enrolled in school? By "school," we mean an elementary school, a junior high or middle school, a high school, or a college or university. Please include home schooling as well.
QD17A	SCHENRL	Are you currently on a holiday or vacation break from school?
QD17B	SCHENRL	Do you plan to return to school when your holiday or vacation is over?
QD18	EDUCATND	What grade or school [are you now attending/will you be attending when your vacation is over]? Please tell me the number from the card.
QD19	SDNTFTPT	[Are you/Will you be] a full-time or a part-time student?
QD20	SCHDSICK	During the past 30 days, that is, from [DATEFILL] up to and including today, how may whole days of school did you miss because you were sick or injured?
QD21	SCHDSKIP	During the past 30 days, that is, from [DATEFILL] up to and including today, how may whole days of school did you miss because you skipped or "cut" or just didn't want to be there?
QD22	HSDIPLMA	Have you received a high school diploma?
QD23	HSGED	Have you received a GED certificate of high school completion?
QD24	LFSCHWHY	Please look at this card and tell me which one of these reasons best describes why you left school before receiving a high school diploma.
QD24SP	LFSCHWHY	What is the main reason why you left school before receiving a high school diploma?
QD25	LFTSCHAG	How old were you when you stopped attending school?

NOTE: Questions QD18 through QD21 applied if respondents reported in QD17 through QD17B that they were currently in school or planned to return to school when their break was over. Questions QD22 through QD25 applied to respondents aged 12 to 25 who reported that they were not currently in school or were on break but did not plan to return to school when their break was over. See Figure 5.1 for additional routing within the noncore education module.

5.2.3.1 Routing Logic for Noncore Education

As shown in Figure 5.1, determining whether respondents were currently attending school was critical for determining the subsequent questions that respondents would be asked about their education. However, respondents could answer the question about their current school attendance (QD17) as "no" if they were on a holiday break or vacation from school. Therefore, respondents who did not report in question QD17 that they were currently attending school were asked follow-up questions (if they were aged 12 to 25 and their highest reported grade from question QD11 was grade 1 to 15) to determine if they were on a holiday or vacation break from school (question QD17B), and if so, whether they intended to return to school once their break was over (question QD17B).

If respondents reported in QD17 that they were not currently attending school (QD17 = 2) but they were on break from school (QD17A = 1) and they intended to return to school once their vacation or break was over (QD17B = 1), the edited school enrollment variable SCHENRL (based on data from QD17, QD17A, and QD17B) was set to a value of 1 ("yes") to indicate that the respondents should be considered enrolled. Otherwise, the response from QD17 was carried over to SCHENRL. That included situations in which respondents reported in QD17A that they were not on vacation break from school, or who reported in QD17B that they were on break but did not intend to return to school once their break was over.

If respondents were currently attending school or were on break but intended to return, the wording of subsequent questions varied according to their current status. For question QD18, respondents who reported in QD17 that they were currently attending school (QD17 = 1) were asked to report their current grade in school (i.e., at the time of the interview). For respondents who were on vacation break from school but intended to return to school once their break was over (QD17B = 1), question QD18 asked for the grade that they would be in once they returned from their vacation break. Similarly, for the question about full-time or part-time student status (question QD19) respondents who were currently attending school were asked, "Are you a full-time student or a part-time student?" Those who were on break from school but intended to return to school were asked, "Will you be a full-time or a part-time student?"

Based on the logic shown in Figure 5.1, a key aspect of editing the noncore education variables involved assigning legitimate skip codes (Section 2.4.2) based on respondents' current enrollment status and age. If respondents were currently enrolled in school, the edited variables corresponding to questions QD22 through QD25 (HSDIPLMA, HSGED, LFSCHWHY, and LFTSCHAG) were assigned legitimate skip codes. Similarly, respondents aged 26 or older were considered to have legitimately skipped out of questions QD22 through QD25 because of the age requirement for administration of these questions, regardless of whether they might not have finished high school. In addition, if respondents were not currently enrolled in school, the edited variables corresponding to questions QD18 through QD21 (EDUCATND, STUDNT, SCHDSICK, and SCHDSKIP) were assigned legitimate skip codes, provided there were no other data to suggest that they were enrolled (Section 5.2.3.2).

5.2.3.2 Edits Involving Current School Enrollment

Because the routing logic in the noncore education module was based on respondents' current enrollment status, an important aspect of editing the variables in this module involved editing of the enrollment variable SCHENRL. Table B.1 in Appendix B describes edits for SCHENRL.

For example, respondents aged 12 to 25 who reported that they were not enrolled in school and that they had not received a high school diploma were asked to report why they left school before receiving a diploma (Figure 5.1). If respondents reported leaving school for some other reason besides the ones that were listed in question QD24, they were asked to specify the main reason why they left school. If the "OTHER, Specify" response (Section 2.3.1.4) indicated that the respondent was still in school or was being home schooled, SCHENRL was assigned special codes to indicate that the respondent was in school (Table B.1). Except for the data on reasons for leaving school that were responsible for the logical inference that the respondent was in school (edited variable LFSCHWHY), the variables HSDIPLMA and HSGED were overwritten with values of 89, and LFTSCHAGE was overwritten with a value of 989 (LEGITIMATE SKIP Logically assigned; see Section 2.4.2) because it was logically inferred that these questions should have been skipped. Also, because questions QD18 through QD21 had been skipped based on the respondent not having reported current school enrollment, the edited variables corresponding to these questions retained missing values (i.e., codes of blank).

In addition, a "hard error" is included in the education section for situations in which the highest grade from QD11 was higher than the current (or anticipated) grade from QD18. FIs were prompted to verify the answers with the respondents and correct any information in QD11 or QD18. If the answers were correct as recorded, the FIs could "suppress" the hard error and continue with the interview.[55] When FIs suppressed a hard error message, however, they were requested to enter a comment documenting why the information that had been entered in QD11 and QD18 was correct. Although most of these comments were relevant to editing involving the last and current grade (Section 5.2.3.3), some comments were relevant to editing the school enrollment variable.

Specifically, if the FI's comments indicated that the respondent was now in some sort of technical or vocational school, the school enrollment variable SCHENRL was set to a value of 4 (No LOGICALLY ASSIGNED). This edit was done because interviewers were instructed not to include vocational or technical schools as types of schools in which respondents could be enrolled. When SCHENRL was set to a value of 4, any data in EDUCATND, STUDNT, SCHDSICK, and SCHDSKIP were overwritten with values of 89 (LEGITIMATE SKIP Logically assigned). Where possible, when respondents were inferred not to be enrolled in school because their current enrollment was in a technical or vocational school, FI comments also were used to edit the variables pertaining to receipt of a high school diploma (HSDIPLMA) or receipt of a GED certificate of high school completion (HSGED). These edits are discussed further in Section 5.2.3.6.

5.2.3.3 General Edits Involving the Last and Current Grade

The current school grade from question QD18 could be inconsistent with the highest grade that the respondent reported completing in question QD11. In most situations, one might expect the current grade in QD18 to be one grade level higher than the response in QD11. In addition, no editing was done when the current grade reported in QD18 was the same as the highest grade reported in QD11 because respondents could have been repeating a grade.

As noted in Section 4.2.9, the core education variable EDUC (highest grade completed) was not edited with respect to data on a respondents' current grade because QD18 was in a noncore section of the interview. However, a second variable, EDTEDUC, was created as part of the editing of the noncore education module. Consequently, the core education variable would not be affected by changes that might occur in the content of noncore education variables over time. Nevertheless, the EDTEDUC variable might in some situations be a more accurate reflection of the highest grade that respondents had completed.

As noted previously, FIs could enter comments if they suppressed a hard error message when the highest grade from QD11 was higher than the current (or anticipated) grade from

[55] Unlike the consistency checks that were described in Section 1.2 and elsewhere in this report (particularly Chapter 6), "hard errors" in the CAI instrument typically required interviewers or respondents to resolve an inconsistency before the interview could proceed to the next available question; consistency checks in self-administered sections of the interview gave respondents an opportunity to resolve an inconsistency but did not require resolution before the interview could proceed. As noted in Section 5.2.3.2, however, interviewers could allow an inconsistency to remain between the last and current grade but were required to provide a reason for doing so.

QD18. These comments were reviewed on a case-by-case basis to determine which of the following held:

- the answers should be accepted and no editing should be done to EDTEDUC (corresponding to QD11) or EDUCATND (corresponding to QD18);
- the value for EDTEDUC or EDUCATND should be edited for consistency with the comments entered by the FI;
- EDTEDUC or EDUCATND should be set to bad data based on the FI comments; or
- predefined editing rules for education should be invoked (see below and Table B.1).

Any edits based on the FI comments were done on a case-level basis using the respondent ID rather than on an automated basis using predefined editing rules. These case-level edits superseded any of the usual edits discussed in Table B.1 that otherwise would have been done. However, this hard error was suppressed for fewer than 30 cases, and specific case-level edits were done for fewer than 15 of these cases.

The general education edits discussed in the remainder of this section were invoked if the hard error between QD11 and QD18 had been triggered, the answers from QD11 and QD18 had not been corrected, or the FI's comments indicated that a correction needed to be made, but what needed to be corrected was not clear from the FI's comments. However, answers to QD11 and QD18 were accepted when FIs provided a plausible reason for the discrepancy between the two answers, such as if respondents were in college and transferred to another school but some prior credits did not transfer.

The following potential patterns of inconsistent or questionable data could occur between QD18 and QD11 despite the presence of the hard error check between the two questions:

- the hard error was triggered, but the case was allowed to proceed through the general education edits for the reasons described above;
- the hard error was not triggered, the current grade in QD18 was exactly two grades higher than the highest grade completed (from QD11), but the respondent was at a current grade level that would be expected for someone at his or her age (e.g., if a 12 year old reported last completing the 4th grade and reported currently being in the 6th grade); or
- the hard error was not triggered, and the current grade in QD18 was more than two grade levels higher than the highest grade from QD11.

An algorithm was developed to handle these types of situations when they occurred. This algorithm is discussed in detail below. In particular, having accurate data on respondents' current grade levels is important for comparing NSDUH data with drug use data from in-school surveys, such as Monitoring the Future, that are administered to students in specific grades.

For respondents aged 12 to 23, a series of arrays was set up that mapped out the highest grade and current grade that would be *expected* relative to a respondent's current age, assuming an orderly progression from one grade level to the next highest level. Table 5.3 shows a matrix that maps the current age with expected grades in the United States.

Table 5.3 Mapping of Current Age with Expected Grades

Current Age	12	13	14	15	16	17	18	19	20	21	22	23
Expected Completed Grade	6	7	8	9	10	11	12	13	14	15	16	17
Expected Current Grade	7	8	9	10	11	12	13	14	15	16	17	17

For example, one might expect most people in the United States to have completed the 6th grade by the time they are 12. It would therefore not be unreasonable for someone to be aged 12 and to be currently in the 7th grade, depending on when the respondent was interviewed. An upper age limit was set at 23 because a grade level of 17 (college or university, 5th year or higher) was the upper limit of the education levels.

In addition, the algorithm allowed for some deviation relative to the expected ages. For example, if a respondent was aged 12, had completed the 5th grade, and was currently in the 6th grade, this would be an acceptable pattern because the respondent might have had his or her 12th birthday at some point during the 6th grade.

Separate edits were done depending on whether a respondent was aged 12 to 18 or was older than 18. The rationale for doing edits separately for these two different age groups was that the typical progression from one grade level to the next would be less likely to hold for adults and at higher educational levels. Suppose, for example, that a respondent completed 3 years of college but changed majors and not all of the prior credits applied to the new major. It would be possible for the respondent to report having completed 3 years of college and to be currently enrolled at a level lower than the third year of college, depending on how the respondent interpreted these questions. Similarly, a respondent who got a bachelor's degree in one field and went back to school for a second bachelor's degree might report having completed 4 years of college but also might report currently being enrolled at some level below the 4th year of college.

5.2.3.4 Edits Involving the Last and Current Grade for Respondents Aged 12 to 18

For respondents aged 12 to 18, the highest grade completed and the current grade were considered to be consistent with the respondent's age if what was reported was within 1 year of the grades shown in Table 5.3. Thus, if a respondent was aged 12, the algorithm considered completion of any grades from the 5th through the 7th to be sufficiently consistent with the respondent's age. Similarly, for respondents aged 12, the algorithm considered current enrollment in any grades from the 6th through the 8th to be sufficiently consistent with the respondent's age.

Therefore, the following four data combinations were possible:

- both the completed grade and the current grade were consistent with the respondent's age;

- the highest completed grade was consistent with the respondent's age, but the current grade was not;
- the current grade was consistent with the respondent's age, but the highest completed grade was not; or
- neither the highest completed grade nor the current grade was consistent with the respondent's age.

Separate edits were done according to the four combinations of data patterns described immediately above. The following edits were done if both the completed grade and current grade appeared to be consistent with the respondent's age:

- If the current grade was more than two grade levels higher than the highest completed grade, the current grade was edited to be consistent with the highest grade because the latter was a core variable. For example, if a respondent was aged 17, reported completing the 10th grade, and reported currently being in the first year in college (QD18 = 13), the edits logically inferred that the respondent currently was in the 11th grade. The edited variable for current grade (EDUCATND) was assigned a code of 31 (i.e., 11th grade LOGICALLY ASSIGNED).
- If the current grade was lower than the highest completed grade, the edit code gave precedence to the reported grade that would yield the most consistent result relative to the respondent's age. In particular, if accepting the report of the highest grade and inferring that the respondent was currently in the next highest grade would yield a current grade that was inconsistent with the respondent's age, then the noncore-created variable EDTEDUC (i.e., edited highest grade completed) was assigned a value consistent with the current grade. Suppose, for example, that a 12-year-old respondent reported currently being in the 6th grade but completed the 7th grade. Accepting the answer that the respondent was currently in the 6th grade and had completed the 5th grade would be more consistent with the respondent's current age than would be the converse (i.e., accepting that this respondent had completed the 7th grade and inferring that he or she was currently in the 8th grade). In this example, EDTEDUC would be assigned a code of 25 (i.e., 5th grade LOGICALLY ASSIGNED).

If the highest completed grade was consistent with the respondent's age but the current grade was not, the highest completed grade was accepted by default. This was done if the current grade was lower than the highest completed grade or the current grade was more than two grade levels higher than the highest completed grade. The edited current grade EDUCATND was therefore assigned a value to indicate a current grade level that was 1 year higher than the highest completed grade. For example, if the respondent reported completing the 10th grade, EDUCATND would be assigned a code of 31 (i.e., 11th grade LOGICALLY ASSIGNED).

If the current grade was consistent with the respondent's age but the highest completed grade was not, the edit procedures accepted the current grade by default. Thus, if a 12-year-old respondent reported last completing the 4th grade and reported currently being in the 6th grade, this edit would identify the current 6th grade as being consistent with an age of 12; completing the 4th grade would not be identified as consistent with an age of 12. In this example, EDTEDUC would be assigned a code of 25 (i.e., 5th grade LOGICALLY ASSIGNED).

If neither the current grade nor the reported highest grade was consistent with the respondent's age, the following was done:

- If the current grade was lower than the highest grade that was reported, the algorithm picked the answer that was closest to the expected grade, based on the matrix shown in Table 5.3. The variable with the more inconsistent data was assigned a bad data code. This edit allowed for situations where respondents may have fallen behind where they would be expected to be grade-wise (e.g., if they had been held back a year).
- If the current grade was exactly two grade levels higher than the reported highest completed grade and the highest completed grade was higher than what would be expected for the respondent's age, no further editing was done. Otherwise, the created noncore variable EDTEDUC was assigned a bad data code. This edit was designed to allow for situations where a respondent might be on an accelerated track.
- If the current grade was more than two grade levels higher than the reported highest completed grade and it was lower than the expected current grade, then the value was retained for the current grade. The variable EDTEDUC was assigned a bad data code. In other situations, both EDTEDUC and EDUCATND (i.e., the edited current grade) were assigned codes of bad data. The rationale for the first edit was that, if EDUCATND was lower than the expected current grade, EDUCATND would be more consistent with the *expected* current grade and the respondent's age than what the reported highest completed grade would be.

5.2.3.5 Edits Involving the Last and Current Grade for Respondents Aged 19 or Older

Minimal editing of EDTEDUC and EDUCATND was done for respondents aged 19 or older. Other than the edits described in this section, no other editing of the educational level data was done for respondents aged 19 or older.

If the current grade was lower than the highest completed grade and the current grade was at the 12th grade or lower, then EDUCATND (i.e., the current grade) was assigned a bad data code. Otherwise, no further editing was done when the current grade was lower than the highest grade. For example, if a respondent reported completing the 12th grade but the answer for the current grade indicated that the respondent was in the 1st grade (QD18 = 1), then the response in question QD18 would probably indicate a typographical error. The first edit described in this paragraph would assign a bad data code to EDUCATND.

If the current grade was more than two grade levels higher than the highest completed grade and the current grade was above the 12th grade, the edits compared what the highest grade completed would be relative to the current grade and examined if the highest completed grade was actually increased by 10 years. If increasing the highest completed grade by 10 years yielded a completed grade that was still less than or equal to the reported current grade, then the variable

EDTEDUC was assigned a code of bad data. In this situation, the interpretation was that a typographical error was made for the highest grade. Otherwise, no further editing was done. The first edit described in this paragraph was based on observed patterns that suggested that keying errors may have been made in QD11 (highest grade completed). For example, there were respondents who reported completing the 1st grade (QD11 = 1) and currently being in their 13th or higher year of school. Again, this pattern suggested that the second digit did not get keyed in QD11. This edit ensured that the respondent was classified as being enrolled in a grade above the high school level.

5.2.3.6 Miscellaneous Edits Involving the Noncore Education Variables

Table B.1 discusses additional editing of variables in the noncore education module other than those pertaining to current enrollment and grade level. These included situations in which the data were consistent with respondents' status as being enrolled in school or not enrolled but other potential inconsistent responses were observed, such as responses for receipt of a high school diploma or GED (if not enrolled) or absences from school in the past 30 days (if currently enrolled).

For example, if a respondent reported that he or she was not enrolled in school, reported having received a high school diploma, but also reported in QD11 that he or she had completed only the 9th grade or lower, the respondent was logically inferred not to have received a diploma. HSDIPLMA was assigned a code of 4 (No LOGICALLY ASSIGNED).

In addition, where possible, when respondents were inferred not to be enrolled in school because of FI comments indicating that their current enrollment was in a technical or vocational school, FI comments also were used to edit the variables pertaining to receipt of a high school diploma (HSDIPLMA) or receipt of a GED certificate of high school completion (HSGED). For example, if the FI comments indicated that respondents had received a high school diploma, HSDIPLMA could be assigned a code of 3 (Yes LOGICALLY ASSIGNED), and the remaining variables HSGED, LFSCHWHY, and LFTSCHAG could be assigned legitimate skip codes. In the absence of information in the FI comments that would permit editing of additional variables, HSDIPLMA, HSGED, LFSCHWHY, and LFTSCHAG were left as blank because these respondents who were logically inferred not to be enrolled were skipped out of questions that were relevant to respondents who were not enrolled.

5.2.4 Employment and Workplace

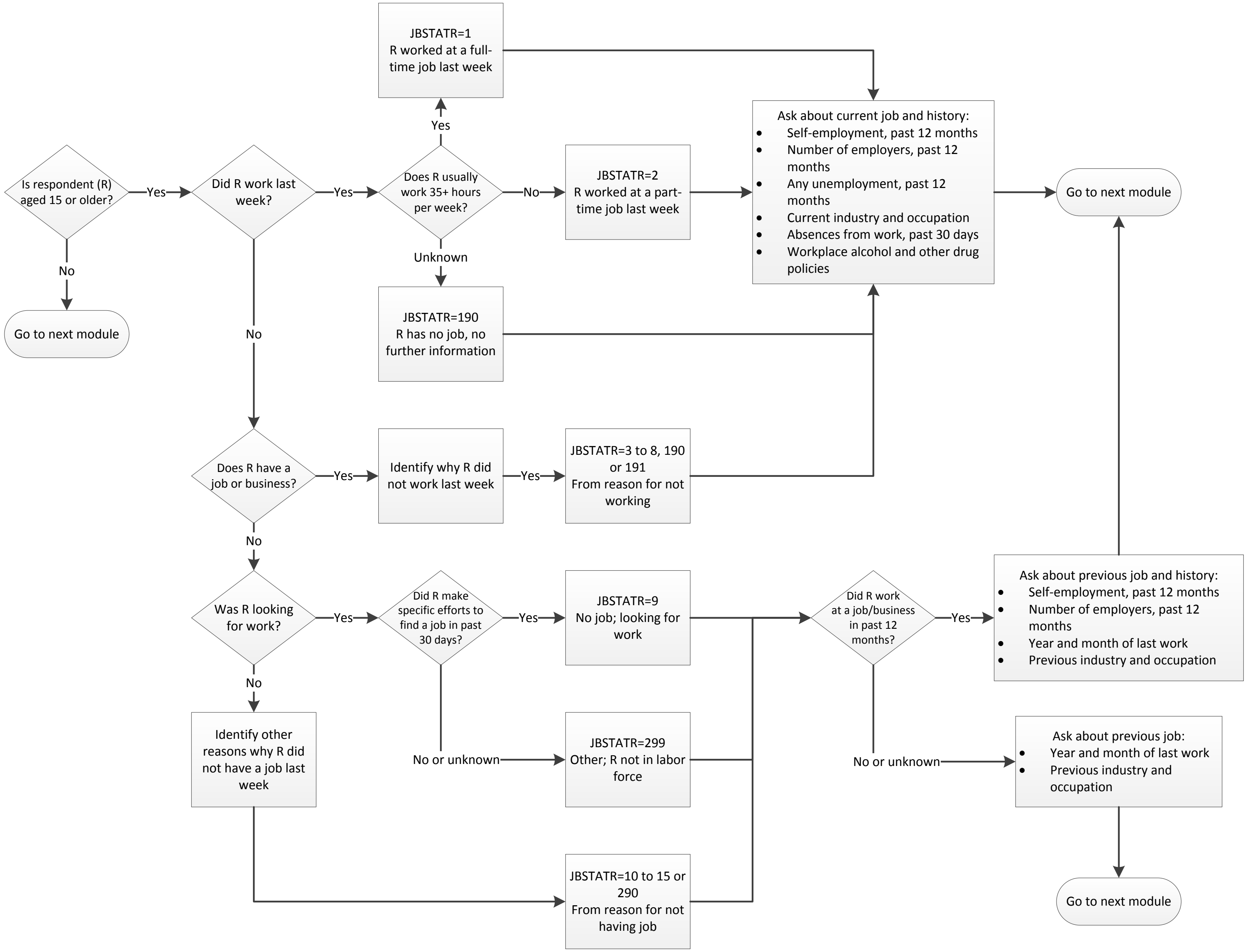
Respondents aged 15 or older were asked questions about their current employment, employment history, and characteristics of their workplace (if applicable). Question QD26 asked whether respondents worked in the week prior to the interview. If respondents reported that they did not work in the past week, they were asked in question QD27 whether they had a job or business. Respondents then were routed through different branches of work-related questions depending on how they answered these two key questions, as shown in Figure 5.2. For example, respondents who worked in the past week were asked questions to determine full-time or part-time work status (e.g., whether they usually worked 35 or more hours per week), whether they ever had a period of unemployment in the past 12 months, the number of days they missed work in the past 30 days because they were sick or because they did not want to be at their workplace, and characteristics of their workplace, particularly with respect to alcohol and other drug policies at their workplace. Similarly, respondents who did not work in the past week and did not have a job were routed into questions relevant for people who currently were not working, such as why they did not have a job, whether they made specific efforts to find work in the past 30 days, and the month and year when they last worked for pay.

The employment and workplace questions and logic underwent important changes prior to the start of the new NSDUH baseline in 2002. Documentation of historical changes to the employment and workplace questions prior to the new baseline in 2002 can be found in the report on editing of interviewer-administered data that was prepared for the 2011 NSDUH methodological resource book (Kroutil et al., 2013).

In addition, coding procedures changed in 2003 for the industry in which respondents worked (currently or in the past year) and for respondents' occupations (for their current or previous job). Since 2003, the Census Bureau has coded these responses. In addition, the industry and occupation (I&O) codes and classification procedures since 2003 have been based on categories from the 2000 census. Therefore, the names of the I&O variables were changed, beginning with the 2003 NSDUH. The names of the I&O variables in 2002 were as follows: WRKINDUS (industry in which the respondent is currently working), WRKOCCUP (respondent's current occupation), WRKINDYR (industry in which the respondent formerly worked in the past year), and WRKOCCYR (respondent's former occupation in the past year). Since 2003, these variables have been called WRKIDSTY, WRKOCUPT, WRKIDSYR, and WRKOCUYR, respectively. Recoded versions of these variables (WRKIDST2, WRKOCUP2, WRKIDSY2, and WRKOCUY2) were created for the NSDUH public use file. Despite these changes, the procedures for editing these I&O variables have not changed.

Also, since 2005, the question about the respondent's month of last employment (question QD39B) has been asked only for those respondents who reported in question QD39A (edited variable WRKLSTYR) that they last worked in the current year or the prior year (i.e., 2013 or 2014). Previously, all respondents who gave a valid year in question QD39A were asked to report the month they last worked in QD39B. Consequently, more respondents in 2005 legitimately skipped out of question QD39B compared with respondents in prior years. For this reason, the edited variable corresponding to question QD39A has been called WRKLSTMN since 2005. In 2004, this variable was called WRKLSTMO. If the year in WRKLSTYR was more than 1 year prior to the current survey year, WRKLSTMN was assigned a legitimate skip code. Because of the changes to the skip logic in 2005, the distribution of WRKLSTMN also was not comparable with the distribution of WRKLSTMO prior to 2005.

Figure 5.2 Routing Logic for Respondents in the Employment Status Module



An important aspect of editing the work-related variables involved identification of situations where questions had been legitimately skipped (Section 2.4.2). A second key aspect of processing the work-related variables was to use the data to establish respondents' current work status (Figure 5.2). As noted previously, a single, recoded work status variable named JBSTATR was created that served as the starting point for creation of a simplified edited variable (EDEMPY) and a final, statistically imputed employment status variable (EMPSTATY). JBSTATR was created from the following final variables: WRKEDWK (whether the respondent worked in the past week), WRKHAVJB (whether the respondent had a job if he or she did not work in the past week), WRKHRSUS (whether the respondent usually worked 35 or more hours per week), WRKNOWRK (reason for not working in the past week despite having a job), WRKNOJOB (reason for not having a job in the past week), WRKEFFRT (made specific efforts to find work in the past 30 days), and WRKEDYR (whether the respondent had a job in the past 12 months). Based on the data in these variables, respondents aged 15 or older were assigned to one of the categories in JBSTATR that are listed in Table 5.4.

Table 5.4 Categories of JBSTATR

Code	Employment Situation	Code	Employment Situation
1	Worked at full-time job, past week	**11**	No job: keeping house full time
2	Worked at part-time job, past week	**12**	No job: in school/training
3	Has job but out: vacation/sick/temp absence	**13**	No job: retired
4	Has job but out: layoff, looking for work	**14**	No job: disabled for work
5	Has job but out: layoff, not looking for work	**15**	No job: didn't want a job
6	Has job but out: waiting to report to new job	**190**	Has full-time job, reason for not working unknown
7	Has job but out: self-employed, no business past week	**191**	Has part-time job, reason for not working unknown
8	Has job but out: in school/training	**199**	Has job, no further information
9	No job: looking for work	**290**	No job, no further information
10	No job: layoff, not looking for work	**299**	Other, not in labor force

NOTE: Additional codes for JBSTATR in the 900 series have their standard meanings in NSDUH: Don't know (994), Refused (997), Blank (998), and Legitimate skip (999).

In addition, respondents who reported in question QD31 that they did not have a job but were looking for work were not classified as being unemployed unless they reported in WRKEFFRT that they had made specific efforts in the past 30 days to find work (such as making contacts with someone about a job, sending out resumes or job applications, or placing or answering ads). If respondents reported that they did not have a job but were looking for work but WRKEFFRT was not answered as "yes," they were classified as not in the labor force (code 299) in JBSTATR.

If respondents did not know or refused to report whether they worked in the past week, WRKEDYR was checked for indications of whether respondents worked in the past year. Respondents who indicated in WRKEDYR that they did not work in the past 12 months were classified as not having a job (JBSTATR = 290). Otherwise, if respondents did not provide information on whether they worked in the past week (i.e., QD26 answered as "don't know" or "refused"), JBSTATR was assigned the corresponding code of "don't know" or "refused."

Table B.2 in Appendix B describes edits for employment variables pertaining to whether respondents had a job in the past week, the number of days absent from work in the past 30 days, the number of weeks without a job in the past 12 months, the number of hours worked in the past week, the year and month that respondents last worked for pay, and self-employment. As noted previously, for example, the question pertaining to the month that respondents last worked for pay was changed in 2005. If respondents reported in question QD39A that they never worked for pay, interviewers were instructed to enter a response of 9991. When the month question QD39B had been skipped because a response of 9991 had been entered in QD39A, the edited month variable WRKLSTMN was assigned a code of 91. Documentation for 9991 (or 91) was as follows: 9991 = NEVER WORKED AT A JOB OR BUSINESS.

A refinement to the editing procedures for the employment and workplace section also has been implemented since 2003 for respondents who reported that they did not work in the past week (WRKEDWK = 2). The variable pertaining to the number of hours that respondents worked in the past week (WRKHRSWK) was assigned a legitimate skip code regardless of how respondents answered the question about having a job (QD27). Prior to 2003, a legitimate skip code was assigned to WRKHRSWK only if respondents reported that they did not work in the past week (WRKEDWK = 2) and QD27 was answered as "no"; the prior logic did not assign legitimate skip codes to WRKHRSWK if respondents answered QD27 as "don't know" or "refused." Logically, however, if respondents reported that they did not work in the past week, they would not have worked any hours at a job during that period, regardless of how they answered question QD27.

A final procedure that is discussed for the editing of the employment variables (not presented in Table B.2) concerns the creation of the base variable EDEMPY, which was used to create the imputation-revised employment status variable. EDEMPY was derived from JBSTATR and WRKHRSUS. WRKHRSUS was used in some cases to determine whether employed respondents were employed full time or part time. Specifically, EDEMPY was derived in the following manner:

EDEMPY =

- 99, if the respondent is 12 to 14 years old; else
- 1 (full time), if JBSTATR = 1 or 190, or if JBSTATR = 3, 6, 7, 8, or 199 and WRKHRSUS = 1; else
- 2 (part time), if JBSTATR = 2 or 191, or if JBSTATR = 3, 6, 7, 8, or 199 and WRKHRSUS = 2; else
- 3 (unemployed), if JBSTATR = 4, 5, 9, or 10; else
- 4 (other), if JBSTATR = 11, 12, 13, 14, 15, 290, or 299; else

- 5 (part time or full time), if JBSTATR = 3, 6, 7, 8, or 199 and WRKHRSUS was missing (i.e., greater than 2); else
- missing.

5.2.5 Field Interviewer Debriefing Questions

The FI debriefing section was to be completed by the interviewer to obtain information about the potential quality of the interview. This included information about factors that might have affected the quality of the data, such as the degree of privacy in the interview setting. These questions were not to be read aloud to the respondent.

Only minimal processing was done to the data in this section. Specifically, unedited variables were replaced with final, mnemonic variable names (e.g., PRIVACY for the variable pertaining to the interviewer's indication of how private the interview was). Where relevant, variables also were assigned legitimate skip codes based on the routing logic in this section.

5.3 Imputation for Noncore Demographics Variables

For the noncore demographics module of the 2014 NSDUH, three imputation-revised variables were created from the base variables EDEMPY, BORNINUS, and ENTRYAG2: the first was an employment status variable, EMPSTATY;[56] the second, IRBORNUS, was an indicator of whether the respondent was born in the United States; and the third, IRENTAG2, recorded the age at which the respondent entered the United States. These three variables were processed in three separate, single-member imputation sets, using the single response propensity (RP)/single prediction (PRD) type of the predictive mean neighborhood (PMN) method described in Chapter 3. The ultimate goal in imputing values for missing data in these variables was to create a data file containing variables that would indicate whether respondents could be included in past year incidence analyses based on when they entered the United States. All three variables tend to have item response rates of more than 99 percent. See Table A.25 in Appendix A for details on the rates of missingness among these three noncore demographics variables.

5.3.1 Employment Status (Imputation Set 1)

The first noncore demographic variable that underwent imputation was the employment status variable. The imputation process was straightforward except for one feature: the handling of cases with EDEMPY = 5, where it was known that the respondents were employed but it was not known whether they were employed part time versus full time. These cases were handled in the hot-deck step in a separate missingness pattern, with a single logical constraint and a modified predictive mean vector. The final imputation-revised variable EMPSTATY had five levels: employed full time, employed part time, unemployed, other, and a skip code for respondents aged 12 to 14. Two age groups were used for RP and PRD modeling: 15 to 25 and 26 or older. The hot-deck step was implemented separately for three age groups: 15 to 17, 18 to 25, and 26 or older.

[56] Unlike other imputation-revised variables, EMPSTATY was not preceded by an "IR" prefix. However, it was accompanied by imputation indicators that did have the requisite "II" prefix: II2EMSTY and IIEMPSTY.

5.3.1.1 Response Propensity Step

The response propensity models for imputation set 1 utilized the preliminary analysis weight, PANALWT. For the first RP model, the domain included all unit respondents aged 15 to 25. For the second RP model, the domain included all respondents aged 26 or older. In both cases, item respondents were those with EDEMPY values of 1, 2, 3, or 4. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

5.3.1.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the employment status variable was modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. For both age groups, the outcome variable had four levels, which mapped to the first four levels of EDEMPY.

5.3.1.3 Hot-Deck Step

The predicted means from the prediction step play a central role in the donor selection algorithm applied in the hot-deck step, but unlike the RP and PRD steps, the hot-deck steps were run separately within three age groups: 15 to 17, 18 to 25, and 26 or older. This was done to allow for parallel processing, which decreases the time required for implementation. Each item nonrespondent in the hot-deck step was assigned one of two missingness patterns. Item nonrespondents with a missing value for EDEMPY were handled in the first missingness pattern, which used the full predictive mean vector and no logical constraints. Item nonrespondents with EDEMPY = 5 were handled in the second missingness pattern, which applied a logical constraint to ensure that the donor was employed (either full time or part time). Also, conditional probabilities were used to take advantage of the partial information that was available. Instead of using the model's predicted probabilities directly, a single predicted mean was derived using a conditional probability, which was the probability that the respondent was employed full time, given that the respondent was employed. In addition to the segment and delta likeness constraints, a third likeness constraint (i.e., donor's age must be within 5 years of recipient's age) was applied in the hot-deck step. See Appendix E for more details on missingness patterns and constraints for employment status.

5.3.1.4 Recodes for Additional Analyses

EMPSTAT4 was a direct recode of EMPSTATY and AGE. For respondents who were younger than 15 or older than 17, EMPSTAT4 and EMPSTATY were equivalent. For 15- to 17-year-olds, responses for EMPSTATY were overwritten with a code indicating that the respondent was too young to have his or her employment status recorded for the variable. This was the same code that was used for 12- to 14-year-olds for EMPSTATY (and EMPSTAT4).

5.3.2 Immigrant Status, Born-in-U.S. Indicator (Imputation Set 2)

The second noncore demographic variable that underwent imputation was the born-in-U.S. variable, BORNINUS. This was a dichotomous variable with very few missing responses. The RP, PRD, and hot-deck steps were all run separately within three age groups: 12 to 17, 18 to

25, and 26 or older. The imputation procedure was straightforward and is described in the next three sections.

5.3.2.1 Response Propensity Step

The response propensity models for imputation set 2 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all unit respondents. Item respondents were those with a nonmissing value for BORNINUS. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

5.3.2.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the born-in-U.S. indicator was modeled using dichotomous logistic regression as implemented by the LOGISTIC procedure in SUDAAN.[57] The outcome variable was BORNINUS. The single predicted mean was the predicted probability that the respondent was born in the United States.

5.3.2.3 Hot-Deck Step

The hot-deck step for the born-in-U.S. indicator included a single predicted mean from the prediction step, no logical constraints, and only the segment and delta likeness constraints. BORNINUS was the base variable for imputation and the imputation-revised version was called IRBORNUS. Details on the hot-deck step are available in Appendix E.

5.3.3 Immigrant Status, Age of Entry (Imputation Set 3)

The age of entry variable was created only for respondents who were not born in the United States as defined by IRBORNUS. This results in a small domain. As a result, imputations were not conducted separately within age groups. The models were likely to be improved when age groups were combined because (1) none of the response categories were sparsely populated, leading to more robust models and (2) sufficiently large donor pools were ensured in the hot-deck step. Details on the procedures applied to the age of entry variable are explained in the next three sections.

5.3.3.1 Response Propensity Step

The response propensity model for imputation set 3 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all respondents who were not born in the United States as defined by IRBORNUS. Item respondents were those domain members with a nonmissing value for ENTRYAG2. See Table D.1 in Appendix D for details of the covariates used in the RP model for this variable.

[57] In SAS-callable SUDAAN, this is the RLOGIST procedure to avoid confusion with SAS's own LOGISTIC procedure.

5.3.3.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the predicted mean for an immigrant's age of entry was estimated using a linear regression model, as implemented by the REGRESS procedure in SUDAAN. To control the upper and lower bounds of predicted means for age of entry, it was necessary to perform a logit transformation on the response variable. The response variable in the model was the immigrant age at entry as a proportion of the continuous version of current age CONTAGE, as described in Section 5.2.2. The expression of the proportion is $P_i = Y_i/N_i$, where $Y_i = Age\ at\ Entry_i$ and $N_i = Continuous\ Age_i$ (CONTAGE).

After the weight adjustment, the following empirical logit transformation was used as the response variable in a weighted linear univariate regression:

$$\log\left[\left(Y_i + 0.5\right)/\left(N_i - Y_i + 0.5\right)\right].$$

This transformation was nearly equivalent to the standard logit transformation,

$$Y_i^* = \log\left[P_i/\left(1 - P_i\right)\right],$$

which was not used because this transformation is unstable for respondents who entered the country at their current age (such that $P_i = 1$).

5.3.3.3 Hot-Deck Step

Two logical constraints were utilized in the hot-deck step for the age of entry variable. Both involved the respondent's age. One required that the donor's age of entry be less than or equal to the recipient's current age. The other required that the difference between the recipient's current age and the donor's age of entry be less than 1 year if the recipient lived in the United States for less than 1 year (as indicated by QD16A) or greater than 1 year if the recipient lived in the United States for more than 1 year. The only base variable was ENTRYAG2 and its imputation-revised counterpart was IRENTAG2. The segment and delta likeness constraints were applied in the hot-deck step. Details on the hot-deck step are available in Tables E.17 through E.19 in Appendix E.

This page intentionally left blank

6. Editing and Imputation for the NSDUH Core Drug Use Variables

6.1 Introduction

This chapter provides documentation of procedures for the editing and statistical imputation (subsequently referred to as "imputation") of core drug variables in the 2014 National Survey on Drug Use and Health (NSDUH).[58] As a prerequisite for reviewing this chapter, readers are encouraged to familiarize themselves with the content and terminology in Chapters 1 through 3 of this report because these three chapters provide an overview to the NSDUH interview and the general approach to editing and imputation of NSDUH data, with information on specific processes that are common across the survey in Chapters 2 and 3.

Consistent with prior years, the core drug use measures collected in the 2014 NSDUH included lifetime use (or nonuse); initiation of use (i.e., age at first use [AFU], and for recent initiates, the month of first use [MFU] and year of first use [YFU]); most recent use; frequency of use in the past 12 months (for alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, and prescription psychotherapeutic drugs); and frequency of use in the past 30 days (for tobacco products except for pipe tobacco, alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, and inhalants). Figure 6.1 provides an overview of the general routing logic for the core drug sections of the interview.[59] The implications of this general routing logic for editing and imputation are discussed in connection with the procedures that are described in the remaining sections of this chapter.

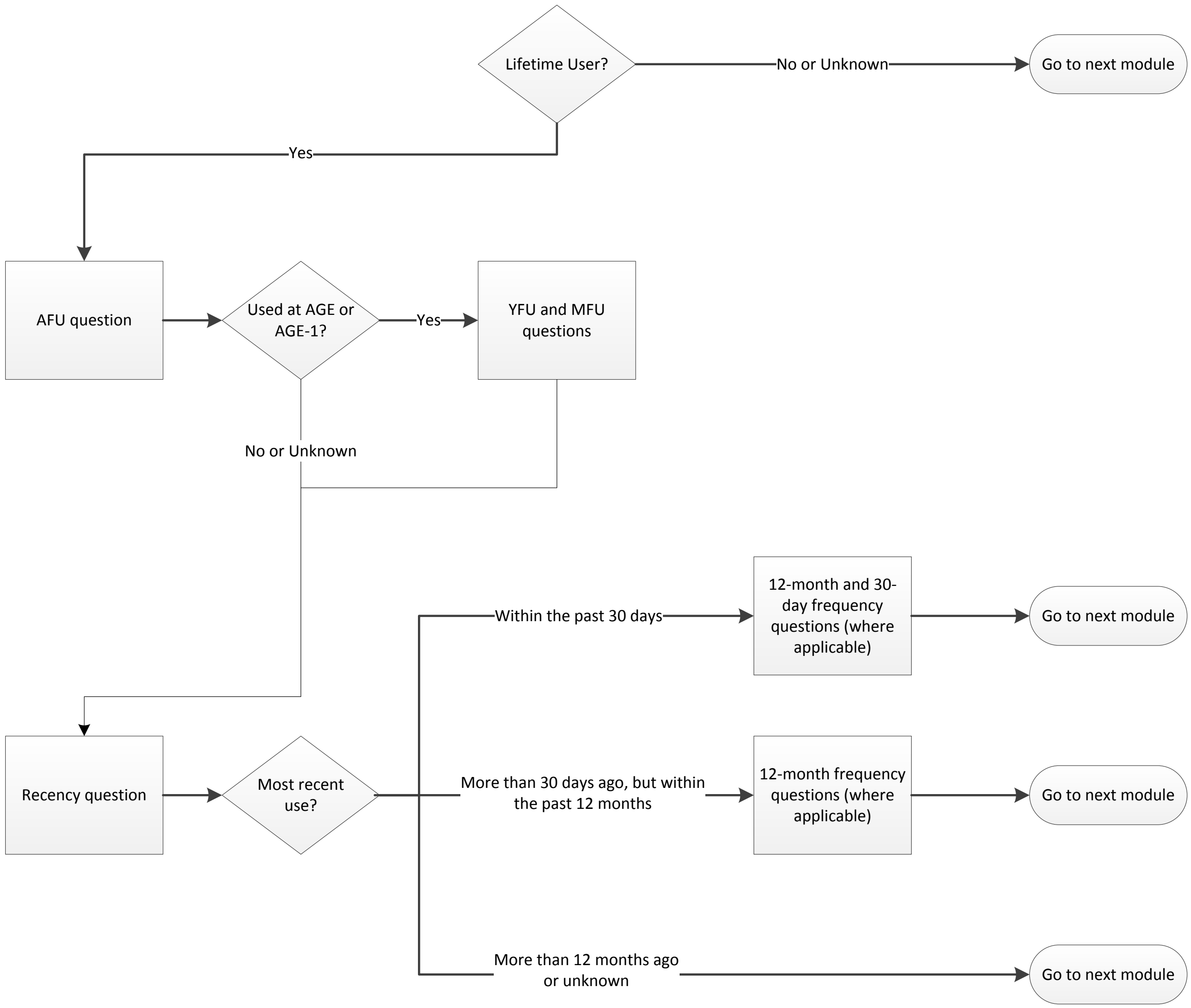
Variables corresponding to questions for these measures first underwent editing to (1) replace missing values with codes that indicated that the questions did not apply, (2) replace missing values with nonmissing values that could logically be inferred from other data, (3) make logical inferences when answers to related questions were inconsistent, and (4) identify inconsistent or ambiguous responses to be resolved through imputation. Edited variables for these measures then underwent imputation to replace missing values with nonmissing values or to replace ambiguous responses (e.g., use at some point in a respondent's lifetime but no definite period for most recent use) with specific ones.

---

[58] A discussion of "core" and "noncore" sections of the NSDUH interview is provided in Chapter 1. Section 6.1.3 discusses an exception to editing and imputation involving core drug variables.

[59] For additional information about the content and logic in the core drug modules, see the computer-assisted interviewing (CAI) specifications in the 2014 NSDUH methodological resource book at http://www.samhsa.gov/data/.

Figure 6.1 Routing Logic for Respondents in the Core Drug Modules



AFU = age at first use; MFU = month of first use; YFU = year of first use.

6.1.1 Edited (but Not Imputed) Drug Use Variables

Some core drug use variables each year undergo editing but no further imputation. Most of these variables pertain to use of tobacco products. Editing procedures for these tobacco variables are discussed in Section 6.2.9. Also, the variable corresponding to the usual number of alcoholic beverages that respondents consumed on days when they used alcohol in the past 30 days (corresponding to question AL07) is used in editing other variables that subsequently undergo imputation (e.g., frequency of consumption of five or more drinks on a single occasion in the past 30 days, also referred to as the frequency of "binge" alcohol use) but the variable corresponding to question AL07 itself is not included in the subsequent imputation procedures. Editing procedures for the variable corresponding to question AL07 are discussed in Section 6.2.4.7.

6.1.2 "Parent" and "Child" Drug Use Categories

In this chapter, some of the drug use measures refer to a general drug category (e.g., hallucinogens), and other measures refer to one or more subcategories within a general category (e.g., LSD, PCP, or Ecstasy as types of hallucinogens). These drug categories are described using the terms "parent drug" for the general drug category and "child drug" for the drug subcategory. Parent/child drug pairs often (but not always) occurred in modules that included multiple gate questions.[60] For example, the hallucinogens module includes questions about lifetime use of specific hallucinogens to determine the respondent's overall lifetime use or nonuse of any hallucinogen. Included in these questions are specific gate questions for LSD (LS01A), PCP (LS01B), and Ecstasy (LS01F). However, parent/child drug pairs also could appear in separate modules, such as for any use of cocaine and specific use of crack cocaine (i.e., if respondents reported lifetime use of any form of cocaine). Table 6.1 shows the drugs with parent/child relationships and the data that were collected for them.

Table 6.1 Core Drugs in the 2014 NSDUH with a Parent/Child Relationship

Parent Drug	Child Drugs	Parent Data Collected	Child Data Collected	"Other" Lifetime Use Indicator[1]
Smokeless Tobacco	Snuff, Chewing Tobacco	None; measures for initiation, recency, and tobacco brands were created from data for the child drugs[2]	Initiation[3], recency, 30-day frequency, tobacco brands	No
Cocaine	Crack	Initiation, recency, 12-month frequency, 30-day frequency	Initiation, recency, 12-month frequency, 30-day frequency	No
Hallucinogens	LSD, PCP, Ecstasy	Initiation, recency, 12-month frequency, 30-day frequency	Initiation, recency	Yes
Pain Relievers	OxyContin®	Initiation, recency, 12-month frequency	Initiation, recency, 12-month frequency	Yes
Stimulants	Methamphetamine	Initiation, recency, 12-month frequency	Initiation, recency, 12-month frequency	Yes

[1] See Section 6.2.8.1.
[2] A 30-day frequency for any smokeless tobacco use could not be determined if respondents were users of both snuff and chewing tobacco in the past 30 days.
[3] Initiation refers to the age at first use and, if applicable, the month and year of first use.

6.1.3 Special Situation for Methamphetamine, Stimulants, and Psychotherapeutics

Questions on methamphetamine use in NSDUH are first asked in the stimulants module in the core section of the questionnaire in the context of questions about nonmedical use of

[60] See Section 2.2.3 for a discussion of the gate question terminology.

prescription stimulants.[61] One concern in measuring methamphetamine use in NSDUH is that some methamphetamine users—particularly those who used it in the past 12 months—may fail to report use if they do not recognize the drug when it is presented in the prescription drug context.

To address this concern, questions were added to the special drugs module in the noncore section of the NSDUH questionnaire beginning in 2005 to capture information from respondents who may have used methamphetamine but did not recognize it as a prescription drug and therefore did not report use in the core stimulants module. Findings from the methamphetamine analysis report in the 2005 NSDUH methodological resource book (MRB; Ruppenkamp, Davis, Kroutil, & Aldworth, 2006) suggested that estimates of methamphetamine use based only on core data could be lower than the true population prevalence. However, larger estimates of methamphetamine use based on both core and noncore answers could be a partial artifact of asking a second set of questions *only* from respondents who did not report use the first time. Repeating questions for any drug only to those who did not report use the first time could artificially increase the positive responses. Doing so only for methamphetamine could result in a disproportionate reporting of that drug relative to the others in the survey. In addition, because the respondents reporting methamphetamine use in the noncore questions essentially had contradicted their prior responses, some may have made mistakes in answering the noncore questions. For these reasons, additional follow-up items have been included since the 2006 NSDUH to identify respondents who failed to report methamphetamine use in response to the earlier question in the core stimulants module because they may not have considered methamphetamine to be a prescription drug. Respondents who did not previously report methamphetamine use because they did not consider it to be a prescription drug have been counted in core-plus-noncore (CPN) estimates of methamphetamine use.

For the purpose of examining trends in nonmedical methamphetamine use, a Bernoulli stochastic imputation (BSI) procedure was used in conjunction with the predictive mean neighborhood (PMN) method (see Chapter 3) to generate comparable estimates for 2002 through 2005 (Ruppenkamp et al., 2007). In addition to CPN variables for methamphetamine, CPN variables were created in the 2014 NSDUH for nonmedical use of stimulants and nonmedical use of any psychotherapeutic drug. Section 6.2.6 discusses editing procedures to create the CPN data for methamphetamine and stimulants. Section 6.3.7 discusses procedures to create the final imputed CPN variables for methamphetamine, stimulants, and any psychotherapeutic drug.

6.2 Editing Procedures for the Core Drug Use Variables

This section discusses procedures for editing the core drug use variables for tobacco (i.e., cigarettes, snuff, chewing tobacco, cigars, and pipe tobacco), alcohol, marijuana, cocaine in any form, crack cocaine, heroin, hallucinogens, inhalants, and nonmedical use of prescription-type pain relievers, tranquilizers, stimulants, and sedatives. Edits of the core drug use variables encompassed the following key activities:

---

[61] Question ST01 in the core stimulants module asks, "Have you ever, even once, used Methamphetamine, Desoxyn, or Methedrine that was not prescribed for you or that you took only for the experience or feeling it caused? Methamphetamine also is known as crank, crystal, ice, or speed." Desoxyn® is available by prescription in the United States but is not commonly prescribed. Methedrine® is no longer available but could be relevant for measuring nonmedical use at some point in a person's lifetime.

- edits of the lifetime use variables (i.e., gate questions), where respondents indicated whether they have ever used the drug of interest (Section 6.2.1);
- edits of the recency-of-use variables, where respondents who indicated lifetime use of the drug indicated when they last used that drug (Section 6.2.2);
- edits involving users of only "child" drugs (Section 6.2.3);
- edits of the 12-month and 30-day frequency variables, where respondents who indicated use of a drug in the 12 months or 30 days prior to the interview indicated the number of days they used that drug in the period of interest (Section 6.2.4);
- edits for incidence (i.e., initiation) variables, where respondents who indicated lifetime use of a drug indicated when they first used it (Section 6.2.5);
- edits for methamphetamine and stimulants to take into account data for methamphetamine from the noncore special drugs module (Section 6.2.6);
- edits for lifetime daily cigarette use (Section 6.2.7);
- edits that were applied as part of the imputation processing (Section 6.2.8); and
- edits for drug use variables that do not undergo imputation (Section 6.2.9).

In connection with each of these edits, the discussion in the remainder of Section 6.2 focuses on relevant issues or inconsistencies in the data that needed to be addressed through logical editing. As noted in Chapter 1, the skip logic in the computer-assisted interviewing (CAI) instrument limited the chances for respondents to be routed to questions where they could give answers that were inconsistent with their answers to previous questions. For example, if respondents reported never using marijuana in question MJ01, the CAI program did not ask additional questions that would presume use of marijuana at least once. However, the CAI program did not completely eliminate all opportunities for respondents to provide inconsistent answers. Nevertheless, most processing of the CAI data was relatively straightforward, and the issues discussed in this section were not widespread relative to the total number of 67,901 respondents in 2014.

6.2.1 Edits of Lifetime Use Variables

As discussed in connection with the usable case criteria (Section 2.2.3), respondents were asked first whether they had ever used a drug of interest. For hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives (hereafter, "hallucinogens through sedatives"), respondents were asked a series of questions to establish whether they had ever used one or more specific types of drugs within that category (e.g., LSD, PCP, peyote). Only those respondents who indicated lifetime use of that drug (or lifetime use of one or more specific drugs within the respective modules for hallucinogens through sedatives) were asked more detailed questions about that drug (including situations in which respondents initially refused to answer a question about their lifetime use of a drug but then changed their answer to "yes" on follow-up; see

Sections 2.2.3 and 2.4.2). Unlike the first six substance use modules[62] (i.e., for tobacco products through heroin), however, there was no overall question about lifetime use or nonuse for the next six substance use modules, (i.e., hallucinogens through sedatives). Rather, lifetime use of any drug in the overall category for these latter six modules could be deduced from one or more affirmative answers regarding lifetime use of any of the specific drugs in that category. Because the modules for hallucinogens through sedatives included a question about use of "any other" drug in that category, answers of "no" for lifetime use of all of the specific drugs in that category (i.e., including lifetime use of "any other") logically indicated that the respondent had never used any drugs in that category.

Processing of the gate variables established whether (1) respondents had used a drug of interest at least once, (2) they had never used the drug, or (3) lifetime use or nonuse of the drug could not be determined. In addition to answering these gate questions as "yes" or "no," respondents could answer them as "don't know" or "refused." As noted in Section 2.4.2, final responses of "don't know" and "refused" were treated by the CAI skip logic as equivalent to situations where respondents never used the drug of interest. For the hallucinogens through sedatives, the CAI skip logic treated combinations of responses of "no," "don't know," and "refused" to the questions about individual drugs in the absence of any affirmative response to these questions in the same manner as if the respondent had answered all of these questions negatively. In these situations where a gate question was answered as "don't know" or "refused," the respondent's lifetime use or nonuse status was treated as unknown because these responses did not provide conclusive information one way or the other.[63] Cases with unknown lifetime use/nonuse status were subsequently imputed to be lifetime users or nonusers, as described in Section 6.3.

This step of the editing procedures also involved assignment of "bad data" codes to lifetime drug use variables (i.e., equivalent to missing data) if potential patterned responses previously had been identified (Sections 2.3.2 and 2.4.3). For the core interview sections pertaining to snuff through heroin,[64] these edits involved assignment of codes for bad data for the lifetime drug question and the follow-up probe, if respondents previously had refused the lead question (e.g., questions AL01 and ALREF for alcohol). For hallucinogens through sedatives, these edits involved assignment of codes for bad data to all lifetime variables in that section (e.g., data in questions LS01A through LS01H for hallucinogens and any associated "OTHER, Specify" data for hallucinogens).

---

[62] The CAI specifications show a single module for all tobacco products. Within this module, there are separate gate questions for cigarettes, snuff, chewing tobacco, cigars, and pipe tobacco. For pipe tobacco, there is only one question following the gate question for this tobacco product (i.e., any use in the past 30 days). In addition, the CAI specifications show a separate module for crack cocaine. As discussed in Section 6.2.1.1, routing of respondents to the crack cocaine module was dependent on whether they reported lifetime use of any form of cocaine.

[63] For multiple gate drugs, this included situations where respondents answered one or more of the gate questions as "don't know" or "refused" but did not answer any of the other gate questions affirmatively.

[64] Because of the requirement of the usable case criteria that respondents had to have defined data for lifetime use or nonuse of cigarettes (Section 2.2.3), assignment of the missing value of bad data for lifetime use of cigarettes would result in a case being reclassified as not usable.

6.2.1.1 Creation of Lifetime Use Variables for Tobacco Products through Heroin

Creation of edited variables for lifetime use or nonuse of tobacco products through heroin typically was straightforward because there were overall questions about lifetime use or nonuse. In most instances, therefore, codes corresponding to respondents' original answers simply were assigned to the relevant variable (e.g., 1 = Yes; 2 = No). When respondents were routed to a follow-up question because they initially refused to answer the first question about lifetime use or nonuse (Section 2.4.2), their response to this follow-up question was taken as the final response for the lifetime use variable.

The exception to these procedures involved lifetime use or nonuse of crack cocaine. Specifically, the logic to question CC01, regarding lifetime use of any form of cocaine, governed whether respondents were asked questions about crack cocaine. Therefore, the lifetime crack use variable CRKEVER (corresponding to question CK01) was assigned a code of 91 (i.e., never used cocaine) when COCEVER (corresponding to CC01) was answered as "no." If respondents reported lifetime use of cocaine but answered question CK01 as "no," then CRKEVER was assigned a code of 2.

6.2.1.2 Creation of Lifetime Use Variables for Hallucinogens through Sedatives

For the variables in the hallucinogens through sedatives sections that indicated lifetime use or nonuse of specific drugs, if respondents answered all gate questions in the series as "no," the edit procedures assigned a code of 91 to the entire series of variables (e.g., LSD through any other hallucinogen).[65] This was done to indicate that the respondents had never used any of the drugs in that category. In contrast, a code of 2 (i.e., "no") in edited variables for lifetime use of specific drugs within a broader category had the following meaning:

- the respondent was a user of at least one drug in the category but had never used the specific drug of interest (e.g., if a respondent was a lifetime user of LSD but not PCP); or
- the respondent reported never using the particular drug of interest but answered other questions in the series as "don't know" or "refused" (e.g., if a respondent did not know whether he or she had ever used LSD but definitely knew that he or she had never used PCP).

Variables also were created to indicate whether respondents had ever used one or more drugs within the overall drug category (e.g., any inhalant). Summary variables for these drugs could take on the following values:

- 1 ("Yes"). The summary variable for a given category was coded as 1 (i.e., "yes") if respondents answered "yes" for lifetime use of at least one specific drug in the overall category. The summary variable also was coded as 1 if respondents initially refused to answer all lifetime questions within one of these sections and then reported on

[65] In the hallucinogens module, this included situations in which respondents initially refused to report whether they had ever used LSD, PCP, or Ecstasy but then changed their answer(s) to "no" on follow-up. Similarly, in the stimulants module, this included situations in which respondents initially refused to report whether they had ever used methamphetamine but changed their answers to "no" on follow-up.

follow-up that they had used some drug in that category at least once. For example, if at least one affirmative answer appeared in questions about lifetime use of inhalants (including the follow-up question INREF) was answered as "yes," INHEVER was coded as 1.

- 91 ("Never Used"). This code of 91 could be assigned to the overall summary variable for one of two reasons: (1) if respondents answered all questions about lifetime use of individual drugs within the category as "no" or (2) if respondents were routed to a follow-up question because they refused to answer *all* questions about lifetime use of individual drugs within a category and then they reported that they never used that type of drug. As noted previously, this code of 91 also was assigned to all of the individual lifetime use variables within that section. For example, if respondents initially refused to answer all questions about lifetime use of specific inhalants, they were asked whether they had ever used any type of inhalant. If respondents answered this follow-up question as "no," the summary variable INHEVER was assigned a code of 91, and all individual lifetime inhalant variables were assigned a code of 91 as well.
- 97 ("Refused"). The summary variable was assigned a code of 97 if respondents initially refused to answer *all* questions about lifetime use for the specific drugs within one of the sections for inhalants through sedatives and then continued on follow-up to refuse to indicate whether they had ever used a drug within that overall drug category. In this case, the edited lifetime use variables for the individual drugs within that module also retained a code of 97. For example, INHEVER was assigned a code of 97 if a respondent refused to answer questions IN01A through IN01L for inhalants and also refused to answer the follow-up question INREF.
- 98 ("Blank"). The summary variable was assigned a code of 98 if respondents had one or more answers of "don't know" or "refused" to questions about lifetime use of specific drugs within the category (i.e., other than the previous situation in which respondents refused all questions) and they had no reports of lifetime use of any of these drugs. In this situation, the code of 98 indicated that lifetime use or nonuse for the overall category was unknown, even if one or more questions about lifetime use of specific drugs was answered as "no." For example, INHEVER was assigned a code of 98 if some questions in IN01A through IN01L were answered as "no" but other questions were answered as "don't know" or "refused."

6.2.1.3 Editing of "OTHER, Specify" Drug Variables for Nonusers

If respondents had never used any of the drugs in a series of multiple gate questions, they would not have been routed to questions where they could specify the use of some other drug in that overall category. In this situation, blank values in the unedited "OTHER, Specify" drug variables were replaced with a code of 9991 to indicate that the questions were skipped because the respondents had never used that class of drugs.

Similarly, if respondents reported in the preceding multiple gate questions that they used at least one drug in the category but they never used some other drug besides the ones they were asked about, they were legitimately skipped out of the questions that asked them to specify what "other" drug they had used. In this situation, blank values in the unedited "OTHER, Specify" drug variables were replaced with a code of 9999 (LEGITIMATE SKIP).

Functionally, the codes of 9991 and 9999 both indicate that respondents legitimately skipped out of the "OTHER, Specify" questions in that drug's section. However, the 9991 code provides for analysts the extra level of detail that the respondents were legitimately skipped out of these questions because they had never used anything within that category of drugs.

6.2.1.4 Use of "OTHER, Specify" Drug Data to Edit Lifetime Use Variables for Hallucinogens through Sedatives

Table B.3 in Appendix B describes edits that were relevant to the multiple gate variables for hallucinogens through sedatives based on "OTHER, Specify" data *within the same module* (Sections 2.4.1 and 2.4.5). For example, if a respondent did not report lifetime use of a specific drug in a module (e.g., PCP) but specified using it as another drug in the category (e.g., "marijuana laced with PCP" as another hallucinogen), the respondent was logically inferred to be a lifetime user of that drug within the relevant module where the "OTHER, Specify" response occurred (e.g., PCP for "OTHER, Specify" data within the hallucinogens module). The following code was assigned to the edited lifetime use variable for that drug: 3 = Yes LOGICALLY ASSIGNED.

This edit was especially relevant to LSD, PCP, Ecstasy, OxyContin®, and methamphetamine, where lifetime use or nonuse was imputed if respondents had missing data for the lifetime use questions. Consequently, no imputation would be needed for these drugs if a respondent reported not knowing whether he or she had used these drugs or refused to report use of these drugs when asked directly but the "OTHER, Specify" data in that module indicated use.

6.2.1.5 Editing Issues for Lifetime Psychotherapeutic Variables

In the sections for the psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives), respondents were asked if they had ever used any of the medications below the red line on that drug's pill card.[66] If respondents answered "yes," they were asked to indicate which of the drugs they had used. If the respondents answered "no," they were skipped out of these follow-up questions. Therefore, consistent with the procedures described at the beginning of Section 6.2.1.2, if respondents had never used *any* prescription-type psychotherapeutic medications in that category (i.e., in addition to never having used any of the medications below the red line on that drug's pill card), the edits assigned a code of 91 (i.e., "never used") to all of the specific drugs that were skipped (e.g., codeine through Ultram® in the PR04A series). In contrast, if respondents reported never using any of the medications below the red line on the pill card but they reported use of at least one other drug (or they answered at least one other gate question as "don't know" or "refused"), the skipped drug questions were assigned a code of 99 (LEGITIMATE SKIP).

---

[66] As an aid in answering the questions for psychotherapeutic drugs, respondents could look at printed "pill cards" that showed pictures of prescription drugs that were included in a given module. Pill cards included a thick red line that separated groups of drugs above and below the line. For example, question PR04 asked, "Please look at the pain relievers shown below the red line on Card A. Have you ever, even once, used any of these pain relievers when they were not prescribed for you or that you took only for the experience or feeling they caused?"

Tables B.3 and B.4 in Appendix B describe edits that were relevant to the gate variables for the psychotherapeutic drugs. As noted in Section 6.2.1.4, a code of 3 (Yes LOGICALLY ASSIGNED) could be assigned to gate variables if respondents did not report nonmedical use of specific psychotherapeutic drugs but they specified nonmedical use in response to question PR05. If respondents answered the lead question about nonmedical use of drugs "below the red line" (e.g., question PR04 for pain relievers) as "no" but they specified nonmedical use of any of these drugs, it was inferred that this question should have been answered as "yes." Therefore, the edited variable (e.g., ANLCARD, corresponding to question PR04) was assigned a code of 3, in addition to specific drugs in the PR04 series (e.g., morphine) being assigned this code.

The following additional codes could be assigned to the lifetime nonmedical use variables for psychotherapeutics:

- 4 = No LOGICALLY ASSIGNED,
- 81 = NEVER USED [DRUG] Logically assigned.

Assignment of both of these codes applied to situations where respondents specified use of over-the-counter (OTC) medications despite being instructed not to report about use of OTCs.

A code of 4 was assigned to the lead question for nonmedical use of any other drug in that category (e.g., ANLNOLST, corresponding to question PR05 for pain relievers) when respondents answered "yes" to at least one other gate question in that section for nonmedical use of psychotherapeutics or if respondents had answers of "don't know" or "refused" in other gate questions. Codes in the "OTHER, Specify" variables were overwritten with a code of 9989 (LEGITIMATE SKIP Logically assigned; see Section 2.4.2) to indicate that respondents logically should have skipped these questions. Assignment of a code of 4 to the lead question for nonmedical use of any other drug in that category did not affect the editing of other gate questions in that section.

A code of 81 was assigned when the respondent answered "no" to all of the questions about lifetime use of specific medications in that category except for use of any other medication in that category (e.g., any other pain reliever besides the ones shown on Pill Card A); and the *only* thing the respondent specified was an OTC medication, subject to the qualifications discussed in the remainder of this section.[67] If respondents who used only OTC drugs correctly followed instructions and answered all gate questions for a given psychotherapeutic category as "no," they would be skipped out of the remaining questions for that category of psychotherapeutic drugs (e.g., first use, most recent use). Consequently, the exclusive OTC users who incorrectly answered the question about nonmedical use of other drugs in the category as

[67] In the pain relievers and stimulants sections, this edit also involved assigning a code of 81 to the lifetime OxyContin® variable OXYCONTN and the methamphetamine variable METHDES, respectively, even though lifetime nonuse was not really logically inferred. In the pain relievers module, respondents already would have answered the lead question PR04 as "no," indicating that they had never used any prescription pain relievers below the red line on Pill Card A, including OxyContin®. Similarly, respondents would already have answered the lifetime methamphetamine question ST01 as "no," indicating that they never used methamphetamine, Desoxyn®, or Methedrine®. However, the code of 81 was assigned to OXYCONTN or METHDES for consistency with the assignment of the code of 81 to the other respective pain reliever or stimulant gate variables.

"yes" would comprise some unknown (and possibly unrepresentative) subset of exclusive OTC users.

Edits that assigned codes of 4 or 81 applied if the only responses in the "OTHER, Specify" data for a type of psychotherapeutic medication were OTCs, with the remaining "OTHER, Specify" variables having values of blank (Section 2.2.1) or bad data (i.e., where the bad data code denoted a nonsensical answer that the respondent keyed). These edits were *not* implemented if respondents had "OTHER, Specify" responses of "don't know" or "refused" in addition to specification of OTCs; such responses were interpreted to mean that the respondent was still a potential nonmedical user of some prescription-type medication, especially in situations where respondents may not have known what they ingested.

If other qualifying prescription-type medications were specified in addition to OTCs, the respondent's status as a nonmedical user was retained (e.g., if a respondent reported nonmedical use of a prescription pain reliever in addition to use of aspirin in the pain relievers section of the interview). Further, the OTC responses were retained in the respondent's "OTHER, Specify" variables. If a respondent reported use of a drug that may be available over the counter in certain strengths but is available in other strengths only by prescription, then the respondent's status as a nonmedical user of that category of prescription-type psychotherapeutics did not change. For example, specification of ibuprofen or Motrin® without a dosage could refer to use in prescription form, and this was assumed to be the case in the editing. However, specification of Advil® (i.e., an OTC dosage of ibuprofen) would be an unambiguous indication of use of an OTC drug. Certain drugs were treated as OTCs if they at one time had been available only by prescription but have become available over the counter without a prescription-strength counterpart (e.g., Benadryl®).

6.2.2 Edits of Recency-of-Use Variables

Edits of the variables that establish when respondents last used a drug of interest are probably the most critical.[68] These recency-of-use variables are the precursors for the final measures that establish the prevalence of use in the past 30 days, past 12 months, and lifetime.

The skip logic in the CAI instrument limited the kinds of information that were available for use in editing the recency-of-use variables. In particular, respondents who answered a gate question (or all multiple gate questions) as "no" (i.e., never used that drug) were not given the opportunity to answer additional questions as though they were users of that drug. Similarly, respondents who reported that they last used a drug "more than 12 months ago" were not given the opportunity to answer further questions in that module about use in the past 12 months or past 30 days, as though they were more recent users than what they had originally indicated (Figure 6.1).

6.2.2.1 Edits to Recency Variables for Nonusers

As noted in Sections 2.2.3 and 2.4.2 in Chapter 2, respondents were skipped out of all remaining questions about use of a particular drug when they answered "no" to a given gate

[68] For brevity, the term "use" in the remainder of Section 6.2 also refers to nonmedical use of prescription psychotherapeutic drugs.

question for cigarettes through heroin or when they answered "no" to all of the individual gate questions for hallucinogens through sedatives.

A code of 91 (i.e., "never used") typically was assigned to the core recency variables when it was determined unambiguously that respondents had never used the drug of interest. For crack cocaine, this situation also held if the lifetime variable CRKEVER was coded as 91 because the respondent reported never using cocaine in any form. Similarly, for hallucinogens through sedatives, the recency variables were assigned a code of 91 if the all of the lifetime use variables corresponding to the gate questions had been assigned a code of 91 because respondents had never used any of the individual drugs (Section 6.2.1.2).

However, editing of the OxyContin® recency variable OXYCREC was handled somewhat differently because the lifetime OxyContin® variable OXYCONTN came from an "enter all that apply" item. In turn, respondents' reports of whether they had ever used OxyContin® nonmedically governed whether they were asked the OxyContin® recency question. Therefore, if respondents had never used any prescription pain reliever nonmedically (and OXYCONTN = 91), then OXYCREC was assigned a code of 91 because the data were conclusive that the respondent had never misused any prescription pain relievers, including OxyContin®. Similarly, OXYCREC was assigned a code of 91 if the edited variable ANLCARD was answered as "no" (i.e., ANLCARD = 2). Because a picture of OxyContin® was shown below the red line on Pill Card A, a response of "no" in ANLCARD was taken as an unambiguous indication that the respondent had never used OxyContin® nonmedically.

In comparison, if the lifetime OxyContin® variable OXYCONTN had a code of 6 ("Response not entered"; see Section 2.4.4), then that was not as strong of an indication that the respondent had never used OxyContin® nonmedically, compared with questions in which respondents explicitly were required to answer "yes" or "no" regarding whether they had ever used the drug of interest. Therefore, when OXYCONTN = 6 because OxyContin® was not chosen from the list of drugs in question PR04A, the recency variable OXYCREC was assigned a code of 81 (NEVER USED OXYCONTIN Logically assigned), as opposed to a code of 91. That is, it was *logically* inferred that respondents had never used OxyContin® when respondents reported nonmedical use of some pain relievers from the PR04A list but not OxyContin®. These indications of 81 were treated the same as the code of 91 in other edits. However, use of the code of 81 introduced the additional detail for analysts that a logical inference had been made about respondents having never used OxyContin®. In contrast, if respondents had been asked, "Have you ever, even once, used OxyContin…" and they answered "no," it would not be necessary to make a logical inference because the respondents would have explicitly reported that they never used it nonmedically.

In addition, a code of 81 was assigned to the relevant edited recency variables for pain relievers (ANALREC), tranquilizers (TRANREC), stimulants (STIMREC), or sedatives (SEDREC) if respondents' only reported nonmedical use of prescription drugs in that module involved use of OTCs (Section 6.2.1.5). A code of 81 also was assigned to the recency variables for OxyContin® (OXYCREC) and methamphetamine (METHREC) if respondents' only reported nonmedical use of pain relievers or stimulants, respectively, involved OTCs. Although respondents would not have been lifetime nonmedical users of OxyContin® or methamphetamine in order to report lifetime nonmedical use of only OTCs, a code of 81 was assigned to these child recency variables (i.e., rather than a code of 91) for consistency with the code that was assigned to the corresponding parent recency variable.

6.2.2.2 Recency Periods in the Questionnaire

In the core modules for alcohol through sedatives, the following standard codes for recency applied, based on the available responses to the question(s) about most recent use:

- 1 = Within the past 30 days,
- 2 = More than 30 days ago but within the past 12 months, or
- 3 = More than 12 months ago.

In the sections of the tobacco module for cigarettes, snuff, chewing tobacco, and cigars, lifetime users first were asked whether they used the relevant tobacco product in the past 30 days (e.g., question CG05 for cigarettes). If respondents reported that they did not use the particular tobacco product in that period, they were asked to report most recent use prior to the past 30 days (e.g., question CG06 for cigarettes). Recency questions for these tobacco products included categories for most recent use (1) more than 30 days ago but within the past 12 months, (2) more than 12 months ago but within the past 3 years, and (3) use more than 3 years ago.[69]

For these tobacco products, single edited recency variables were created that combined the data for the questions about use in the past 30 days and most recent use in later periods, if applicable. Codes of 1 and 2 in the edited recency variables for these tobacco products had the same meaning as for alcohol through sedatives. In addition, standard codes of 3 and 4 in these edited tobacco recency variables had the following meanings based on respondents' answers to the follow-up question (e.g., question CG06 for cigarettes) if they did not report use in the past 30 days:

- 3 = More than 12 months ago but within the past 3 years, and
- 4 = More than 3 years ago.

The CAI instrument included follow-up probes for respondents who were lifetime users of a given drug but did not know or refused to report when they last used it.[70] Respondents who initially did not know when they last used a drug were asked to give their "best guess" of when they last used it. Respondents who initially refused to report when they last used a drug were asked to reconsider answering the question. If respondents changed their initial answer of "don't know" or "refused" to report a definite period when they last used the drug of interest, then that information served as the starting point for subsequent editing of the drug's recency variable. The standard recency codes described previously in this section were assigned according to the specific period of most recent use that was reported in these follow-up probes. In the absence of any inconsistencies between the recency-of-use answers in the follow-up probes and other data

[69] A recency variable was not created for pipe tobacco because lifetime users of pipe tobacco were asked only if they smoked tobacco in a pipe in the past 30 days.

[70] For cigarettes, snuff, chewing tobacco, and cigars, these follow-up probes were asked if respondents reported that they did not use a given tobacco product in the past 30 days but they answered "don't know" or "refused" to the follow-up question about most recent use more than 30 days ago.

within a given drug's module, these answers from the follow-up probes were accepted as final and were incorporated within the edited recency variable.

6.2.2.3 Logically Assigned Recency Periods among Lifetime Users

The following code was assigned to an edited recency variable for alcohol through sedatives if respondents reported lifetime use of a drug but continued on follow-up to answer "don't know" or "refused" regarding when they last used it: 9 = Used at some point in the lifetime LOGICALLY ASSIGNED.

These respondents were eligible to be statistically imputed to be users in any period, including in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago. Similarly, for respondents whose recency-of-use questions for LSD, PCP, Ecstasy, OxyContin®, or methamphetamine had been skipped but they were logically inferred to be lifetime users of these drugs based on their responses to "OTHER, Specify" items (Section 6.2.1.4), their edited recency variables at least initially were assigned a code of 9.

In the recency variables for cigarettes, snuff, chewing tobacco, and cigars, the following code was assigned if respondents reported that they did not use the tobacco product in the past 30 days but they had missing data for the period more than 30 days ago when they last used it: 19 = Used more than 30 days ago LOGICALLY ASSIGNED.

These respondents were eligible to be statistically imputed to have most recently used a given tobacco product more than 30 days ago but within the past 12 months, more than 12 months ago but within the past 3 years, or more than 3 years ago.

The following additional codes could be assigned to the edited recency variables when respondents reported lifetime or more recent use:

- 8 = Used at some point in the past 12 months LOGICALLY ASSIGNED,
- 11 = Used in the past 30 days LOGICALLY ASSIGNED,
- 12 = Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED,
- 14 = Used more than 12 months ago LOGICALLY ASSIGNED,[71]
- 29 = Used more than 30 days ago but within the past 3 years LOGICALLY ASSIGNED, and
- 39 = Used at some point within the past 3 years LOGICALLY ASSIGNED.

Sections 6.2.2.4, 6.2.2.5, and 6.2.2.6 discuss situations in which these codes were assigned based on the editing procedures. In particular, codes of 11 or 12 could be assigned to a parent recency variable (e.g., any hallucinogen) based on data from a child recency variable (e.g., LSD). Recency variables did not require further statistical imputation when codes of 11 or 12 were assigned.

[71] Codes of 14, 29, and 39 applied only to the recency variables for cigarettes, snuff, chewing tobacco, and cigars.

Recency variables that were assigned codes of 8, 9, 14, 19, 29, or 39 underwent further statistical imputation to assign a specific period of most recent use. These codes were used to place constraints during the subsequent imputation process for assigning the period of most recent use. For example, assigning a code of 8 to an edited recency variable constrained the imputation for the final recency to be within the past 30 days or more than 30 days ago but within the past 12 months.

6.2.2.4 Application of "Flag and Impute" Procedures to Editing of Recency-of-Use Variables

The procedures used to edit the CAI recency-of-use variables were referred to as the "flag and impute" procedures (Section 2.4.6). Under these procedures, the limited situations where potential inconsistencies existed between a respondent's answer to a drug's recency question and other data in that module were identified and flagged.

Table 6.2 lists the usual types of inconsistencies that could occur between a drug's recency variable and other variables in that drug's module, and how these inconsistencies were handled through the flag and impute procedures. In the situations described in Table 6.2, these inconsistencies were handled by statistically imputing final values for the affected recency variable and the other variable(s) where the data were inconsistent with the respondent's original answer to the recency question. Most of the codes that were described in Section 6.2.2.3 were assigned to the edited recency variables as a result of these edits. For example, if a respondent reported first use of a marijuana at his or her current age but also reported most recent use more than 12 months ago, both answers logically cannot be true. In this situation, the edited recency variable MJREC was assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) and the inconsistent initiation variables were assigned codes for bad data (Sections 2.3 and 2.4.3).

Table 6.2 How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables

Type of Inconsistency	Edits Implemented
Recency originally indicates use in the past 30 days, but use on 0 days in the past 30 days is confirmed (suggesting less recent use).	**The edited recency variable was assigned a code of 8** (i.e., Used at some point in the past 12 months LOGICALLY ASSIGNED) to indicate that the respondent (R) was at least a user in the past 12 months and potentially a user in the past 30 days (Section 6.2.2.3). The 30-day frequency was set to missing (i.e., bad data).
Recency originally indicates use more than 30 days ago but within the past 12 months, but the 12-month frequency indicates use on more than 335 days in that period (suggesting past month use).	**The edited recency variable was assigned a code of 8**. The 12-month frequency data were set to missing.

Table 6.2 How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables (continued)

Type of Inconsistency	Edits Implemented
Recency does not indicate use in the past 30 days, but the R reports first using the drug (or smoking cigarettes daily) in the same month as the interview took place (suggesting past month use).	**If the recency originally indicated use more than 30 days ago but within the past 12 months, it was assigned a code of 8. If the recency originally indicated use more than 12 months ago (or was missing), it was assigned a code of 9** (i.e., Used at some point in the lifetime LOGICALLY ASSIGNED) to indicate that the R was at least a lifetime user (and potentially a user in the past 12 months or past 30 days). The values in the month of first use (MFU) and year of first use (YFU) that triggered the inconsistency with the recency-of-use answer were overwritten with bad data codes. The MFU and YFU variables also were set to bad data if the recency was missing and had been assigned a code of 9 because the initiation data would have suggested use in the past 30 days.
Recency does not indicate use in the past 30 days, but the R has other data suggesting initiation of use in the past 30 days (e.g., if first use was indicated at the R's current age and the R's last birthday was fewer than 30 days ago, or based on a comparison of the 12-month frequency and the maximum number of days that the R could have used the drug).	**The edited recency variable was assigned a code of 11** (Used in the past 30 days LOGICALLY ASSIGNED) to infer that the R was a past month user. No further editing was done to the variables indicating use in the past 30 days. This edit was an exception to the general "flag and impute" procedures.
Recency does not indicate use in the past 12 months, but the age at first use (AFU; or age at first daily use of cigarettes) equals the R's current age (suggesting past year use).	**For alcohol through sedatives, the edited recency variable was assigned a code of 9**. **For tobacco products, if the lead 30-day question** (e.g., CG05 for cigarettes) **indicated that the R did not use in the past 30 days, then the recency was assigned a code of 19** (Used more than 30 days ago LOGICALLY ASSIGNED) to indicate that the R was a user more than 30 days ago (but was potentially a user in the past 12 months). **Otherwise, the relevant tobacco recency variable was assigned a code of 9.** The AFU that triggered the inconsistency with the recency-of-use answer was set to bad data. If the MFU and YFU were answered (i.e., not blank), the values in these variables were overwritten with bad data codes. (Month and year data were considered to be linked with the AFU data. Therefore, if the AFU was questionable, then the month and year were considered to be questionable as well.)
Recency does not indicate use in the past 12 months, but the R reported first using in a month and year that falls within 12 months of the interview date (including data for the month and year when the R reported first smoking cigarettes daily).	**For alcohol through sedatives, the edited recency variable was assigned a code of 9**. **For tobacco products, if the lead 30-day question** (e.g., CG05 for cigarettes) **indicated that the R did not use in the past 30 days, then the recency was assigned a code of 19** (Used more than 30 days ago LOGICALLY ASSIGNED) to indicate that the R was a user more than 30 days ago (but was potentially a user in the past 12 months). **Otherwise, the relevant tobacco recency variable was assigned a code of 9.** The values in the MFU and YFU that triggered the inconsistency with the recency-of-use answer were overwritten with bad data codes.

Table 6.2 How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables (continued)

Type of Inconsistency	Edits Implemented
For tobacco products, recency indicates use more than 3 years ago, but AFU (or first daily use of cigarettes) indicates that the first use was within 2 years of the R's current age (suggesting use within the past 3 years). (Note that edits checking for indications of use at the R's current age were given priority over the condition described here because the former response pattern suggests use in the past year.)	**The recency was assigned a code of 14** (Used more than 12 months ago LOGICALLY ASSIGNED) to indicate that the R last used at some point more than 12 months ago (but potentially in the past 3 years). The AFU that triggered the inconsistency with the recency-of-use answer was set to bad data. If the MFU and YFU were answered (i.e., not blank), the values in these variables were overwritten with bad data codes.
For tobacco products, the R did not use in the past 30 days but did not report a specific period of most recent use more than 30 days ago. However, the R reported first use within 1 or 2 years of his or her current age (i.e., but not at the R's current age).	**The recency was assigned a code of 29** (Used more than 30 days ago but within the past 3 years LOGICALLY ASSIGNED) to indicate that the R last used at some point more than 30 days ago but within the past 3 years (but potentially within the past 12 months).
For tobacco products, the R did not indicate whether he or she used in the past 30 days. However, the R reported first use within 1 or 2 years of his or her current age (i.e., but not at the R's current age).	**The recency was assigned a code of 39** (Used at some point in the past 3 years LOGICALLY ASSIGNED) to indicate that the R last used at some point within the past 3 years (but potentially within the past 30 days or past 12 months).
For chewing tobacco and snuff, the brand of chewing tobacco that Rs reported using most often in the past 30 days was really a snuff brand, or vice versa.	**Created a recoded any smokeless tobacco recency (SLTREC)** that incorporated data from the chewing tobacco and snuff recency variables CHEWREC and SNFREC, respectively. Thus, for example, Rs who reported using chewing tobacco in the past 30 days but specified a snuff brand as the brand they used most often in that period would still be considered a past month user of some type of smokeless tobacco product.

CAI = computer-assisted interviewing.

NOTE: Indications of most recent use include answers from follow-up probes for the recency questions.

Prior to implementing these flag and impute rules, initial edits checked for situations where a respondent's original answer to an AFU question might have been inconsistent with his or her recency of use, but a revised AFU was not. For example, if a respondent reported first using a drug at his or her current age, the respondent was asked to verify this AFU. If the respondent reported that this AFU was *not* correct but then on follow-up did not know at what age he or she first used, or refused to answer, the edits updated the AFU to reflect this "don't know" or refusal response. The rationale for this edit was that the respondent indicated that the initial answer was not correct. A final answer of "don't know" or "refused" to an AFU question would not necessarily be inconsistent with a reported recency of use more than 12 months ago.

6.2.2.5 Most Recent Use of Smokeless Tobacco

Table 6.2 also lists edits that applied to a special situation for chewing tobacco and snuff. When the CAI instrument was first fielded in 1999, considerable cross-reporting of chewing tobacco and snuff brands was observed among users in the past 30 days, suggesting that respondents were not always clear about the differences between these two types of smokeless tobacco. For example, respondents could report using chewing tobacco in the past 30 days but specify a snuff brand as the brand of "chewing tobacco" they used most often in that period. However, this cross-reporting was identifiable only for respondents who reported use in the past 30 days of either smokeless tobacco product but was assumed to be operating for respondents who reported less recent use. For this reason, a recoded smokeless tobacco recency variable SLTREC was created from the respective chewing tobacco and snuff recency variables (CHEWREC and SNFREC, respectively). Thus, if a respondent reported use of chewing tobacco in the past 30 days but specified use of a snuff brand in the past 30 days, the respondent was still a smokeless tobacco user in that period.

In creating the recoded SLTREC, indications of more recent use of chewing tobacco or snuff were given precedence over indications of less recent use. In situations where one recency variable indicated use in a definite period (e.g., more than 30 days ago but within the past 12 months) and the second recency variable indicated use in an indefinite period (e.g., use at some point in the lifetime, which could have included use in the past 30 days, past 12 months, or past 3 years), the final assignment to SLTREC indicated a less definite recency value. The rationale for this procedure was that the respondent was potentially a user in a more recent period. For example, if a respondent indicated use of chewing tobacco more than 30 days ago but within the past 12 months and the flag and impute rules had assigned a code of 9 to the snuff recency to indicate that the respondent last used snuff at some point in his or her lifetime, the recoded SLTREC was assigned a code of 8 (Section 6.2.2.3) to indicate use at some point in the past 12 months. That is, the report of chewing tobacco use in the past 12 months (but not the past 30 days) could be used to narrow down the use of any smokeless tobacco to some point in the past 12 months, but the respondent could still have used in the past 30 days. Similarly, if one of the recency variables had a missing value but the other did not, the SLTREC variable was assigned a code to indicate that there was some uncertainty about when the respondent last used smokeless tobacco. Suppose, for example, that a respondent reported last using chewing tobacco more than 12 months ago but within the past 3 years, but refused to report whether he or she had ever used snuff. In this situation, the SLTREC variable was given a code to indicate that the respondent used smokeless tobacco at some point in the lifetime because the respondent may have used snuff within the past 12 months or past 30 days.

6.2.2.6 Edits to Parent and Child Recency Variables Other than Smokeless Tobacco

Important exceptions to the general flag and impute principles involved situations where inconsistencies existed between parent and child recency variables (any cocaine and crack cocaine; any hallucinogen use and LSD, PCP, or Ecstasy use; any pain reliever use and OxyContin®; any stimulant use and methamphetamine use). These are presented in Table 6.3, along with a description of how the data were edited when specific types of inconsistencies occurred between related recency variables. In these special situations, indications of use of the child drug (e.g., crack cocaine) that were more recent than that indicated for the parent drug category (e.g., cocaine in any form) were used to logically infer more recent use of the parent drug category. For example, not all respondents might make the connection that crack cocaine fits within the broader category of cocaine in general.

Table 6.3 How the Flag and Impute Edit Procedures Handled Inconsistencies between Parent and Child Recency Variables

Child Recency (i.e., crack, LSD, PCP, Ecstasy, OxyContin®, methamphetamine) Reported by Respondent	Parent Recency (i.e., any cocaine, any hallucinogen, any pain reliever, any stimulant) Reported by Respondent	Edited Child Recency (i.e., crack, LSD, PCP, Ecstasy, OxyContin®, methamphetamine)	Edited Parent Recency (i.e., any cocaine, any hallucinogen, any pain reliever, any stimulant)
(1) Indicates use in past month.	Indicates use that is less recent than the past month.	Retains the recency reported by the respondent (R).	Logically infers the R to be a past month user. Assigns a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED).
(2) Indicates use more than 30 days ago but within the past 12 months.	Coded as 8, indicating use at some point in the past 12 months (Section 6.2.2.3).	Retains the recency reported by the R.	Retains the code of 8 to indicate that the R has used at some point in the past 12 months.
	Indicates use more than 12 months ago.	Retains the recency reported by the R.	Logically infers the R to have last used more than 30 days ago but within the past 12 months. Assigns a code of 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED).
	Coded as 9, indicating use at some point in the lifetime (Section 6.2.2.3).	Retains the recency reported by the R.	Logically infers the R to be *at least* a past year user. Assigns a code of 8.
(3) Indicates use more than 12 months ago.	Coded as 8, indicating use at some point in the past 12 months.	Retains the recency reported by the R.	Retains the code of 8 to indicate that the R is *at least* a past year user.
	Coded as 9, indicating use at some point in the lifetime.	Retains the recency reported by the R.	Retains the code of 9 to indicate that the R is *at least* a lifetime user.
(4) Coded as 8, indicating use at some point in the past 12 months (see Table 6.1).	Indicates use more than 30 days ago but within the past 12 months.	Retains the code of 8 to indicate use at some point in the past 12 months.	Assigns a code of 8 to indicate use at some point in the past 12 months.
	Coded as 8, indicating use at some point in the past 12 months.	Retains the code of 8 to indicate use at some point in the past 12 months.	Retains the code of 8 to indicate use at some point in the past 12 months.
	Coded as 9, indicating use at some point in the lifetime.	Retains the code of 8 to indicate use at some point in the past 12 months.	Assigns a code of 8 to indicate use at some point in the past 12 months.

Table 6.3 How the Flag and Impute Edit Procedures Handled Inconsistencies between Parent and Child Recency Variables (continued)

Child Recency (i.e., crack, LSD, PCP, Ecstasy, OxyContin®, methamphetamine) Reported by Respondent	Parent Recency (i.e., any cocaine, any hallucinogen, any pain reliever, any stimulant) Reported by Respondent	Edited Child Recency (i.e., crack, LSD, PCP, Ecstasy, OxyContin®, methamphetamine)	Edited Parent Recency (i.e., any cocaine, any hallucinogen, any pain reliever, any stimulant)
(5) Coded as 9, indicating use at some point in the lifetime.	Indicates use more than 30 days ago but within the past 12 months.	Retains the code of 9 to indicate use at some point in the lifetime.	Assigns a code of 8 to indicate use at some point in the past 12 months.
	Indicates use more than 12 months ago.	Retains the code of 9 to indicate use at some point in the lifetime.	Assigns a code of 9 to indicate use at some point in the lifetime.
	Coded as 9, indicating use at some point in the lifetime.	Retains the code of 9 to indicate use at some point in the lifetime.	Retains the code of 9 to indicate use at some point in the lifetime.

NOTE: These edits take place after inconsistencies have been identified between a recency variable and nonrecency variable (e.g., between the recency and the age at first use). For hallucinogens/LSD/PCP/Ecstasy, pain relievers/OxyContin®, and stimulants/methamphetamine, these edits also take place after the R has revised one or more answers in response to a consistency check. Further, for hallucinogens/LSD/PCP/Ecstasy, pain relievers/OxyContin®, and stimulants/methamphetamine, any inconsistencies that remain between a given recency variable and other nonrecency variables following inconsistency resolution are transferred back into the recency variables prior to implementation of these edits. For example, if the original answer to the hallucinogen recency disagreed with the age at first use but the revised recency in response to the consistency check did not, then the recency would be updated to reflect the revised value. Prior to implementation of the edits shown in the table, however, if the revised recency still disagreed with the age at first use, then the recency would be edited further to reflect the fact that the previous inconsistency still remained.

Therefore, if a respondent reported last using any cocaine more than 30 days ago and also reported last using crack cocaine in the past 30 days, the edit procedures assigned a code of 11 to the edited cocaine recency variable COCREC (Section 6.2.2.3). This edit indicated logical inference that the respondent had used cocaine in any form in the past 30 days. Overall, however, imputation played a more prominent role than editing in resolving inconsistencies with respect to the most recent use of a drug.

In addition, special patterns could remain in the data for cocaine, hallucinogens, or stimulants after most inconsistencies had been addressed and some related data elements were missing. Specifically, respondents could indicate that they first used any cocaine, any hallucinogen, or any stimulant within 12 months of the interview date (e.g., first use at their current age) and indicate that they last used that drug more than 30 days ago but within the past 12 months (e.g., HALLREC = 2 for any hallucinogen use). Logically, then, if respondents who had used a child drug within a given category (e.g., LSD, PCP, or Ecstasy) had missing data on first use for one or more of their child drugs, it followed not only that they had to have *first* used a child drug at some point in the past 12 months, but they also had to have *last* used the child drug at some point within the past 12 months. In this situation, both the parent recency (e.g., HALLREC) and child recency variable(s) (e.g., LSDREC, PCPREC, or ECSREC) were set to values of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) so that they would be imputed consistently. If the parent recency indicated past month use (e.g., HALLREC

= 1) when initiation of the parent occurred in the past 12 months but a child recency (e.g., LSDREC) did not indicate use in the past 12 months, only the child recency was set to a value of 8.

Similarly, respondents could indicate that they first used any cocaine, any hallucinogen, any pain reliever, or any stimulant in the same month that they were interviewed and indicate that they last used the drug in the past 30 days but have missing data on first use for the child drug(s) within a category. In this situation, respondents who had used a child drug within that category (e.g., LSD) also were inferred to have last used that drug in the past 30 days. The edits assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED) to the child recency variables (e.g., LSDREC = 11).

6.2.3 Edits Involving Users of Only Child Drugs

In addition to the situations described in Table 6.2, special edits were applied in situations in which respondents were users of only the child drug (for pain relievers and stimulants) or of only one child drug (for hallucinogens). The following specific patterns indicated that respondents were users of only the child drug:[72]

- for hallucinogens, a gate question for only one of the child drugs (i.e., LSD, PCP, or Ecstasy) was answered as "yes" and all other gate questions for hallucinogens were answered as "no";
- for pain relievers, only the response in question PR04A for OxyContin® was chosen and all other gate questions for pain relievers were answered as "no";
- for stimulants, only the gate question for methamphetamine was answered as "yes" and all other gate questions for stimulants were answered as "no"; or
- for hallucinogens, pain relievers, or stimulants, only the gate question for use of any other drug in the category was answered as "yes" (i.e., including the gate questions for the child drugs) but the *only* report in the "OTHER, Specify" data for that module was for use of a child drug.[73]

Respondents were *not* considered to be users of only the child drug if they reported lifetime use of only the child drug but they had responses of "don't know" or "refused" for any of the other gate questions. For example, if a respondent answered the question about lifetime use of LSD as "yes" and had some responses of "don't know" or "refused" for other hallucinogen gate questions and no reports of lifetime use of hallucinogens from the other gate questions, the

[72] For brevity, "use of only the child drug" in the remainder of this section also refers to use of only one child drug for hallucinogens.

[73] Respondents were considered to be users of only the child drug if they had codes only for "blank" or "bad data" in their "OTHER, Specify" data (see Sections 2.2.1 and 2.3) in addition to the report for a child drug. However, respondents were considered *not* to be users of only the child drug if they had codes for "don't know" or "refused" in their "OTHER, Specify" data in addition to the report for the child drug; responses of "don't know" or "refused" were considered to be potential indications of use of another drug besides the child drug. This procedure is consistent with the editing procedures described in Section 6.2.1.5 for identifying respondents who reported use of only OTCs in a prescription drug section.

respondent could have been a lifetime user of any of the specific hallucinogens for which he or she had missing data.

For hallucinogens, users of only one child hallucinogen who answered the gate question for one of the child drugs as "yes" and answered all remaining gate questions as "no" were asked questions about initiation and recency for the overall parent category. The corresponding questions were skipped for the child drug. This also was the pattern in the data if respondents reported lifetime use only of "other" hallucinogens and specified lifetime use of only a single child drug.

For pain relievers and stimulants, users of only OxyContin® or methamphetamine, respectively, who reported nonmedical use in the relevant gate question (e.g., question ST01 for methamphetamine) were asked questions about initiation, recency, and frequency of use in the past 12 months for the *child* drug. The corresponding questions were skipped for the parent drug. In contrast, for respondents who were identified as nonmedical users of only OxyContin® or methamphetamine based on their "OTHER, Specify" data (i.e., they reported use only of "other" pain relievers or stimulants, respectively but specified only the child drug), the questions were asked for initiation, recency, and 12-month frequency of use of the *parent* drug, and the child drug questions were skipped.

6.2.3.1 Edits to Child Hallucinogen Variables for Users of Only One Child Drug

If respondents reported use of only one child hallucinogen, the values from the parent initiation and recency variables were assigned to the corresponding variables that had been skipped for the child drug. For example, if a respondent reported lifetime use of only LSD, the value from the edited AFU variable for hallucinogens (HALLAGE) and the value from the edited hallucinogen recency variable (HALLREC) were assigned to the corresponding variables for LSD (LSDAGE and LSDREC, respectively). Similarly, any values for the MFU and YFU for recent initiates of any hallucinogen (HALYFU and HALMFU, respectively) were assigned to the corresponding year and month variables for LSD. These values also were retained in the edited variables for use of any hallucinogen. Section 6.2.5 provides further details about the content of the questions in the core drug modules for initiation of use.

6.2.3.2 Edits to Pain Reliever and Stimulant Variables for Users of Only the Child Drug

If respondents reported use of only the child drug and this report was not logically inferred from the "OTHER, Specify" data, the values from the child initiation, recency, and 12-month frequency variables were assigned to the corresponding variables that had been skipped for the parent drug. These values also were retained in the edited variables for nonmedical use of OxyContin® or methamphetamine. For example, if all gate questions for pain relievers except PR04 had been answered as "no" and OxyContin® was the only pain reliever that had been chosen from the list in PR04A, the values from the edited variables for OxyContin® (e.g., OXYCREC for most recent use) were assigned to the corresponding variables for pain relievers (e.g., ANALREC). Section 6.2.4 provides further details about the content of the questions in the core drug modules that established the frequency of use in the past 12 months. As noted previously, Section 6.2.5 provides further details about the content of the questions in the core drug modules for initiation of use, including initiation of nonmedical use of pain relievers or stimulants.

If respondents reported use of only the child drug but this report was logically inferred from the "OTHER, Specify" data, the values from the *parent* initiation, recency, and 12-month frequency variables were assigned to the corresponding variables that had been skipped for the *child* drug. For example, if all gate questions for pain relievers except PR05 (nonmedical use of any other pain reliever) had been answered as "no" and OxyContin® was the *only* pain reliever that had been specified, the values from the edited variables for pain relievers (e.g., ANALREC for most recent use) were assigned to the corresponding variables for OxyContin® (e.g., OXYCREC).

6.2.4 Edits of Frequency-of-Use Variables

The CAI instrument included questions about the number of days that respondents used different drugs in the past 30 days or past 12 months (or the average number of days per week or days per month that they used in the past 12 months). These are referred to in this section as 30-day and 12-month frequency variables, respectively. Data from these frequency questions can be used to distinguish between occasional and more frequent users of a drug. For example, frequent users of alcohol and illicit drugs may represent a group who are potentially in need of substance abuse treatment or other services for their substance use. Similarly, regular users of tobacco products, such as people who smoked cigarettes every day in the past 30 days, probably represent a group that would have greater difficulty stopping their use of tobacco. In addition, the alcohol section included a question about the number of days that respondents consumed five or more drinks per occasion in the past 30 days, a question which is used to construct measures of binge and heavy alcohol use in that period.[74]

6.2.4.1 Preferred Ways of Reporting 12-Month Frequency

For the 12-month frequency determinations, respondents first were asked how they preferred to report their frequency of use in the past 12 months. Respondents could indicate a preference to report their frequency of use in one of three ways: (1) use on an average number of days per week in the past 12 months, (2) use on an average number of days per month in the past 12 months, and (3) the total number of days they used in the past 12 months. In particular, respondents who used a drug regularly in the past 12 months might find it easier to report their frequency of use in one of the first two ways as opposed to figuring the total number of days they used in that entire period. Conversely, respondents who used on only a few days in the past 12 months might prefer the third reporting method.

Individual variables (subsequently referred to as "source variables") were created for the preferred way of reporting the 12-month frequency and for the associated follow-up questions (i.e., average number of days per week, average number of days per month, or total number of days in the past 12 months). An overall measure of the number of days that respondents used a substance in the past 12 months was created from these individual source variables.

[74] Binge alcohol use refers to the consumption of five or more drinks on a single occasion on at least 1 day in the past 30 days. Heavy alcohol use refers to the consumption of five or more drinks on a single occasion on 5 or more days in that period.

For respondents who chose to report a total number of days that they used a substance, the 12-month frequency was the actual number of days that the respondent reported using the drug in the past 12 months (assuming no inconsistency with the 30-day frequency; these inconsistencies are discussed in Section 6.2.4.4). For respondents who chose to report an average use in days per week or days per month, the overall number of days that they used in the past 12 months was a calculated value. Specifically, answers in terms of the average number of days used per week in the past 12 months were multiplied by 52, and answers in terms of the average number of days used per month in the past 12 months were multiplied by 12 in order to yield a calculated 12-month frequency. Because these latter two response options were averaged responses over the past 12 months, no further adjustments were made to the calculated 12-month frequency value when respondents used the drug more than 30 days ago but within the past 12 months, and they did not initiate use at some point in the past 12 months.

6.2.4.2 Editing of 12-Month and 30-Day Frequency Variables for Nonusers or Less Recent Users

If the lifetime gate question(s) and edited recency-of-use variable indicated that the respondent had never used the drug of interest,[75] then edits at this step assigned a code of 91 to the 30-day frequency variable (where applicable)[76] and a code of 991 to the final 12-month frequency variable (where applicable). For questions on drugs where respondents were asked to report their frequency of use in the past 12 months, a code of 91 (or 991) was assigned to the source variables pertaining to the preferred method of reporting the 12-month frequency (i.e., average number of days per week, average number of days per month, or total number of days used in the past 12 months), the average number of days per week, the average number of days per month, and total number of days used in the past 12 months.[77]

Similarly, if the edited recency of use indicated that the respondent had used the drug but not in the period of interest, edits at this step assigned a code of 93 to the 30-day frequency variable and a code of 93 (or 993) to the 12-month frequency variable and related source variables that were used to create the 12-month frequency.

6.2.4.3 Editing of 12-Month and 30-Day Frequency Variables When the Recency Was Indefinite

If the respondent was potentially a user in the period of interest (i.e., there was some question about when the respondent last used the drug) and the CAI program had skipped the 30-day or 12-month frequency questions, then the skipped variables retained a blank code. For example, if respondents reported lifetime use of a substance but they did not indicate a specific period for most recent use (Section 6.2.2.3), the frequency-of-use questions retained missing values. If the recency was imputed to indicate use in the past 30 days or past 12 months, the

[75] For hallucinogens, inhalants, and the psychotherapeutics, this meant that the respondent had never used any of the drugs in that category.

[76] For alcohol, this edit also applied to other 30-day variables, including the variable on the number of days in the past 30 days that respondents had five or more drinks in a single occasion.

[77] If a respondent was logically inferred not to have used a drug and the recency variable had been assigned a code of 81, the corresponding edited 30-day or 12-month frequency variables were assigned codes of 81 or 981, where applicable.

imputation procedures also replaced the missing data in these frequency-of-use variables with nonmissing values (Section 6.3).

Also, if a respondent reported last using marijuana more than 12 months ago, the CAI program skipped the questions pertaining to frequency of marijuana use in the past 12 months and past 30 days. However, reporting first use of marijuana at the respondent's current age would be inconsistent with the reported recency. As discussed in Section 6.2.2.4 and Table 6.2, the flag and impute edit rules assigned a value to the recency variable to indicate that this respondent was a user at some point in his or her lifetime, which could include use in the past 30 days or past 12 months. In this situation, the 12-month and 30-day marijuana frequency variables that had been skipped retained a blank value in case subsequent imputation might assign the respondent to a more recent category.

In addition, as discussed in Section 2.4.3, if respondents refused the lifetime gate question(s) and were skipped out of the 12-month and 30-day frequency questions (where applicable),[78] the edits at this step assigned refusal codes to the skipped frequency questions (i.e., the refusal was propagated). However, if respondents were skipped out of the 12-month and 30-day frequency questions because they answered the lifetime gate question(s) as "don't know," the edits retained codes of "blank" in the frequency variables, for the reasons given in Section 2.4.3.

6.2.4.4 Editing of the Frequency-of-Use Variables Based on Consistency Checks

Modules that contained both 12-month and 30-day frequency variables included consistency checks between these variables. A consistency check was triggered in situations where the number of days that respondents reported using the drug in the past 30 days exceeded the number of days that the respondent used in the past 12 months.

If respondents did not know how many days they used a drug in the past 30 days or refused to give an answer, they have been asked to give their best estimate of the number of days that they used. Respondents could estimate their 30-day frequency by choosing the category most likely to contain the number of days they used the drug: 1 or 2 days, 3 to 5 days, 6 to 9 days, 10 to 19 days, 20 to 29 days, or all 30 days. A consistency check also was triggered if the number of days that respondents reported using a drug in the past 12 months was lower than the minimum value for the number of days that respondents estimated using that drug in the past 30 days. For example, it would be inconsistent for a respondent to report using marijuana on 6 to 9 days in the past 30 days and also to report using it on fewer than 6 days in the past 12 months.

If the respondent revised either the 12-month or 30-day frequency data (or both) to make them consistent (i.e., such that the 12-month frequency was greater than or equal to the 30-day frequency following any updates done by the respondent), data from the consistency checks were taken as final. This included situations in which respondents resolved inconsistencies between their 12-month frequency and the minimum value for their estimated frequency of use in the past 30 days. If the 30-day frequency (or the minimum value for an estimated 30-day frequency) still was greater than the computed 12-month frequency despite a consistency check having been triggered, then the 12-month frequency was assigned a bad data code.

---

[78] For the tobacco variables through heroin, such a situation would occur if respondents initially refused the gate question and then refused again on follow-up.

These edits based on data from consistency checks also applied to data that respondents entered in the follow-up questions for the 12-month and 30-day frequencies for hallucinogens. For example, when respondents were asked follow-up questions about their 30-day frequency of use (Section 6.2.4.6), consistency checks existed between the 12-month hallucinogen frequency and answers to the follow-up questions for the hallucinogen 30-day frequency. Thus, if respondents were routed to the follow-up questions for the hallucinogen 30-day frequency and the resulting 30-day frequency continued to be greater than the 12-month frequency after respondents were prompted to resolve the inconsistency, the hallucinogen 12-month frequency was assigned a bad data code.

In addition, a consistency check was triggered in the alcohol module if respondents reported that they had five or more drinks in a single occasion on more days in the past 30 days than they reported for the number of days in which they drank any alcohol in that period. If respondents made their overall 30-day frequency of alcohol use and frequency of consumption of five or more drinks consistent with one another, then these answers were taken as final.

6.2.4.5 Editing of the Frequency-of-Use Variables in Response to Data Patterns Not Involving Recency, Parent/Child Data, or Binge Alcohol Use

Table B.5 in Appendix B lists detailed edits for the 12-month and 30-day frequency variables. The edits that are described pertain to data patterns that do not involve inconsistencies (1) between the frequency data and most recent use, (2) between frequency-of-use data for parent and child drugs, or (3) involving the frequency of binge alcohol use in the past 30 days. The edits in Table B.5 also pertain to any inconsistent reports that did not trigger a consistency check during the interview or situations in which a consistency check was triggered but respondents did not resolve the inconsistency.

For example, if the value for the 12-month frequency fell within the range of a respondent's estimate for the 30-day frequency (e.g., if a respondent reported using on 8 days in the past 12 months and on 6 to 9 days in the past 30 days), maximum and minimum values were created for the estimated 30-day frequency. In this example where a respondent reported use on 8 days in the past 12 months but estimated using the drug on "6 to 9" days in the past 30 days, use on 6 to 8 days (as opposed to 6 to 9 days) in the past 30 days would be consistent with the respondent's answer to the 12-month frequency. Information on the maximum and minimum possible number of days that a respondent could have used a drug in the past 30 days was used subsequently during statistical imputation to assign a final value to the 30-day frequency.

6.2.4.6 Editing of the Frequency-of-Use Variables in Response to Parent/Child Data Patterns for Cocaine, Pain Relievers, and Stimulants

Table B.6 in Appendix B lists detailed edits for the 12-month frequency variables for the parent/child pairs of cocaine and crack, pain relievers and OxyContin®, and stimulants and methamphetamine. This includes description of some special edits to the variable for the preferred way of reporting the 12-month frequency when respondents were nonmedical users of only OxyContin® or users of only methamphetamine (Section 6.2.3). Table B.6 also lists edits for the 30-day frequency variables for cocaine and crack. Because the hallucinogens module did not include questions on the frequency of use in the past 12 months or past 30 days for the child drugs LSD, PCP, and Ecstasy, inconsistencies in parent/child data for the frequency of use did not occur for hallucinogens.

For example, respondents could report use of a child drug on a number of days in the past 12 months that was greater than the number of days they reported using the parent drug in the past 12 months. In these situations, the higher value from the 12-month frequency for the child drug was assigned to the 12-month frequency for the parent drug. The source variables for the parent 12-month frequency also were edited to indicate the movement of data from the 12-month frequency of the child drug to the 12-month frequency for the parent drug.

Also, since 2003, the hallucinogens and stimulants modules have included follow-up questions for the 12-month frequency variables. Similar follow-up questions have been included in the pain relievers module since 2005. In the hallucinogens module, for example, these questions were asked if respondents originally reported that they last used any hallucinogen more than 12 months ago but subsequently reported more recent use of any hallucinogen, LSD, PCP, or Ecstasy. Respondents' original answer of use of the parent more than 12 months ago would cause them to be skipped out of the 12-month frequency-of-use questions. Therefore, when respondents gave some updated indication of use in the past 12 months, they were asked to fill in previously missing information about their frequency of use in the past 12 months. Similar logic was in place in the pain relievers and stimulants modules. For example, respondents were asked follow-up questions for their frequency of use of any pain relievers in the past 12 months if they originally reported that they last used any pain reliever more than 12 months ago but they subsequently indicated that they last used OxyContin® in the past 12 months, and they continued to be past year users when prompted to resolve the inconsistent data for their period of most recent use of any pain relievers and OxyContin®.

In addition to the 12-month frequency follow-up questions, the hallucinogens module had similar follow-up questions for the 30-day frequency of use. These questions were asked when respondents originally indicated that they last used any hallucinogen more than 30 days ago but subsequently reported that they last used any hallucinogen, LSD, PCP, or Ecstasy within the past 30 days. Thus, since 2003, data on the frequency of hallucinogen use in the past 30 days have been intended to be supplied by respondents (instead of through statistical imputation) if they subsequently indicated some hallucinogen use in the past 30 days but did not originally report using hallucinogens in the past 30 days.

6.2.4.7 Editing of the Frequency-of-Use and Related Variables Involving Binge Alcohol Use

Table B.7 in Appendix B includes edits related to the data for the frequency of binge alcohol use in the past 30 days. For example, question AL07 in the alcohol module (edited variable NODR30A) asked respondents to report the usual number of drinks that they consumed in a given day in the past 30 days. Although this is not a frequency variable per se, information from this variable was used to edit the 30-day frequency data for alcohol and the frequency data for binge alcohol use. For example, if a respondent reported having five or more drinks per occasion on *exactly* the same number of days that he or she reported drinking any alcohol in the past 30 days, then it would logically follow that the respondent's usual number of drinks per day had to have been five or more. If the respondent reported usually having fewer than five drinks on the days when he or she drank alcohol in the past 30 days, NODR30A was assigned a special code of 975 (AT LEAST 5 Logically assigned) to indicate usual consumption of at least five drinks.

Similarly, if a respondent drank on only 1 day in the past 30 days and reported having fewer than five drinks on that 1 day in question AL07, but the respondent answered question AL08 as "don't know" or "refused," it would logically follow that the respondent could not have had five drinks on any occasion in the past 30 days. When this occurred, the edited variable DR5DAY (corresponding to question AL08) was assigned a special code of 80 (NO OCCASIONS OF 5 OR MORE DRINKS IN THE PAST 30 DAYS Logically assigned) to denote that the respondent logically could be inferred not to have had five or more drinks on an occasion in the past 30 days. Consequently, the respondent could subsequently be defined as not being a binge alcohol user in the past 30 days.

6.2.4.8 Frequency of Smokeless Tobacco Use

Although a recoded recency variable SLTREC was created for any smokeless tobacco use, a variable was not created for the number of days that respondents used any smokeless tobacco in the past 30 days. If respondents reported use of both chewing tobacco and snuff in the past 30 days, it would have been possible for use of both smokeless tobacco types to have overlapped to varying degrees in the past 30 days. However, this degree of overlap was unknown. To create a recoded 30-day frequency of any smokeless tobacco use, assumptions would have been required (e.g., picking the maximum of the two) that could not have been confirmed from the data.

6.2.5 Incidence (Age at First Use, Month of First Use, and Year of First Use)

In all core modules except for pipe tobacco, respondents were asked how old they were when they first used the drug of interest. If respondents reported first using the drug within 1 year of their current age, they were asked to report the specific month and year when they first used, with the allowable years ranging from 2012 to 2014. If respondents reported first using the drug at their current age and their birth month was earlier than the interview month (i.e., they reached their current age in the same year that they were interviewed), the CAI program assumed that the first use of the drug occurred in the current year (i.e., 2014). These respondents were asked only for the month that they first used in the current year. The remaining respondents who first used a drug within 1 year of their current age could be routed to one of two possible questions on the specific year they first used. They then were routed to a question to report on the specific month that they first used the drug in the year they had reported previously.

Because the routing logic to the different versions of the MFU and YFU questions was mutually exclusive, a single, composite set of MFU and YFU variables was created from the individual unedited variables. In addition, if respondents indicated a specific year that they first used a drug, the final YFU variables for 2014 were recoded to replace unedited codes with values for the specific years (i.e., 2012 through 2014). If respondents confirmed that they first used a drug at their current age and were interviewed subsequent to their birthday in the current year, a code of "2014" was assigned to the YFU; this was done even if respondents did not know what month they first used in the current year, or if they refused to report what month they first used in the current year. If the MFU and YFU questions had been skipped because respondents first used the drug more than 1 year younger than their current ages, legitimate skip codes were assigned to the final MFU and YFU variables.

Since 2002, consistency checks have been included in the instrument if the values for the MFU and YFU were inconsistent with the AFU. Specifically, for recent initiates of a given drug, the CAI program calculated a second AFU use based on the MFU and YFU by comparing these data with the respondent's date of birth. This comparison was not done if the respondent reported first use of the drug in the same month that he or she was born; a unique AFU could not be determined from the MFU and YFU in these situations because it was not known whether the drug use occurred before or after the respondent's birthday. Similarly, a consistency check was not triggered if the respondent had missing data in either of the month or year questions, such as if the respondent knew the year when he or she first used a drug but did not know the MFU.

In remaining situations in which respondents provided complete data for the MFU and YFU, a consistency check was triggered if the MFU and YFU suggested that respondents initiated use of the drug at an earlier or a later age than what they had previously reported. For example, a consistency check was triggered if a 16-year-old respondent reported first using a drug at age 16 but then reported first using the drug in a month and year that would have meant the respondent was 15 years old when he or she first used the drug. No editing needed to be done if respondents indicated twice in a row that the AFU that was calculated from the MFU and YFU was correct. The CAI program updated the value for the AFU (e.g., AGE1STCG for cigarettes) to agree with the values for the MFU and YFU.

If respondents indicated at some point in the consistency check sequence that the value they had reported for their AFU (e.g., question CG04 for cigarettes) was correct, they had an opportunity to revise the values for their YFU and their MFU. If a consistency check was triggered between the AFU and data in the MFU and YFU, the MFU and YFU were updated with any year and month data that the respondent entered in the consistency checks (e.g., CGCC21 and CG221a for any cigarette use). These data were used in subsequent editing steps. Otherwise, the MFU and YFU data were picked up from the original source variables (e.g., CG04A through CG04D for any cigarette use) for use in editing, such as for the "flag and impute" edits that were described in Section 6.2.2.4 and Table 6.2.

Table B.8 in Appendix B lists edits pertaining to the incidence variables when consistency checks had been triggered because of inconsistent data between the AFU and the YFU and MFU. The default when a respondent did not resolve an inconsistency between the AFU and the MFU and YFU was to favor the AFU in subsequent editing decisions. Table B.9 presents information on additional edits involving the incidence variables other than for parent/child relationships, such as situations where the AFU was inconsistent with the respondent's current age. Table B.10 presents information on edits involving parent/child relationships in the incidence data. In addition to parent/child relationships that have been described previously for the cocaine, pain relievers, and stimulants modules, parent/child relationships for incidence included first use of any cigarette and initiation of daily cigarette use.

Also, recoded variables were created for respondents' ages when they first used any smokeless tobacco product (i.e., chewing tobacco or snuff), and the month and year when they first used, if applicable. If respondents had a missing value for one of the types of smokeless tobacco (i.e., "don't know," "refused," "bad data"), the missing value was retained in the recoded smokeless tobacco variables for the AFU, YFU, or MFU. For example, if a respondent had used both chewing tobacco and snuff and reported an age when he or she first used chewing tobacco but refused to report the age when he or she first used snuff, the respondent may have used snuff at a younger age than was reported for chewing tobacco. If the AFU did not have a missing value, the recoded MFU and YFU for smokeless tobacco subsequently were edited to be consistent with the AFU that was chosen. If respondents initiated use of both types of smokeless tobacco at the same age and were asked the month and year that they first used (i.e., the first use was within 1 year of their current age), the recoding procedures picked the earliest year. If they reported first using both types of smokeless tobacco in the same year, the recoding procedures picked the earliest month.

6.2.6 Editing of Noncore Methamphetamine Use Data

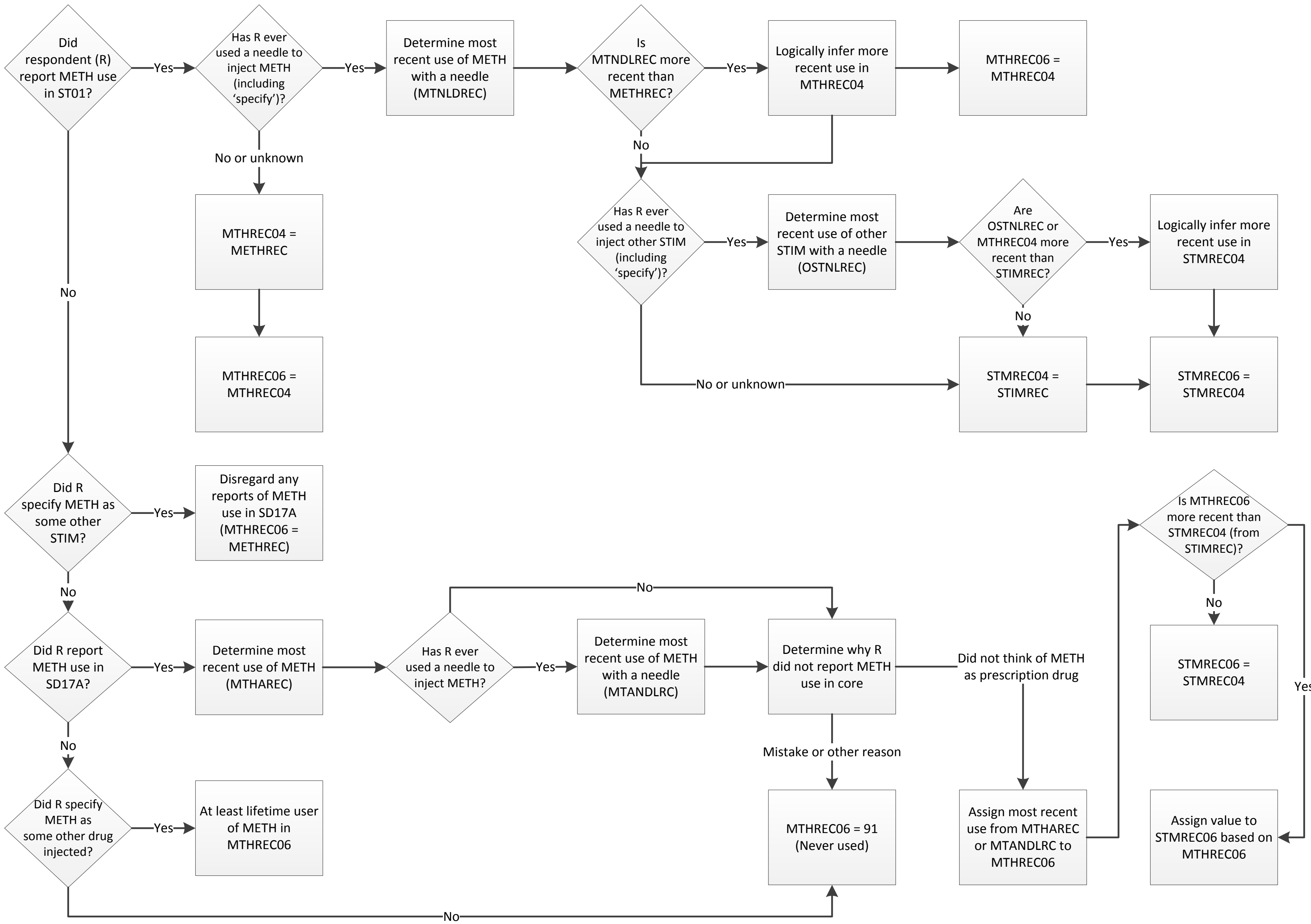
As noted in Section 6.1.3, questions have been included in the noncore special drugs module since 2005 to capture information from respondents who may have used methamphetamine but did not recognize it as a prescription drug and therefore did not report use in the core stimulants module. Additional follow-up items have been included since 2006 to identify those respondents who specifically did not report methamphetamine use in the core stimulants module because they did not consider methamphetamine to be a prescription drug. This section describes the editing procedures for the noncore methamphetamine data in the special drugs module that were used to create the CPN variables for methamphetamine and stimulants. Editing of methamphetamine variables from the special drugs module that was not relevant to creating the CPN variables is discussed in Section 7.4.1 in Chapter 7.

Figure 6.2 shows routing logic for the methamphetamine and stimulant variables in the special drugs module that were used to create the edited CPN variables for most recent use of methamphetamine and stimulants. The figure also includes information about decision making for creating the edited CPN recency variables.

Table 6.4 shows the mapping of questions in the noncore special drugs module to edited variables that were used to create the edited CPN recency variables for methamphetamine and stimulants. In particular, questions SD17A through SD18B module captured information about methamphetamine use from respondents who did not report methamphetamine use in the core stimulants module. Data from questions SD10A, SD10B, SD10C, and SD11 that were present in the special drugs module prior to 2005 also were used in creating the edited CPN variables. Unlike the questions that were described previously, SD10A and SD10B were asked if respondents reported lifetime use of methamphetamine in the core stimulants module.

The general issues that were described in Section 6.2 for the core recency variables also applied to the CPN variables. For example, if respondents reported lifetime methamphetamine use in question SD17A (edited variable MTHAMP) but did not know or refused to report in question SD17B when they last used it, the edited methamphetamine recency variable MTHAREC was assigned a code of 9 (Used at some point in the lifetime, LOGICALLY ASSIGNED).

Figure 6.2 Routing Logic and Editing Overview for Methamphetamine and Other Stimulant Questions in the Special Drugs Module



METH = methamphetamine; STIM = stimulants.

Table 6.4 Mapping of Noncore Methamphetamine Questions in the Special Drugs Module to Edited Variables

Question(s)	Edited Variable	Comments
SD10A	MTHNEEDL	Ever used a needle to inject methamphetamine; asked if respondents (Rs) reported methamphetamine use in the core stimulants module.
SD10B	MTNDLREC	Most recent use of methamphetamine with a needle, if SD10A was answered as "yes."
SD10C	OSTNEEDL	Ever used a needle to inject stimulants/any other stimulant; asked if Rs reported nonmedical use of stimulants in the core stimulants module. SD10C was worded as "any other stimulant" if lifetime methamphetamine use was reported and was worded as "any stimulant" if methamphetamine use had not been reported.
SD11	OSTNLREC	Most recent use of stimulants/other stimulants with a needle, if SD10C was answered as "yes."
SD17A	MTHAMP	Ever used methamphetamine; asked if Rs did not report lifetime use in the core stimulants module.
SD17B	MTHAREC	Most recent use of methamphetamine, if SD17A was answered as "yes."
SD18A	MTHANEDL	Ever used a needle to inject methamphetamine; asked if Rs did not report lifetime use in the core stimulants module.
SD18B	MTANDLRC	Most recent use of a needle to inject methamphetamine, if SD18A was answered as "yes."
SD17A1 SD17ALT	MTHEVCK	Consistency checks to determine which answer was correct: the report of lifetime nonuse of methamphetamine from the core stimulants module or the report of methamphetamine use in the special drugs module.
SD17A2	MTHNORSN	Reason for not previously reporting methamphetamine use in the core stimulants module; asked if Rs confirmed their use of methamphetamine.
SD17A2SP	MTHNOSP	"OTHER, Specify" response for not reporting methamphetamine use in the core stimulants module; asked if Rs reported "some other reason" for not reporting use.

6.2.6.1 Editing of the Methamphetamine Recency Variables in the Special Drugs Module

Table B.11 in Appendix B includes edits for the noncore methamphetamine and stimulant variables that were relevant for creating the edited CPN recency variables. The focus is on edits pertaining to most recent use of methamphetamine (SD17B) or most recent use of methamphetamine or stimulants with a needle (SD10B or SD18B for methamphetamine; SD11 for other stimulants). However, Table B.11 also includes details about editing of the needle use variables—including logical inference of use or nonuse of methamphetamine or other stimulants with a needle—to allow all of the issues for editing of these variables to be included in the same table. In particular, if respondents had never used a needle to inject methamphetamine or other stimulants, then the edited variables pertaining to most recent use of methamphetamine or stimulants with a needle did not contribute to the creation of the edited CPN recency variables.

In contrast, it was possible for respondents to report in the "OTHER, Specify" questions SD05A through SD05E pertaining to use of other drugs with a needle that they had injected methamphetamine or other stimulants at some point in their lifetime. These responses overruled any denial of methamphetamine use (or use of methamphetamine with a needle) or any denial of use of stimulants with a needle. In these situations, the general practice was to assign a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED), with the following exceptions:

- If respondents were lifetime users of methamphetamine in the core stimulants module but the most recent use of methamphetamine from the variable METHREC was more than 12 months ago, the respondent was logically inferred to have last used a needle to inject methamphetamine more than 12 months ago. MTNDLREC was assigned a code of 13 (More than 12 months ago LOGICALLY ASSIGNED).
- If respondents were routed to question SD17A because they did not report methamphetamine use in the core stimulants module and the most recent use of methamphetamine from MTHAREC (corresponding to SD17B) was more than 12 months ago, then MTANDLRC was assigned a code of 13 to indicate that the respondent logically had last injected methamphetamine more than 12 months ago.
- If respondents were logically inferred to have used a needle to inject other stimulants in their lifetime but the most recent use of stimulants from the variable STIMREC was more than 12 months ago, then OSTNLREC was assigned a code of 13 to indicate that the respondent logically had last used a needle to inject other stimulants more than 12 months ago.

6.2.6.2 Creation of the Edited CPN Recency Variables for Stimulants and Methamphetamine

In creating the CPN recency variables for methamphetamine and stimulants, the core recency variables METHREC and STIMREC were used as the "base" variables. Consequently, if data from the special drugs module did not indicate use or more recent use than was indicated in the core recency variables, then the CPN recency variables retained the information from the core variables.

The following edited CPN recency variables were relevant to determining most recent use of methamphetamine and stimulants if respondents had reported use of methamphetamine in the core stimulants module:

- MTHREC04: Most recent use of methamphetamine, based on the core recency variable METHREC and the variable MTNDLREC for most recent use of methamphetamine with a needle; and
- STMREC04: Most recent use of stimulants, based on the core recency variable STIMREC and the variable OSTNLREC for most recent use of (other) stimulants with a needle.

In these variables, the number "04" represented editing based on variables that were in the special drugs module in 2004 (i.e., prior to addition of the follow-up questions in 2005).

In creating MTHREC04 and STMREC04, MTHREC04 was initially set to the value from METHREC, and STMREC04 was initially set to the value from STIMREC. The following edits were implemented for MTHREC04 and STMREC04:

- If respondents were lifetime nonusers of methamphetamine in the core stimulants module and respondents did not specify in the special drugs module that methamphetamine was "some other drug" that they injected with a needle, then MTHREC04 retained the value of 91 (Never used) from METHREC. A similar edit applied to STMREC04 if respondents had reported never using stimulants nonmedically and did not specify injecting stimulants as "some other drug."
- If METHREC and STIMREC had been assigned a code of 81 (Never used; logically assigned) because respondents reported use of only OTCs in the stimulants module (Section 6.2.1.5), then MTHREC04 and STMREC04 retained a code of 81. However, this edit did not apply if respondents reported in the special drugs module that these were "some other drug" that they injected.
- Respondents were logically inferred to be more recent users in MTHREC04 and STMREC04 if the needle recency variables MTNDLREC and OSTNLREC indicated more recent use of these drugs with a needle than was indicated in the corresponding core recency variables METHREC and STIMREC. Codes of 11 (Used in the past 30 days LOGICALLY ASSIGNED) or 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED) were assigned to MTHREC04 and STMREC04 based on these edits.
- If METHREC indicated use more than 12 months ago or nonuse and MTNLDREC had been set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) because respondents specified use of methamphetamine as "some other drug" that they injected, then MTHREC04 was set to 9. A similar edit was applied to STMREC04. In addition, if MTHREC04 had been set to 9 because of these edits but STIMREC indicated use more than 12 months ago, then STMREC04 was set to 9.
- If METHREC indicated use more than 30 days ago but within the past 12 months and MTNLDREC had been set to a value of 9, then MTHREC04 was set to a value of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED). A similar edit was applied to STMREC04. In addition, if MTHREC04 had been set to 9 because of the edits that were described previously but STIMREC indicated use more than 30 days ago but within the past 12 months, then STMREC04 was set to 8.
- If MTHREC04 indicated more recent use of methamphetamine (including use of methamphetamine with a needle) than was indicated in STIMREC or OSTNLREC, then STMREC04 was assigned a value indicating more recent use.

The following edited CPN recency variables were the final CPN recency variables that were used in the subsequent imputation procedures that are described in Section 6.3.7:

- MTHREC06: Most recent use of methamphetamine, based on MTHREC04, MTHAREC, or MTANDLRC; and
- STMREC06: Most recent use of stimulants, based on STMREC04 and MTHREC06.

In these variables, the number "06" represented editing based on variables that have been included in the special drugs module since 2006. MTHREC06 was initially set to the value from MTHREC04 and STMREC06 was initially set to the value from STMREC04 to capture the results of any editing when respondents had reported lifetime use of methamphetamine in the core stimulants module.

As noted in Section 6.1.3, however, giving respondents a second opportunity to report methamphetamine use could bias the estimates if respondents who had made a mistake in answering the previous question about methamphetamine use in the core stimulants module could change their answer on follow-up to indicate use—something that is not done for other drugs in NSDUH. Rather, the aim of asking the follow-up methamphetamine questions if respondents had not previously reported use in the core stimulants module was to identify respondents who had not reported methamphetamine use in the context of questions about *prescription* stimulants. Therefore, if respondents who previously did not report methamphetamine use confirmed in the special drugs module that they were indeed users, they were asked why they had not reported methamphetamine use when they were asked about it earlier in the stimulants module. Respondents could indicate one of the following reasons why they had not previously reported methamphetamine use:

- the earlier question in the core stimulants module asked about prescription drugs, and they did not think of methamphetamine as a prescription drug;
- they made a mistake when they answered the earlier question about methamphetamine; or
- there was some other reason.

If respondents reported that there was some other reason why they had not previously reported methamphetamine use, they were asked to specify what this other reason was.

In creating MTHREC06, only those respondents who were routed to question SD17A and indicated that they did not think of methamphetamine as a prescription drug (or who specified something similar to that as their other reason for not previously reporting use) were counted as additional methamphetamine users. Otherwise, data from MTHAREC and MTANLREC were not used in creating MTHREC06 if respondents reported that they made a mistake in not previously reporting methamphetamine use in the core or the other reason they specified did not pertain to their not thinking of methamphetamine as a prescription drug. However, if respondents had specified using methamphetamine with a needle as "some other drug," then this was reflected in the edits that were mentioned previously for MTHREC04. In turn, these edits to MTHREC04 affected the final value in MTHREC06.

The following edits were implemented for MTHREC06 based on data from questions SD17A to SD17ASP. These edits that are described for MTHREC06 when respondents reported methamphetamine use assume that respondents had not reported methamphetamine use in the core stimulants module because they did not think of it as a prescription drug.

- If respondents were lifetime nonusers of methamphetamine in the core stimulants module, they indicated in SD17A that they never used methamphetamine, and they did not specify use of methamphetamine with a needle, then MTHREC06 retained the value of 91 (Never used) from METHREC.
- If MTHAREC had been set to a value of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) because respondents reported in SD17B that they had last used methamphetamine more than 30 days ago but within the past 12 months and MTANDLRC had a value of 9, then MTHREC06 was assigned a value of 8.
- If MTHAREC had been set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) because respondents reported in question SD17B that they last used methamphetamine more than 12 months ago but MTANDLRC had a value of 9, then MTHREC06 was assigned a value of 9.
- If the methamphetamine recency variables were consistent (i.e., MTHAREC indicated as recent or more recent use of methamphetamine than was indicated in MTANDLRC, or respondents never used a needle to inject methamphetamine), then the value from MTHAREC was assigned to MTHREC06.
- If MTHAREC indicated that respondents were logically inferred to be more recent users of methamphetamine based on indications of more recent use of methamphetamine with a needle in MTANDLRC (Table B.11), then MTHREC06 was assigned the corresponding code of 11 (Used in the past 30 days LOGICALLY ASSIGNED) or 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED) from MTHAREC.

The following edits were implemented for STMREC06 based on data from MTHREC06 according to the edits that were described previously:

- If MTHREC06 had been set to a value of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) and STMREC04 did not indicate use in the past 30 days, then STMREC06 was assigned a value of 8.
- If MTHREC06 had been set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) and STMREC04 did not indicate use in the past 12 months, then STMREC06 was assigned a value of 9.
- If MTHREC06 was consistent with STMREC04 (i.e., STMREC04 indicated as recent or more recent nonmedical use of stimulants than was indicated in MTHREC06, or MTHREC06 indicated that respondents never used methamphetamine), then the value from STMREC04 was assigned to STMREC06.
- If MTHREC06 and STMREC04 indicated that they last used methamphetamine or stimulants in a definite period (i.e., in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago) and MTHREC06 indicated more recent use of methamphetamine, then STMREC06 was assigned a code of 11

(Used in the past 30 days LOGICALLY ASSIGNED) or 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED).

If MTHREC06 indicated that respondents last used methamphetamine in a definite period but STMREC04 indicated nonuse or had a missing value, then STMREC06 was assigned a code of 11, 12, or 13 (Used more than 12 months ago LOGICALLY ASSIGNED) based on the most recent use from MTHREC06. This included situations in which STMREC04 retained a code of 81 (Never used; logically assigned) from STIMREC because respondents reported use of only OTCs in the stimulants module (Section 6.2.1.5). In this situation, the data from MTHREC06 (based on MTHAREC) indicating methamphetamine use overruled the logical inference that respondents were never lifetime nonmedical users of stimulants.

6.2.7 Editing of Lifetime Daily Cigarette Use Status

Lifetime users of cigarettes were asked in question CG15 whether they ever had a period when they smoked cigarettes every day for at least 30 days. Respondents were not asked question CG15 if they had already reported that they smoked cigarettes on all 30 days in the past 30 days.

Edits associated with determining whether respondents had ever smoked cigarettes daily for at least 30 days (edited variable CIGDLYMO) are presented in Table B.12 in Appendix B. In particular, if question CG15 had been skipped because respondents smoked cigarettes on all 30 days in the past 30 days, then CIGDLYMO was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern). A code of 5 was assigned rather than a legitimate skip code (Section 2.4.2) to indicate that it could be logically inferred from the skip logic that the respondent had a lifetime period of having smoked cigarettes daily for at least 30 days.

6.2.8 Additional Edits Applied During Imputation Processing

In addition to the logical edits applied as described previously, edits to selected variables were applied during the imputation process and are discussed in this section. In general, these edits affected only a few records. They were implemented mostly to resolve residual inconsistencies that prevented the determination of a valid interval for the assignment of the date of first use (Section 6.3.3.4).

6.2.8.1 "Other" Hallucinogens, "Other" Pain Relievers, and "Other" Stimulants Variables

For respondents who were known to have never used "other" hallucinogens, "other" pain relievers, and "other" stimulants, certain logical deductions could be made regarding the relationship between the parent drug data and the child drug data if all the necessary conditions that also are described were met.[79] Note that these edits also could have been applied to respondents who were imputed to lifetime nonuse of the "other" variable.

1. If the respondent was known never to have used "other" hallucinogens, the overall hallucinogens recency was missing, and none of the recencies for the child drugs

[79] The creation of these "other" indicators is described in Section 6.1.2.

associated with hallucinogens were missing, then the overall hallucinogens recency was assigned to the most recent of its child drug recencies. This also was applied for pain relievers and stimulants.

2. If the respondent was known never to have used "other" hallucinogens, the overall hallucinogens recency was past month, one of the child recencies was past year (where past month vs. not past month use could not be determined), and no other child recency was past month, then the child recency that was past year (where past month vs. not past month use could not be determined) was edited to past month.
3. If the respondent was known never to have used "other" hallucinogens, the parent AFU value was nonmissing, only one child AFU value was missing, and the minimum of the nonmissing child AFU values was greater than the parent AFU value, then the missing child AFU value was edited to the parent AFU value.
4. If the respondent was known never to have used "other" hallucinogens, the parent AFU value was nonmissing, only one child AFU value was missing, the minimum of the nonmissing child AFU values was equal to the parent AFU value, and the earliest of the nonmissing child months and years of first use was later than the parent MFU and YFU, then the missing child AFU value was edited to the parent AFU value.[80]

6.2.8.2 Respondents Imputed to Lifetime Use for Child Drug Variables

As discussed in Section 6.3, the first imputation set consisted of the lifetime drug use measures. The results of these imputations could restrict the range of plausible values for other drug use measures, and, therefore, based on this additional information obtained from the imputations, certain editing rules that were applied to the unedited recency and frequency data had to be reapplied. The list of these edits follows:

1. If the parent drug recency of use was known to be lifetime but not past year, and the respondent was imputed to lifetime use of the child drug(s), then the child drug recency was set to lifetime but not past year. This was done because the respondent could not have used the child drug more recently than the parent drug.
2. If the respondent used the parent drug on exactly 1 day in the past 12 months, and the respondent was imputed to lifetime use of the child drug, then the child drug recency of use was set equal to the parent drug recency of use, and the child drug 12-month frequency of use was set to 1 day. This was done because the respondent could not have used the child drug on any days when the parent drug was not used, so the recencies and frequencies cannot differ. This edit only applied to OxyContin®, methamphetamine, and crack, which are the only child drugs with 12-month frequencies.
3. If the parent drug incidence data indicated a date of first use in the past year, the parent drug recency of use was past year but not past month, and the respondent was imputed to lifetime use of the child drug(s), then the recency of use for the child drug was set to past year but not past month. This was done because the respondent could

[80] These cases occur rarely, so they are handled on a case-by-case basis. The procedures do not automatically apply this edit. They flag cases like these for further examination.

not have used the child drug more recently than the parent drug (eliminating the possibility of past month recency), and the respondent also could not have started using the child drug before the parent drug (eliminating the possibility of lifetime but not past year recency).

4. Similarly, if the parent drug incidence data indicated a date of first use in the past year, the parent drug recency of use was past month, and the respondent was imputed to lifetime use of the child drug(s), then the recency of use for the child drug was set to past year (whether the respondent had used in the past month could not be determined). This was done because the respondent could not have started using the child drug before the parent drug (eliminating the possibility of lifetime but not past year recency).

6.2.8.3 Age-at-First-Use Variables

The edits that are described in this section were implemented in preparation for the imputation process. These edits are designed to reduce processing time and to set the imputation indicators properly (i.e., to "logically assigned" instead of to "statistically imputed"). As with the other edits that were described previously in Section 6.2.8, these edits affected only a small number of records.

1. If the parent AFU value was missing and the minimum of the child AFU values was 3 years, then the parent AFU value was edited to 3 years. This was done because respondents with AFU values of less than 3 years were ineligible to be donors (Section 6.3.3.1).[81] This edit applied to all parent AFU variables: cigarettes, overall hallucinogens, overall pain relievers, overall stimulants, and cocaine.
2. If the parent AFU was equal to the respondent's current age, all missing child AFU values were edited to the same age. This edit applied to all child AFU variables: daily cigarettes, LSD, PCP, Ecstasy, OxyContin®, methamphetamine, and crack.
3. If the parent AFU was equal to 1 less than the respondent's current age, the child recency[82] was lifetime but not past year (or, for cigarettes, past 3 years but not past year), and the child AFU value was missing, then the child AFU value was assigned to 1 less than the respondent's current age. This was done because the child AFU cannot be less than AGE − 1, because the parent AFU is AGE − 1, and the respondent could not have begun using a child drug before using the parent drug. It also cannot be greater than AGE − 1, because the child drug recency of lifetime but not past year indicates that the respondent did not use the drug while at his or her current age (because he or she did not use the drug at all in the past year). If the respondent did not use the drug at all in the past year, then he or she could not have begun using the drug in the past year. Because the child AFU cannot be less than AGE − 1 or greater

[81] This could be considered to be an imputation rule rather than an editing rule. Nevertheless, once the decision is made to prevent those with age-at-first-use values of less than 3 years from being donors, it is clear that cases like these do not require any sort of stochastic imputation.

[82] Because there was no recency question associated with daily cigarettes, the overall cigarette recency was used instead.

than AGE − 1, it must be equal to AGE − 1. This edit also applied to all child AFU variables.

4. If the age at first cigarette use was equal to AGE − 3, cigarette recency was lifetime but not past 3 years, and age at first daily cigarette use was missing, then age at first daily cigarette use was assigned to AGE − 3. The logic is similar to the above edit: the age at first cigarette use precludes the possibility that the age at first daily cigarette use was less than AGE − 3, and the cigarette recency precludes the possibility that the age at first daily cigarette use was greater than AGE − 3.

6.2.9 Edits for Drug Variables that Do Not Undergo Imputation

The following tobacco variables were edited but did not undergo further imputation:

- for adolescents aged 12 to 17 who had never smoked a cigarette: their likelihood of smoking a cigarette if their friends offered them one or of smoking a cigarette in the next 12 months;
- lifetime smoking of 100 or more cigarettes;
- usual brands of cigarettes, snuff, chewing tobacco, or cigars that respondents used in the past 30 days;
- among respondents who smoked cigarettes in the past 30 days:
  - the average number of cigarettes that they smoked per day;
  - the type of cigarette they smoked (light, ultralight, medium, or full flavor);
  - whether the brand of cigarettes that was smoked most often was a menthol cigarette;
  - for respondents who usually smoked Marlboro cigarettes in the past 30 days, the length of the Marlboro cigarettes that they smoked (shorts, regulars or king-sized, or 100s); and
  - whether respondents smoked a "roll-your-own" cigarette in the past 30 days.

6.2.9.1 Editing of Likelihood Variables for Adolescent Nonsmokers

If respondents were aged 12 to 17 and had never smoked a cigarette, they were routed to questions CG02 and CG03. Question CG02 (edited variable CIGOFRSM) asked adolescents if they would smoke a cigarette if one of their friends offered them a cigarette. Question CG03 (edited variable CIGWILYR) asked adolescents how likely they thought they would be to smoke a cigarette in the next 12 months.

No editing was done to CIGOFRSM and CIGWILYR if adolescents were routed to the corresponding questions because they never smoked cigarettes. Otherwise, a legitimate skip code of 99 (Section 2.4.2) was assigned if the questions had been skipped because (1) respondents were aged 18 or older or (2) they were aged 12 to 17 but they had already smoked a cigarette in their lifetime.

6.2.9.2 Editing of Lifetime Smoking of 100 or More Cigarettes

Question CG16A (edited variable CIG100LF) asked cigarette smokers whether they had ever smoked 100 or more cigarettes in their lifetime. Minimal editing was done to CIG100LF. As indicated in Table B.12, respondents were not asked this question if their answers to previous questions indicated that they had smoked 100 or more cigarettes. For example, respondents who smoked cigarettes in the past 30 days were asked in question CG07 to report the number of days that they smoked cigarettes in the past 30 days. Respondents who smoked cigarettes on more than 1 day in the past 30 days were asked in question CG08 to report the usual number of cigarettes that they smoked on those days when they smoked cigarettes. Question CG08 was a categorical variable that gave ranges of numbers of cigarettes that were smoked per day and, where relevant, the equivalent number of packs of cigarettes (e.g., 16 to 25 cigarettes per day, or about 1 pack).

If the product of the number of days that respondents smoked cigarettes and the lower bound of the range was 100 or greater, then question CG16A was skipped. For example, if a respondent smoked cigarettes on all 30 days in the past 30 days and usually smoked about a pack a day (i.e., 16 to 25 cigarettes per day), then a conservative estimate of the number of cigarettes that he or she smoked in the past 30 days (i.e., based on the lower bound) would be $30 \times 16 = 480$.

In this situation, the respondent would not be asked question CG16A because the number of cigarettes that the respondent logically smoked in the past 30 days was greater than 100. As indicated in Table B.12, if question CG16A had been skipped because had been skipped because respondents the quantity and frequency of cigarette use in the past 30 days indicated that respondents had smoked 100 or more cigarettes, then CIG100LF was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).

6.2.9.3 Editing of Tobacco Brand Variables

As noted in Section 2.3.1.2, the CAI instrument included questions to identify the specific brands of tobacco that were used most commonly by respondents who reported use in the past 30 days of cigarettes, chewing tobacco, snuff, or cigars. Respondents could choose from a list of brands or they could indicate use of "a brand not on this list." Procedures for coding the "OTHER, Specify" data when respondents reported use of a brand not on the list were described in Section 2.3.1.2.

The coding approach that was described in Section 2.3.1.2 also applied to situations where respondents chose tobacco brands from the lists of brands that they were shown in the questions. For cigarettes, for example, codes of 101 through 126 were used for Basic through Winston, corresponding to the brands that had been listed in question CG11 prior to the 2005 survey.[83] Cigarette brand codes of 127 through 160 were used for Alpine through True,

[83] Reference is made to response choices in 2004 or earlier because new response choices were added in 2005. Prior to 2005, for example, the first category in question CG11 was Basic. American Spirit was added to question CG11 in 2005, and this became category 1. To preserve continuity with the codes prior to 2005, these codes were not renumbered in 2005 to reflect any changes to the tobacco brand questions.

corresponding to the brands listed in question CG11A prior to the 2005 survey.[84] Thus, the edited variable CIG30BRN for the brand of cigarettes that respondents smoked most often in the past 30 days could be assigned a code of 101 if respondents reported in question CG11 that they usually smoked Basic cigarettes or if they specified Basic cigarettes as "a brand not on this list."

If respondents entered a brand from an available listing, they were asked to confirm their answer. If they confirmed their answer, they were asked no further questions about the brand they used for that particular type of tobacco. However, if respondents indicated that their previous answer was not correct, they were routed back through the series of 30-day brand questions for that type of tobacco. Thus, respondents had the opportunity to make corrections in situations where they may have miskeyed a number, such as if they keyed the number immediately above or below the number of the brand they meant to choose. For each type of tobacco that respondents reported using in the past 30 days, they were allowed to make corrections up to a total of three times. Respondents exited the loop once they confirmed an answer or specified use of a brand not on the list. Respondents also exited the loop if they answered "don't know" or "refused" when asked to confirm their answer.

Because of this routing logic, the brands that respondents confirmed that they used most often in the past 30 days were assigned to the edited variables CIG30BRN (for cigarettes), SNF30BRN (for snuff), CHW30BRN (for chewing tobacco), and CGR30BRN (for cigars). If respondents were rerouted through the series of questions and confirmed their answer on their second or third pass through the questions, the final tobacco brand coding procedures retained the final answer that respondents confirmed and disregarded whatever previous answers the respondent had given but did not confirm. Respondents who answered "don't know" or "refused" when asked to confirm what brand they used were assigned that corresponding code to the final brand variable for that type of tobacco. If respondents did not confirm what brand of a given tobacco type they used most often in the past 30 days despite three passes through the series of questions, a final code of 9000 to the edited variable for that tobacco type's brand. The code of 9000 indicated that these respondents did not confirm their brand despite three opportunities to do so.

As noted in Section 2.3.1.2, respondents sometimes specified that the brand of chewing tobacco they used most often was actually a snuff brand, or vice versa. Respondents also could specify that the cigarette brand that they smoked most often was actually a brand of little cigars. No editing was done to the codes for tobacco brands that applied to a different type of tobacco. For example, if respondents specified a little cigar brand as the brand of cigarettes that they smoked most often in the past 30 days, the edited variable CIG30BRN retained a code in the 400 or 4000 series for cigars. Continuing this example, if respondents did not report use of cigars in the past 30 days, no editing was done to the data for cigars to indicate use of cigars in the past 30 days in the edited recency variable CIGARREC or to assign the cigar brand code from cigarettes to CGR30BRN.

---

[84] The numbering of codes corresponding to responses in question CG11A started with 127 instead of 128 because prior to 2005, the response option of 27 in question CG11 meant "a brand not on this list." This response option was simply a toggle to question CG11A. Because the coding of brands resumed at 127 for brands that were listed in question CG11A, there was no break in the codes.

In addition, a recoded smokeless tobacco brand variable SLT30BRN was created because confusion sometimes existed in terms of what constituted chewing tobacco and snuff. If respondents reported use of both snuff and chewing tobacco in the past 30 days, they were asked to indicate which they had used most often. A final code was assigned to SLT30BRN according to the answer to this question. For example, if a respondent reported that the brand in the chewing tobacco section was the brand that he or she used most often, but this brand of "chewing tobacco" was really a snuff brand, SLT30BRN indicated that the respondent used a particular snuff brand most often in the past 30 days, even though this response came from the chewing tobacco brand questions.

In addition, the precoded response options in the questions for cigarette, chewing tobacco, and cigar brands since 2005 have included more prevalent brands that were mentioned in "OTHER, Specify" data from prior years. Since 2005, response options also are no longer present for less prevalent brands from 1999 to 2004. Documentation of these changes since 2005 is provided in the report on general principles and procedures for editing drug use data in the 2011 NSDUH MRB (Kroutil, Handley, & Bradshaw, 2013).

These changes in 2005 did not affect the creation of the edited tobacco brand variables. Nevertheless, analysts are advised that this change could affect analyses comparing trends for certain brands prior to 2005 and from 2005 onward. For example, significant differences in prevalence between some tobacco brand estimates prior to 2005 and after 2005 could occur if respondents prior to 2005 needed to type in the brand as "a brand not on this list" but they could choose it from a list in 2005 and beyond.

6.2.9.4 Editing of Miscellaneous Cigarette Use Variables for the Past 30 Days

As noted previously, respondents who reported that they smoked cigarettes in the 30 days prior to the interview were asked a series of additional questions about their cigarette use in that period (i.e., other than the brand of cigarettes that they smoked most often). These questions covered the following topics:

- the average number of cigarettes that they smoked per day (Section 6.2.9.4.1);
- the type of cigarette they smoked (light, ultralight, medium, or full flavor) and whether the brand of cigarettes that respondents smoked most often was a menthol cigarette (Section 6.2.9.4.2);
- for respondents who usually smoked Marlboro cigarettes in the past 30 days, the length of the Marlboro cigarettes that they smoked (shorts, regulars or king-sized, or 100s) (Section 6.2.9.4.3); and
- whether respondents smoked a "roll-your-own" cigarette in the past 30 days (Section 6.2.9.4.4).

6.2.9.4.1 Average Number of Cigarettes Smoked Per Day

Respondents who smoked cigarettes in the past 30 days were asked one of two possible questions regarding the number of cigarettes they smoked per day. Respondents who smoked on only 1 day in the past 30 days were asked to report the number of cigarettes they smoked on that

1 day. Respondents who smoked on more than 1 day (or who estimated the number of days they smoked in the past 30 days) were asked to report the average number of cigarettes they smoked per day. A single, composite variable (CIG30AV) was created from these two questions using data from whatever question the respondents were asked. No further editing was done to the data from these two questions.

6.2.9.4.2 Type of Cigarettes Including Menthol Cigarettes

As for the questions about the number of cigarettes that were smoked per day, past month cigarette users were routed to one of two possible questions about the type of cigarette they usually smoked in the past 30 days (CGTAR1 or CGTAR2). Since 2005, these questions have included a new response option for respondents who smoked "mediums" most often in the past 30 days. Review of data from quarter 1 of 2005 indicated that the addition of this new level for mediums affected the distribution of responses in CGTAR1 and CGTAR2 relative to the distribution in 2004, such that the data would not be comparable between 2004 and 2005. Therefore, the name of the edited variable corresponding to CGTAR1 and CGTAR2 has been CIG30TPE since 2005; prior to 2005, this variable was called CIG30TYP.

Past month cigarette users also were routed to one of two possible questions regarding whether the brand of cigarette they usually smoked was menthol (CGMENTH1 or CGMENTH2). Routing to CGTAR1 or CGTAR2 and to CGMENTH1 or CGMENTH2 was mutually exclusive for respondents who smoked cigarettes in the past 30 days (i.e., respondents were routed to one or the other question in a set but not both). Therefore, composite variables were created for the cigarette type (CIG30TPE) and whether the cigarette brand that respondents smoked most often was menthol (CIG30MEN).

No attempt was made to edit CIG30TPE or CIG30MEN for consistency with the cigarette brand from the variable CIG30BRN. In developing these items, instrument development staff consulted with tobacco research experts regarding which brands offered or did not offer menthol, light, or ultralight varieties. No conclusive information was obtained. As was discussed in Section 2.3.1.2, the "OTHER, Specify" data for tobacco brands did not capture information for particulars such as regular or menthol forms, or light, ultralight, or full-flavor varieties. Therefore, in situations where respondents specified that level of detail regarding the brand of cigarette that they smoked most often in the past 30 days, that information was not used to edit CIG30TPE or CIG30MEN.

6.2.9.4.3 Type of Marlboro Cigarettes

Respondents who reported in questions CG11, RCG11, or RRCG11 that they usually smoked Marlboro cigarettes in the past 30 days (and who confirmed this report of smoking Marlboro cigarettes) have been asked in question CGLNTH since 2005 about the length of the Marlboro cigarettes they smoked most often in the past 30 days: (1) shorts, (2) regulars or king-sized, or (3) 100s. The edited variable corresponding to CGLNTH was CIG30MLN. CIG30MLN was assigned a legitimate skip code of 99 (Section 2.4.2) if the cigarette brand variable CIG30BRN did not have a missing value and did not indicate that respondents smoked Marlboro cigarettes most often in the past 30 days; this included situations in which respondents reported that they smoked "a brand not on this list" and did not specify that they smoked Marlboro cigarettes.[85] Because respondents who reported that they smoked some other brand of cigarettes in the past 30 days were not asked CGLNTH, CIG30MLN retained values of 98 (blank) for any respondents who specified that the "other" brand of cigarettes was Marlboro; however, CIG30MLN was blank for this reason for fewer than 25 of approximately 6,100 respondents in 2014 who reported that they smoked Marlboro cigarettes most often in the past 30 days. In addition, CIG30MLN had missing values if respondents who smoked cigarettes in the past 30 days did not know or refused to report at the outset what cigarette brand they smoked most often in the past 30 days, or if they failed to confirm the brand that they smoked most often.

6.2.9.4.4 "Roll-Your-Own" Cigarettes

The cigarette section also included a question (CG14) about whether respondents smoked part or all of a "roll-your-own" cigarette in the past 30 days. The edited variable CIG30ROL corresponded to this question. The cigarette brand question CG11A[86] included response categories for two roll-your-own brands of cigarette tobacco. Respondents who chose either of these roll-your-own brands were skipped out of question CG14; by choosing a roll-your-own brand from the list of cigarette brands, these respondents already had indicated that they had smoked a roll-your-own cigarette in the past 30 days. Therefore, if question CG14 had been skipped and the cigarette brand was one of the roll-your-own brands from CG11a, a code of 5 (Yes LOGICALLY ASSIGNED [from skip pattern]) was assigned to the edited variable CIG30ROL.

However, respondents could specify a cigarette brand that was not on the list in questions CG11 and CG11A and then specify a roll-your-own brand in question CG12. In this situation, respondents were routed to question CG14. If respondents specified a roll-your-own brand and CG14 was already answered as "yes," no further editing needed to be done. If CG14 was not answered as "yes" when a respondent had specified a roll-your-own cigarette brand, the respondent's original answer was replaced with a code of 3 (Yes LOGICALLY ASSIGNED). This code of 3 signified to analysts that the respondent's original answer in question CG14 had been overwritten to make the roll-your-own data consistent with the cigarette brand information that was recorded in the variable CIG30BRN. However, this edit was implemented in 2014 for fewer than 10 respondents out of approximately 14,000 who reported cigarette use in the past 30 days.

6.3 Imputation of the Core Drug Use Variables

The predictive mean neighborhood (PMN) imputation methodology used in the imputation of drug variables beginning in 1999 was applied in a similar manner to the 2014

[85] This assignment of legitimate skip codes included situations in which the "other" cigarette brand was for a different type of tobacco (e.g., if respondents specified that the brand of "cigarettes" they smoked most often was a cigarillo or little cigar) or if respondents reported smoking a brand of cigarettes "not on this list" and they had missing values in their "OTHER, Specify" data. In this latter situation, CIG30BRN continued to have a code of 1999 (Cigarette; brand otherwise unspecified), and it was inferred in the editing of CIG30MLN that this otherwise unspecified cigarette brand was not Marlboro.

[86] Respondents could be routed to questions RCG11, RRCG11, RCG11A, RRCG11A, RCG12, or RRCG12 if they cycled through the cigarette brand questions more than once. For brevity, however, reference here is limited to the first set of cigarette brand questions: CG11, CG11A, or CG12.

National Survey on Drug Use and Health (NSDUH) drug data. Consistent with prior years, the drug use measures collected in the 2014 NSDUH included lifetime usage, recency of use, frequency of use in the past 12 months, frequency of use in the past month, and AFU, MFU, and YFU. However, depending on the drug in question, only a subset of these measures were collected and imputed.

Table 6.5 summarizes the drugs and drug use measures that were imputed. This table also indicates how these measures were segregated into units referred to as imputation sets. See Section 3.4 for more information on imputation sets.

Table 6.5 Drugs and Drug Use Measures, Imputation Sets

	Drug Use Measure					
Drug	**Lifetime Usage**	**Recency of Use**	**12-Month Frequency of Use**	**30-Day Frequency of Use**	**Age at First Use**	**Age at First Daily Use**
Cigarettes	NM	Set 2 (12-Month Frequency N/A)			Set 3	Sets 4, 5*
Smokeless Tobacco[1]	Set 1	Set 6 (12-Month Frequency N/A)			Set 7	N/A
Cigars		Set 8 (12-Month Frequency N/A)			Set 9	
Pipes		Set 10 (12-Month and 30-Day Frequency N/A)			N/A	
Alcohol		Set 11[5]			Set 12	
Inhalants		Set 13			Set 14	
Marijuana		Set 15			Set 16	
Hallucinogens[2]		Set 17			Set 18	
Pain Relievers[3]		Set 19 (30-Day Frequency N/A)			Set 20	
Tranquilizers		Set 21 (30-Day Frequency N/A)			Set 22	
Stimulants[4]		Set 23 (30-Day Frequency N/A)			Set 24	
Sedatives		Set 25 (30-Day Frequency N/A)			Set 26	
Cocaine and Crack		Set 27			Set 28	
Heroin		Set 29			Set 30	
Core-Plus-Noncore Stimulants and Methamphetamine	Set 31	Set 32 (12-Month and 30-Day Frequency N/A)			N/A	

N/A = not applicable; NM = never missing. Lifetime cigarette use is used to define a unit respondent and is therefore never missing.
*Prior to imputing age at first daily cigarette use, lifetime daily cigarette use must first be imputed.
[1] Includes chewing tobacco and snuff.
[2] Includes LSD, PCP, and Ecstasy.
[3] Includes OxyContin®.
[4] Includes methamphetamine.
[5] Includes binge drinking frequency.

Because there are numerous sets, some of the set-specific descriptions are explained as deviations from the procedures applied to an earlier set. For example, imputation set 2 is described in detail in Section 6.3.2, and the rest of the recency and frequency sets are described in Section 6.3.5 as deviations from the procedures described in Section 6.3.2.

Because drug use was highly correlated with age, and to facilitate more timely implementation of the imputation procedures, the model building and final assignment of imputed values for all drug use variables were performed separately within three distinct age groups: 12 to 17, 18 to 25, and 26 or older.

6.3.1 Lifetime Drug Use (Imputation Set 1)

The lifetime drug use variables were imputed using the single response propensity (RP)/multiple prediction (PRD) type of PMN, as outlined in Section 3.4.3. In general, the response rates for lifetime drug use variables were very high with less than 1 percent of cases requiring imputation. These high response rates were observed, in part, because of the usable case rule that requires that a respondent answer "yes" or "no" to the question on lifetime use of cigarettes and "yes" or "no" to at least nine additional lifetime use questions.

Because the single RP/multiple PRD type of PMN was used for the lifetime usage imputations, decisions had to be made on the order in which to fit the PRD models. Drugs later in the sequence would have more covariates in their models, because drugs earlier in the sequence were used as covariates after provisional imputation. The order in which the lifetime indicators of use were imputed is shown in Table 6.5, with the exception of lifetime cigarette use.[87] The lifetime use or nonuse of cigarettes was used to define a unit respondent for the NSDUH and, therefore, did not contain any missing values.

6.3.1.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for imputation set 1 was the preliminary analysis weight, PANALWT. As with the 1999–2013 surveys, the 2014 survey implemented automatic routing of the respondent through the questionnaire based on the respondent's answers, thereby skipping over (i.e., not asking the respondents) specific questions. Within each drug module, one (e.g., for marijuana) or multiple (e.g., for hallucinogens) questions were asked in order to establish whether the respondent had ever used the drug in question during his or her lifetime. For more information on gate questions and multiple gate questions, see section 6.2.1.

For an individual to be considered a lifetime use item respondent, he or she must have complete data for all of the drug module gate questions: cigarettes; cigars; chewing tobacco; snuff; pipes; alcohol; marijuana; cocaine; crack; heroin; inhalants; LSD; PCP; Ecstasy; hallucinogens other than LSD, PCP, and Ecstasy; OxyContin®; pain relievers other than OxyContin®; tranquilizers; methamphetamine; stimulants other than methamphetamine; and sedatives. See Table D.3 in Appendix D for details of the covariates used in the RP models for these variables.

6.3.1.2 First Prediction Step (Lifetime Smokeless Tobacco Use)

Many respondents who indicated lifetime use of smokeless tobacco seemed to be confused regarding the difference between chewing tobacco (chew) and snuff, as was

[87] See Section 3.4.2 for a brief discussion of how order is determined for imputation sets that use the multiple RP/multiple PRD or single RP/multiple PRD type of PMN.

demonstrated by their responses to questions regarding specific brands. For example, many respondents who indicated use of chewing tobacco entered a snuff brand, such as Copenhagen™, when asked about the specific brand of chew they used. As a result, one model for smokeless tobacco (a combination of the chew and snuff responses) was fitted, rather than individual models for chew and snuff. The probability of lifetime smokeless tobacco use was modeled for item respondents within each age group, using the nonresponse-adjusted weights. SUDAAN's RLOGIST procedure was used to perform dichotomous logistic regression[88] to determine the parameter estimates and probability of use for both respondents and nonrespondents.

6.3.1.3 First Provisional Hot-Deck Step (Lifetime Smokeless Tobacco Use)

In order to use lifetime usage of a given drug as a covariate for a drug later in the sequence, it was necessary to create temporary imputed values in cases where the original lifetime usage indicator was missing. Lifetime indicators for both chew and snuff were used as covariates for later models, so it was necessary to create these provisional values. In the first provisional hot-deck step, matching was done on a single predicted mean from the PRD step, but missing values for both chew and snuff were replaced with the values from a donor within this neighborhood.

If possible, donors and recipients were required to be from states with the same level of smokeless tobacco usage (state rank[89]), where the level of usage was defined in terms of the weighted proportion of a given state's residents who were lifetime users of the drug.[90] An additional likeness constraint required the donor to match the recipient on any nonmissing lifetime use indicators for child drugs. For example, if the lifetime use indicator for overall smokeless tobacco was missing, but the recipient was known to be a lifetime nonuser of snuff, then the donor must also have been a lifetime nonuser of snuff. If insufficient donors were available within these constraints, they were loosened in the following order: (1) the delta constraint was removed, and (2) both the state-rank and child lifetime drug indicator constraints were removed, and the delta constraint was reapplied.

No logical constraints were placed on the neighborhoods for any of the lifetime usage indicators. Even in the case of smokeless tobacco where more than one substance was associated with a single predicted mean, leading to a multivariate assignment of provisional imputed values, no logical constraints were necessary.

6.3.1.4 Analogous Prediction and Provisional Hot-Deck Steps for Remaining Drugs

PRD models and provisional hot-deck steps were completed in a manner similar to that described above for cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine and crack, and heroin, with the following deviations:

---

[88] SAS®-callable SUDAAN® was used to fit all dichotomous and polytomous logistic regression models. Details about the logistic regression model and additional references can be found in RTI International (2013). SAS software is a registered trademark of SAS Institute Inc. SUDAAN is a registered trademark of Research Triangle Institute.

[89] See Section 3.5 for a general discussion of state-rank variables.

[90] Those with a missing lifetime use indicator for the drug were treated as lifetime nonusers.

- For cigars, pipes, alcohol, inhalants, marijuana, tranquilizers, and sedatives, only one substance was associated with the predicted mean from the modeling stage. In these cases, the donor directly supplied the overall drug use value rather than providing values for child drugs that were then combined into a final usage measure as was the case for smokeless tobacco.
- Because cocaine and crack were in two separate back-to-back modules in the 2014 NSDUH questionnaire, separate models were fitted for the two substances. However, crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood with multivariate matching.[91] This was true regardless of whether the item nonrespondent was missing only crack, only cocaine, or both crack and cocaine. Once the neighborhood was defined, missing values for crack and/or cocaine were replaced with the values from one donor within this neighborhood.
- For hallucinogens, pain relievers, and stimulants, predicted probabilities were calculated for the parent drugs, and these probabilities were used to determine neighborhoods for each group of drugs. Lifetime usage indicators were assigned for LSD, PCP, Ecstasy, and "other" hallucinogens; OxyContin® and "other" pain relievers; and methamphetamine and "other" stimulants. Lifetime usage indicators for the parent drugs were created later by combining the constituent parts, including the "other" group of substances.
- Heroin did not undergo a provisional imputation step, because it was the last variable in the imputation set.

6.3.1.5 Final Hot-Deck Step

Tables E.20 through E.23 in Appendix E provide details on the final hot-deck step for the lifetime use indicators. Although the predictive mean vector could be large if several indicators were missing, the hot-deck step included fairly simple constraints. Only one logical constraint was used for lifetime use of pain relievers: those item nonrespondents who were known to have used pain relievers, but both their OxyContin® and "other" pain reliever indicators were missing, were required to have a donor who was a lifetime user of pain relievers. This pattern of nonresponse occurs either when respondents respond affirmatively to PR04 (lifetime use of one of the drugs appearing on the card, which includes OxyContin®) but fail to select any drugs from the card in PR04A or when respondents refuse to respond to each individual gate question (that specifically ask about certain pain relievers) but then respond affirmatively to the probe question PRREF.

No final imputation-revised variables indicating lifetime usage alone were created, because this information was recorded in the final imputation-revised recency-of-use variables. Imputation indicators also were not created, though temporary variables indicating that lifetime usage was imputed were maintained to inform the creation of the recency-of-use imputation indicators.

[91] This provisional hot-deck program actually uses a multivariate delta constraint, but the distance from donor to recipient is based only on the predicted probability of lifetime use of cocaine. This was done to avoid the complexity of the Mahalanobis distance calculation. Strictly, this is neither univariate nor multivariate matching.

6.3.2 Imputation-Revised Cigarette Recency and Frequency of Use (Imputation Set 2)

As indicated in Table 6.5, the second set of drug use variables to undergo imputation were the cigarette recency and 30-day frequency variables. The multiple RP/multiple PRD type of PMN was used to process these variables.

6.3.2.1 Sequence of Imputation

Because recency-of-use and frequency-of-use variables for a given drug were in the same imputation set, the calculation of predicted means for the frequency-of-use variables required the item nonrespondents to be identified as provisional past month and/or past year users. For this reason, cigarette recency was modeled prior to 30-day frequency, and provisional imputations were performed to allow for the identification of provisional past month users of cigarettes.

6.3.2.2 First Response Propensity Step (Cigarette Recency of Use)

The input to the weight adjustment model in the response propensity step for imputation set 2 was the preliminary analysis weight, PANALWT. To impute for missing recency-of-use values for cigarettes, it was first necessary to define the domain within each of the three age groups. Using the imputation-revised lifetime indication of use, the file was reduced to lifetime cigarette users. Among these lifetime users, item respondents and nonrespondents were identified across recency-of-use and 30-day frequency-of-use measures. If a valid response was provided for each drug use measure, the individual was deemed an item respondent for cigarettes. Otherwise, he or she was an item nonrespondent. See Tables D.4, D.5, and D.6 in Appendix D for details of the covariates used in the RP models for this variable.

6.3.2.3 First Prediction Step (Cigarette Recency of Use)

Using the adjusted weights, the probability of selecting each cigarette recency-of-use category was modeled within each age group using polytomous logistic regression. SUDAAN's MULTILOG procedure was used to estimate the parameters from the appropriate logistic model from which predicted probabilities for each of the cigarette recency categories were calculated for both item respondents and item nonrespondents. The four recency categories were the following:

1. past month;
2. past year, not past month;
3. past 3 years, not past year; and
4. lifetime, not past 3 years.

6.3.2.4 First Provisional Hot-Deck Step (Cigarette Recency of Use)

In order to define the domain for the cigarette 30-day frequency-of-use variable, it was necessary to create temporary imputed values in cases where the original cigarette recency value was missing. In order to save time and resources and because the imputation was only provisional, a univariate matching procedure was implemented. The only predicted mean used was the predicted probability of past month use, because past month use was the most critical measure of recency of cigarette use.

If possible, donors and recipients were required to be from states with the same level of usage of a given drug (state rank; see Section 3.5), where the level of usage was defined in terms of the weighted proportion of a given state's residents who had used cigarettes in the past month.[92] If insufficient donors were available within these constraints, they were loosened in the following order: (1) the delta constraint was removed, and (2) donors and recipients were no longer required to be from states with similar usage levels.

The only logical constraints placed on the neighborhoods involved cases where a general recency category was available for a respondent and imputation was required to determine the specific recency categories. The general recency categories that appeared are shown in Table 6.6. Logical constraints ensured that only donors with allowable specific recency categories were included in the neighborhood.

Table 6.6 General Incomplete Recency Categories for Cigarettes

General Incomplete Recency Category	Allowable Specific Recency Categories
Lifetime	1. Past month 2. Past year but not past month 3. Past 3 years but not past year 4. Lifetime but not past 3 years
Past Year	1. Past month 2. Past year but not past month
Lifetime, Not Past Year	1. Past 3 years but not past year 2. Lifetime but not past 3 years
Lifetime, Not Past Month	1. Past year but not past month 2. Past 3 years but not past year 3. Lifetime but not past 3 years
Lifetime, Not Past Month but within Past 3 Years	1. Past year but not past month 2. Past 3 years but not past year
Past 3 Years	1. Past month 2. Past year but not past month 3. Past 3 years but not past year

6.3.2.5 Second Response Propensity Step (Cigarette 30-Day Frequency)

The input to the weight adjustment model in the response propensity step for imputation set 2 was the preliminary analysis weight, PANALWT. The modeling of cigarette 30-day frequency followed that of recency. The file was first reduced to the domain, which was past month cigarette users, as defined by the provisional recency variable. Next, item respondents and nonrespondents were defined according to the same criterion used for the cigarette recency

[92] Those individuals whose past month use status was unknown were treated as if they were not past month users.

imputations. To be an item respondent, the individual had to have provided valid responses to both the cigarette recency and 30-day frequency measures. The item response propensity adjustment was then computed so that the respondents' weights accurately represented all past month users of cigarettes. See Tables D.4, D.5, and D.6 in Appendix D for details of the covariates used in the RP models for this variable.

6.3.2.6 Second Prediction Step (Cigarette 30-Day Frequency)

As stated in the previous section, only past month users of cigarettes were used to build the 30-day frequency-of-use model. The response variable of interest in the 30-day frequency-of-use models, prior to a normalizing transformation, was the proportion of the days in a month (30 days) on which a respondent used cigarettes. The range of values for the proportion was from (greater than) 0 to 1. Hence, to model 30-day frequency of use, the following empirical logit transformation was computed for all respondents:

$$\log\left[\left(Y_i + 0.5\right) / \left(N - Y_i + 0.5\right)\right],$$

where Y_i was the observed 30-day frequency for respondent i and N was 30, the total number of days in the month that the respondent could have used the substance. This transformation was nearly equivalent to the standard logit transformation:

$$Y_i^* = \log\left[P_i / \left(1 - P_i\right)\right],$$

where P_i was defined as the proportion of days in the past month on which respondent i used the drug. The standard logit transformation was not used because it was not defined for daily users.[93] Using the adjusted weights, a linear univariate regression model was then fitted using SUDAAN software for the log-transformed variable Y_i within each age group.

For cigarettes, the empirical distribution for 30-day frequency of use was in fact a mixture distribution, with a positively skewed distribution from 1 to 29 and a spike at 30. This substance was modeled using two separate models. One was a logistic model for daily use versus nondaily use among past month users. For the nondaily past month users (i.e., those who had used between 1 and 29 days), the model described above was used. In this case, the response variable in a linear regression model was a logit of the proportion of the period (30 days) during which a respondent used the substance. Both the predicted probability of daily use and the logit of the proportion of the month used (assuming nondaily use) were used as predicted means in the subsequent hot-deck step. The logit was back-transformed into a proportion before use in the hot-deck step.

6.3.2.7 Final Hot-Deck Step (Cigarette Recency and 30-Day Frequency)

The full predictive mean vector for cigarettes contained probabilities associated with several of the recency-of-use categories, a probability of daily use, and a predicted probability of

[93] If the respondent was a daily user of the substance, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[(N + 0.5)/0.5]$ with N = 30 so that it was defined for all respondents. See Cox and Snell (1989) for a discussion of the empirical logit transformation.

use on a given day in the past month. Each element in the full vector of predicted means was adjusted so that all elements were conditioned on the same usage status whenever possible. The elements in the predictive mean vector that could have potentially resulted are shown in Table 6.7, with the assumption that only the lifetime usage is known. If other information about the recency of use is known (e.g., past year user), the predictive mean vector is adjusted accordingly. The portion of the full predictive mean vector used to determine the neighborhood for a particular item nonrespondent was dependent on the pattern of missingness for that item nonrespondent. If partial information was available regarding recency of use, then that information was used to adjust the recency-of-use probabilities. The portions of the full predictive mean vector used for each missingness pattern, with accompanying adjustments, are provided in Table E.28 in Appendix E. The Mahalanobis distance was then calculated using only the portion of the predictive mean vector that was associated with the given missingness pattern, with elements appropriately adjusted. The likeness and logical constraints applied to each missingness pattern are also available in Table E.28.

Table 6.7 Elements of Full Predictive Mean Vector: Cigarettes

Drug Use Measure and Category of Interest	Predicted Mean
Recency of Use, Past Month Use[1]	*P*(past month user | lifetime user)
Recency of Use, Past Year but Not Past Month Use[1]	*P*(past year but not past month user | lifetime user)
Recency of Use, Past 3 Years but Not Past Year Use[1]	*P*(past 3 years but not past year user | lifetime user)
30-Day Frequency of Use for Nondaily Users over Past 30 Days	*P*(use on a given day in the month | past month user, not a daily user) × *P*(not a daily user | lifetime user) × *P*(past month user | lifetime user)[2]
Daily User over Past 30 Days	*P*(daily user | past month user) × *P*(past month user | lifetime user)[2]

[1] The final category for recency (lifetime but not past year or lifetime but not past 3 years) was not needed in the predictive mean vector, because the multinomial probabilities summed to 1, and this probability was determined by the other probabilities.

[2] Interpreting the proportion of the month used as a probability of use on a given day in the month assumed that the probability of use on each day in the month was equal, which was not true.

6.3.2.8 Final Variables (Cigarette Recency and 30-Day Frequency)

The final imputation-revised recency-of-use and 30-day frequency variables were identified with the prefix IR, followed by a five-letter identifier, where a three-letter code identified the drug (CIG) and the final two letters identified the measure (RC = recency of use, FM = 30-day frequency). Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR. The levels for the imputation indicators were the standard levels used for all imputation-revised variables: 1 = questionnaire data, 2 = logically assigned, 3 = statistically imputed, and 9 = legitimate skip (not a lifetime user).

6.3.2.9 Recodes for Additional Analyses

From the final imputation-revised recency-of-use variable, three dichotomous indicator variables were created to indicate cigarette use in the lifetime (CIGFLAG), past year (CIGYEAR), or past month (CIGMON).

6.3.3 Imputation-Revised Cigarette Age at First Use (Imputation Set 3)

As indicated in Table 6.5, the third imputation set consisted of the cigarette AFU variable. Unlike the recency and 12-month frequency-of-use variables, age at first drug use was not statistically imputed in the surveys prior to 1999. Instead, missing values were excluded from subsequent analyses. However, as with the 30-day frequency, missing AFU values have been replaced using imputation since the 1999 survey. Also, recent drug initiates (i.e., those whose current age was equal to or 1 year greater than the reported AFU) were asked the year and month of their first use. To have this information for all users, both missing MFU and missing YFU for less recent initiates (and recent initiates who did not report MFU and YFU) were replaced by assigning values consistent with the respondent's current age, interview date, imputation-revised AFU, and imputation-revised recency and frequency variables. To have complete date-of-first-use information, day of first use (DFU) was randomly assigned for all users. The combined data gave the respondent's AFU along with the date of first use.

6.3.3.1 Response Propensity Step (Cigarette Age at First Use)

The input to the weight adjustment model in the response propensity step for imputation set 3 was the preliminary analysis weight, PANALWT. To impute for missing cigarette AFU, it was necessary to define the eligible population. Using the imputed recency of use, the files were reduced to lifetime users of cigarettes. If a valid response was provided for the AFU measure,[94] the individual was deemed an item respondent. See Tables D.4, D.5, and D.6 in Appendix D for details of the covariates used in the RP models for this variable.

6.3.3.2 Prediction Step (Cigarette Age at First Use)

The response variable in the model for AFU, before a normalizing transformation, was the AFU as a proportion of the current age. The numerator in this proportion was an integer representing AFU. However, because this integer was in fact a truncated version of the real AFU, the value was made continuous by adding a random component between 0 and 1. Hence, expressing the proportion as $P_i = Y_i/N_i$, the numerator was given as

$$Y_i = \textit{Age at First Use}_i + \textit{Uniform(0,1) random number}\,.$$[95]

[94] Respondents who reported an age at first use of 1 or 2 were treated as item nonrespondents in the response propensity and prediction steps, because of the implausibility of such a young age at first use. In the hot-deck step, their response was left unchanged, but they were not allowed to be donors.

[95] In the event that the age at first use was equal to the age, Y_i was constrained so that it was equally likely to be anywhere on the interval $[\textit{Age at First Use}_i, N_i]$. Thus, Y_i was prevented from being greater than N_i.

The denominator in the proportion was the total age. The true age was known, based on the interview date and birth date. Expressing it in years rather than days required dividing by the number of days in the year:

$$N_i = (\text{Interview Date} - \text{Birth Date} + 1)/365.25.$$

After a weight adjustment, the empirical logit transformation was used as the response variable in a weighted linear univariate regression:

$$\log\left[\left(Y_i + 0.5\right)/\left(N_i - Y_i + 0.5\right)\right].$$

This transformation was nearly equivalent to the standard logit transformation:

$$Y_i^* = \log\left[P_i/\left(1 - P_i\right)\right],$$

which was not used, because it might be unstable for respondents who started using at their current age.

One unusual covariate used in the PRD model for cigarette AFU was a modified 30-day frequency variable for cigarettes. It was defined as follows:

new30$_i$	=	0	if respondent *i* did **not** use cigarettes in the past month
	=	30-day frequency	if respondent *i* used cigarettes in the past month

Naturally, the full model for AFU did not include the lifetime indicator for the drug in question, because the model was built on cigarette users. A summary of the starting and final models can be found in Tables D.4, D.5, and D.6 in Appendix D.

From the final model, a predicted value (based on the *Y* variable) was calculated for each cigarette user, which was then back-transformed to produce a predicted cigarette AFU.

6.3.3.3 Hot-Deck Step

The imputation-revised cigarette AFU assignment was conducted using a single predicted mean: the predicted AFU. Tables E.58 through E.60 provide a complete list of likeness and logical constraints applied to the cigarette AFU imputations. The likeness constraints for AFU were more stringent than those for the other drug use measures. Therefore, it was often necessary to loosen the constraints. Once these likeness constraints were removed, some complex logical constraints remained, based on the interview date, the birth date, and imputation-revised recency and frequency values.

6.3.3.4 Date-of-First-Use Assignments

After the AFU imputations, all lifetime users of cigarettes had nonmissing AFU values. Using this AFU, users were assigned values for MFU, YFU, and DFU. Recent initiates, or those respondents whose AFU was within 1 year of his or her age, were asked for their MFU and YFU.

The DFU was not collected in the questionnaire and was missing for all respondents. The MFU, YFU, and DFU data contained four patterns of missingness:

Pattern 1: Recent initiates: missing day of first use only;

Pattern 2: Recent initiates: missing month/day of first use;

Pattern 3: Recent initiates: missing year/month/day of first use; and

Pattern 4: Less recent initiates: missing year/month/day of first use.

For each missingness pattern, upper and lower bounds on the date of first use (i.e., the earliest possible date of first use and the latest possible date of first use) were determined. Once the earliest and latest possible dates of first use were determined, a day was randomly selected from this interval. The imputation-revised month/day/year values were then extracted from this date of first use.

6.3.3.4.1 Missingness Pattern 1

In this missingness pattern, a recent initiate provided all the information asked by the questionnaire (i.e., both the MFU and YFU). However, to obtain a complete date of first use, a DFU also was needed. Thus, a DFU was randomly assigned, given the respondent's MFU and YFU, in a way that was consistent with both the 30-day frequency/recency and AFU. Below is a brief description of the process used to obtain a date of first use in such cases. The imputed YFU, MFU, and DFU were extracted from the date, as defined below:

Final date of first use = *Earliest possible date* + [(*Days between earliest and latest date*) × (a random number generated from a Uniform (0,1) distribution)],

where

Days between earliest and latest date = Latest possible date − Earliest possible date + 1;

Earliest possible date = maximum [(AFUth birthday), (first day of the month indicated by MFU/YFU)]; and

Latest possible date =

- minimum [(Interview date − 30-day frequency + 1), (1 day before the (AFU + 1)th birthday)], *if recency* = 1;
- minimum [(Interview date − 30), (1 day before the (AFU + 1)th birthday)], *if recency* = 2; and
- minimum [(Interview date − 1 year), (1 day before the (AFU + 1)th birthday)], *if recency* = 3.

Note that it is impossible for recent initiates to have recency = 4 (lifetime but not past 3 years). Recent initiates had to have begun using the drug no earlier than their (AFU)th birthday.

Because AFU = current age, or AFU = current age − 1, their $(AFU)^{th}$ birthday was within the past 2 years. Respondents who had begun using the drug within the past 2 years must logically have last used the drug within the past 2 years, and therefore could not have had recency = 4.

In rare cases, the *earliest possible date* was set to 29 days before the interview. This occurred for respondents meeting all of the following conditions:

1. The *latest possible date* was within 29 days of the interview.
2. The *earliest possible date* determined by the above rule was within a year of the interview.
3. The recency = 1.
4. The 12-month frequency = 30-day frequency (if applicable), or the 12-month frequency = 1.

Logically, all the lifetime usage of the drug for these respondents occurred in the past 30 days (including the interview date). The first condition ensures that the application of this rule does not cause an inconsistency. The second condition implies that the drug was not used by these respondents more than 1 year ago. The third and fourth conditions imply that the drug was not used by these respondents in the interval (1 year before the interview, 1 month before the interview). Therefore, these respondents did not use the drug more than 1 month ago. All their lifetime use must have occurred in the past month.

6.3.3.4.2 Missingness Pattern 2

The second missingness pattern occurred when a recent initiate provided his or her YFU but did not provide an MFU. In such cases, a month and day were randomly assigned that were consistent with both the respondent's frequency/recency and with the AFU range. The imputed MFU and DFU were derived in the same manner as the date of first use in Missingness Pattern 1, except with the following changes:

- For the *earliest possible date*, replace "first day of the month indicated by MFU/YFU" with "January 1^{st} of the YFU."
- For the *latest possible date*, replace "last day of the month indicated by MFU/YFU" with "December 31^{st} of the YFU."

6.3.3.4.3 Missingness Pattern 3

Similar to Missingness Pattern 2, the third missingness pattern occurred when recent initiates provided neither an MFU nor a YFU value. In these cases, the MFU, YFU, and DFU were randomly assigned from a uniform distribution in a way that was consistent with both the cigarette 30-day frequency/recency and the AFU. Again, the imputed MFU, YFU, and DFU were derived in the same manner as described in Missingness Pattern 1.

6.3.3.4.4 Missingness Pattern 4

The fourth missingness pattern occurred when the respondent reported, or was imputed to, an AFU at least 2 years less than his or her age. This case is analogous to data prior to the 1999 survey, where MFU and YFU were not asked in the questionnaire. In this missingness pattern, the 30-day frequency was immaterial to the final date of first use because the respondent could not have begun using in the past month:

Earliest possible date = AFUth birthday; and

Latest possible date =

- 1 day before the (AFU + 1)th birthday, *if recency* < 4; or
- minimum [(Interview date − 3 years), (1 day before the (AFU + 1)th birthday)], *if recency* = 4.

6.3.3.5 Final Age and Date-of-First-Use Variables

As with all other imputation-revised variables, the final imputation-revised date-of-first-use variables were identified with the prefix IR, followed by a six-letter identifier, where a three-letter code identified the drug[96] and the final three letters identified the measure (AGE = age at first use, MFU = month of first use, YFU = year of first use, DFU = day of first use). Each IR variable was accompanied by an imputation indicator with the requisite II prefix. The levels for the imputation indicators were the standard levels used for all imputation-revised variables: 1 = questionnaire data, 2 = logically assigned, 3 = statistically imputed, and 9 = legitimate skip (not a lifetime user). Because survey respondents are not asked for the specific day on which they first used the drug of interest, all respondents in the domain receive IIxxxDFU = 3. Also, as indicated above, only recent initiates are asked for the year and month of first drug use. Subsequently, these questions have high rates of nonresponse because of the skip logic embedded in the questionnaire, as all other individuals in the domain require imputation for their MFU and YFU.

6.3.4 Imputation-Revised Age at First Daily Cigarette Use (Imputation Sets 4 and 5)

In addition to AFU, the cigarettes module also included a question asking for the respondent's age at first daily cigarette use, where a daily user was defined as someone who reported having at some time smoked cigarettes every day for a period of at least 30 days. Imputation procedures for age at first daily cigarette use were similar to AFU, with some key exceptions as discussed below.

One such exception involved the domain of the AFU variable. Whereas the AFU question was asked of all cigarette users, the age-at-first-daily-use question was asked of only daily users. The "daily use" indication came from two sources. If a respondent answered either the 30-day frequency or estimated 30-day frequency with a "30," or if the respondent had a "yes" value for

[96] Exceptions to this rule occurred with marijuana and cigarette daily use. For historical reasons, marijuana contained a two-letter code (MJ). Marijuana variables therefore ended with a five-letter identifier rather than a six-letter identifier. The code for cigarette daily use was CDU, which differed from the general cigarette code of CIG. Details about cigarette daily use are provided in Section 6.3.4.

the edited variable associated with the "ever daily used" question (CIGDLYMO), then he or she was considered a daily user. For more information about CIGDLYMO, see Section 6.2.7. The "ever daily used" question (CIGDLYMO) can be thought of as a lifetime "child" drug to the "parent" lifetime cigarette use question (CIGEVER). However, anyone who answers the 30-day frequency or estimated 30-day frequency with a "30" is automatically skipped out of this question. Therefore, it is necessary to obtain the imputation-revised cigarette 30-day frequency (IRCIGFM) prior to imputing the lifetime-daily-cigarette-use variable (IRCDULF) so that it is not included with the other lifetime drug indicators as part of imputation set 1. Instead, as indicated in Table 6.5, the age at first daily cigarette use actually contains two separate imputation sets. Imputation set 4 includes the lifetime indicator of daily cigarette use and imputation set 5 includes the age-at-first-daily-cigarette-use variables. At this stage in the process, there should be no missing responses to the 30-day frequency question, which were imputed as part of imputation set 2 as discussed above. Daily users, based on 30-day frequency, should be either known (based on a response in the survey) or imputed. However, responses for the ever-daily-used question (CIGDLYMO) could still be missing, and, therefore, it was first necessary to impute these values to define the domain for the age-at-first-daily-use variable.

6.3.4.1 Response Propensity Step (Ever-Daily-Used Cigarettes)

The input to the weight adjustment model in the response propensity step for imputation sets 4 and 5 was the preliminary analysis weight, PANALWT. To impute for missing values in the ever-daily-used variable, it was necessary to define the domain: lifetime users of cigarettes who had an imputation-revised 30-day frequency[97] of fewer than 30 days (includes legitimate skip codes for lifetime but not past month users). If a valid response was provided in the ever-daily-used variable, the individual was deemed an item respondent. See Tables D.4, D.5, and D.6 in Appendix D for details of the covariates used in the RP models for this variable.

6.3.4.2 Prediction Step (Ever-Daily-Used Cigarettes)

After the weights were adjusted, the ever-daily-used variable was modeled using weighted logistic regression in SUDAAN. The predicted mean from this model was the predicted probability of ever smoking cigarettes daily.

6.3.4.3 Hot-Deck Step (Ever-Daily-Used Cigarettes)

The predicted mean from the prior step was used in a straightforward hot-deck step, which is summarized in Tables E.24 and E.25.

6.3.4.4 Hot-Deck Step (Age at First Daily Cigarette Use)

Instead of separately modeling age at first daily cigarette use, the predicted means from the age-at-first-cigarette-use models were used to determine neighborhoods. The imputation-revised age-at-first-daily-use assignment was conducted using univariate matching and univariate assignment.

---

[97] The imputation-revised 30-day frequency included responses from the 30-day frequency question (CG07), as well as the estimated 30-day frequency question (CG07DKRE).

All the logical constraints applied to age at first cigarette use were also applied to age at first daily cigarette use. Besides those logical constraints, an additional logical constraint was applied specifically to age at first daily cigarette use. If the cigarette AFU was not missing for a recipient with a missing age at first daily use, the donors were prevented from having an age at first daily use earlier than the preexisting cigarette AFU. This constraint was applied as daily cigarette users constitute a subset of all cigarette users. Therefore, daily use of cigarettes can be thought of as a child drug for cigarettes, with a lifetime indicator and an AFU measure but no recency or frequency. This association required that these variables remain internally consistent for each respondent.

The likeness constraints were nearly identical to those used for cigarette AFU. There was only one difference: an additional step was employed if no donor was found after loosening all of the likeness constraints. In particular, if the AFU and age at first daily use were both initially missing, the imputed AFU was set back to missing and reimputed simultaneously with the age at first daily use so that they were mutually consistent.[98]

Subject to these constraints, the age-at-first-daily-use value of the randomly selected donor was then assigned to the recipient.

6.3.4.5 Assignment of Date of First Daily Cigarette Use

After the imputation-revised age at first daily cigarette use was created, all daily cigarette users had a valid age at first daily cigarette use. From this age, a year/month/day of first daily use was assigned. The date assignment procedure was identical to the procedure described in Section 6.3.3.4 with the following exception. In the setting of the *earliest possible date* for daily cigarette use, the overall cigarette date of first use was used as an additional bound. This was done for cigarettes and other substances to ensure that the child drug's date of first use was never earlier than the parent drug's date of first use.

6.3.4.6 Final Variables

The final imputation-revised date-of-first-daily-cigarette-use variables were named in the same manner as described in Section 6.3.3.5. However, the three-letter identifier for cigarette daily use was CD2. As with the general cigarette use variables, each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

6.3.5 "Other" Drugs Recency and Frequency

Imputation for the parent and child recency and frequency variables for imputation sets 6, 8, 10, 11, 13, 15, 17, 19, 21, 23, 25, 27, and 29 in Table 6.5 (smokeless tobacco, cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine, and heroin, respectively) was done in a manner similar to that described above for cigarettes. The following deviations from the process described for cigarettes applied to these "other" drugs.

---

[98] Though it has occurred in prior years, the situation where no donors were available, even after loosening all constraints, did not occur in the 2014 NSDUH.

The order of imputation for smokeless tobacco and cigars was identical to that for cigarettes, with recency of use being modeled first, followed by 30-day frequency. However, not all imputation sets included the same variables. Alcohol, inhalants, marijuana, hallucinogens, cocaine and crack, and heroin also included a measure for 12-month frequency of use. For these drugs, imputation proceeded in the following order: recency of use, 12-month frequency of use, and 30-day frequency of use. For a given drug, this ordering allowed recency of use to be included in the model for 12-month frequency of use and allowed 12-month frequency of use to be included in the model for 30-day frequency. Further, this ordering allowed the provisional recency of use to define the domains for the frequency measures. Alcohol also had a measure for binge drinking frequency, which was modeled after the 30-day frequency-of-use variable so that the provisionally imputed value could be used as a covariate in the binge drinking frequency model. For pain relievers, tranquilizers, stimulants, and sedatives, the respondents were not asked to report their 30-day frequency of use. For these imputation sets, recency of use was completed first, followed by the 12-month frequency-of-use variable. For pipes, the respondents were only asked about their most recent use, and no information was collected regarding frequency of use in the past year or month. Therefore, only the recency-of-use variable required modeling and imputation.

6.3.5.1 Recency of Use

6.3.5.1.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the parent and child recency-of-use variables was the preliminary analysis weight, PANALWT. Similar to cigarettes, the eligible population for the recency-of-use models included all lifetime users of the drug of interest as identified by the imputation-revised lifetime drug use variables. However, the identification of respondents and nonrespondents for the purpose of imputation differed from drug to drug depending on the information collected in the questionnaire. In general, an individual had to have provided a valid response to all variables included in the imputation set to be classified as a respondent. If the imputation set included both "parent" and "child" drugs (Table 6.1), then this requirement extended across all observed measures of drug use. For example, to be classified as a respondent for the hallucinogens imputation set, a valid response must be provided for the overall hallucinogens recency of use, 12-month frequency of use, and 30-day frequency of use, as well as the recency of use for LSD, PCP, and Ecstasy questions. See Tables D.7 through D.45 in Appendix D for details of the covariates used in the RP models for these variables.

6.3.5.1.2 Prediction Step

Only cigarettes, cigars, chewing tobacco, and snuff included a recency category for past 3 years but not past year. For all other drugs except pipes, the outcome variable was a three-level categorical variable with the following levels:

1. past month
2. past year, not past month
3. lifetime, not past year

For pipes, the outcome variable had only two levels:

1. past month
2. lifetime, not past month

Because cigarettes were the first recency/frequency imputation set, it was not possible to include the recency information for other drugs as covariates in the PRD model. However, for drugs other than cigarettes, recency-of-use covariates for cigarettes, alcohol, and marijuana replaced the lifetime indicators where applicable. For example, the PRD model for alcohol included recency indicators for cigarettes but only included the lifetime usage indicator for marijuana because this drug comes later in the sequence.

6.3.5.1.3 Provisional Hot-Deck Step

For certain cases, a general recency category was assigned during the editing process, and the specific recency was then determined during imputation. However, the categories available for both the general recency and the specific recency varied from drug to drug depending on the number of levels included in the recency-of-use measure. The allowable general and specific recency categories for cigarettes are shown in Table 6.6. The same categories apply to cigars, chewing tobacco, and snuff. For all other drugs except pipes, the only general incomplete recency categories that were applicable were lifetime and past year (the first two rows). For pipes, only the lifetime category was applicable. Logical constraints ensured that only donors with allowable specific recency categories were included in the neighborhood of potential donors.

Occasionally, more than one substance was associated with a single predicted mean, leading to a multivariate assignment of imputed values. However, for the provisional imputed values, a multivariate assignment was necessary only if the substances associated with a single predicted mean were of equal standing. This occurred with smokeless tobacco, which consists of chewing tobacco and snuff. No provisional imputed values were determined for substances that were a subset of the substance associated with the predicted mean (parent/child drugs). Examples of such situations included cocaine (parent) and crack (child); pain relievers (parent) and OxyContin® (child); stimulants (parent) and methamphetamine (child); and hallucinogens (parent) and LSD, PCP, and Ecstasy (child).

As with lifetime use, one model for smokeless tobacco (a combination of the chew and snuff responses) was fitted rather than individual models for chew and for snuff. The nearest neighbor hot-deck neighborhood was then based on the predicted probability of past month use of smokeless tobacco. Missing recency-of-use values for chew and/or snuff were replaced with the (provisional) values from a donor within this neighborhood. The provisional recency of use for smokeless tobacco was obtained by combining the recency-of-use information from chew and snuff.

6.3.5.1.4 Hierarchical Modeling

For certain drugs, the proportion of users who were past year users was quite small when compared with the total number of lifetime users. The lopsided distributions[99] for these drugs caused convergence problems when fitting polytomous logistic models. This problem occurred with the following set of drugs that were either rare overall or were rare within one or more age groups: inhalants, hallucinogens, sedatives, stimulants, tranquilizers, and heroin. To alleviate this problem, the single polytomous logistic model was replaced with two dichotomous logistic models[100] that were fit hierarchically.

As with the polytomous logistic model, the first dichotomous logistic model was fit among lifetime users, but the past month and past year but not past month categories in the response variable were collapsed into a single level. In a similar manner to other recency-of-use models, respondents' weights were adjusted so that they represented all lifetime users. The predicted probability of past year use given lifetime use was obtained from this model.

The second model was limited to past year users, where the response variable had two levels: past month and past year but not past month users. For the second model, respondents' weights were adjusted so that they represented all past year users. In order to do this, it was necessary to completely define the domain of past year users. Missing values were provisionally imputed to past year or not past year use by randomly allocating the response using the predicted means from the first model.

From the two dichotomous logistic models, both the probability of past month use and the probability of past year but not past month use were obtained and used in the provisional hot-deck program for recency. Once the predicted means were determined from the two models, a single vector of predicted means conditional on lifetime usage, as with the polytomous logistic models, was determined as follows:

$$P(\text{past month use} \mid \text{lifetime use}) = P(\text{past month use} \mid \text{past year use}) \times P(\text{past year use} \mid \text{lifetime use}),\ \text{and}$$

$$P(\text{past year, not past month use} \mid \text{lifetime use}) = P(\text{past year, not past month use} \mid \text{past year use}) \times P(\text{past year use} \mid \text{lifetime use}).$$

6.3.5.2 12-Month Frequency of Use

The modeling of 12-month frequency sequentially followed that of recency of use for alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine and crack, and heroin.

---

[99] A "lopsided distribution" in the context of recency of use is where, among the categories of past month use, past year but not past month use, and lifetime not past year use, only a small minority of respondents gave a response of "past month use."

[100] The set of covariates used for these dichotomous logistic models were the same as those for logistic modeling given earlier in this section.

6.3.5.2.1 *Response Propensity Step*

The input to the weight adjustment model in the response propensity step for the 12-month drug frequency-of-use variables was the preliminary analysis weight, PANALWT. The eligible population for the imputation of 12-month frequency of use was past year users of the drug in question (as defined by the provisional recency of use). The item response indicator and the response propensity adjustment were defined among the past year users of each drug. Item respondents were defined using the same criterion as was used in the recency-of-use imputations. Namely, the respondent had to have a valid response to all of the applicable measures for the drug of interest. The item response propensity adjustment was then computed so that the respondents' weights accurately represented all past year users of the drug. See Appendix D for details of the covariates used in the RP models for these variables.

6.3.5.2.2 *Prediction Step*

As indicated in the previous section, only past year users of the drug of interest were used to build the 12-month frequency-of-use model. The response variable of interest in the 12-month frequency-of-use models for most respondents, prior to a normalizing transformation, was the proportion of the days in a full year (365.25) on which a respondent used a particular drug. For example, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100/365.25. Some respondents, however, started using the drug within the past year. If they responded to the MFU question, the difference between the MFU and the date of the interview indicated the total time period during which they could have been using drugs.[101] If the date of the interview was July 10, for example, and the MFU was March of the same year, the maximum period during which the respondent could have used is the number of days between March 1 and July 10 (inclusive), or 101. Thus, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100/101 instead of 100/365.25. The range of values for the proportion was from (greater than) 0 to 1. Hence, in order to model 12-month frequency of use, the following empirical logit transformation was computed for all respondents:

$$\log\left[\left(Y_i+0.5\right)/\left(N_i-Y_i+0.5\right)\right],$$

where Y_i is the observed 12-month frequency for respondent i and N_i is the total number of days in the year that respondent i could have used the substance. This transformation is nearly equivalent to the standard logit transformation:

$$Y_i^*=\log\left[P_i/\left(1-P_i\right)\right],$$

where P_i is defined as the proportion of days in the past year in which respondent i used the drug. The standard logit transformation was not used because it was not defined for daily

[101] If a respondent initiated use in the past year (according to his or her age-at-first-use response), but did not answer the month-at-first-use question, then the maximum period the respondent could have been using drugs was assumed to be 365.25 because no other information was available.

users.[102] Using the adjusted weights, a linear univariate regression model using SUDAAN software was then fitted for the log-transformed variable Y_i within each age group.

- **Covariates:** Because the 12-month frequency models were limited to past year users, only two recency categories could have resulted: past month use and past year but not past month use.[103] Hence, recency of use for the drug being modeled was represented as a covariate in the 12-month frequency-of-use model by a single indicator variable representing these two categories. Imputation-revised recency of use for other drugs was used if available. If the missing values for a given drug's recency of use had not yet been imputed, a single covariate was used that indicated lifetime usage of that drug. To control for state variations in drug use, the state-rank groups defined for the recency-of-use imputations were included as covariates in the 12-month frequency-of-use models.[104] Appendix D provides a complete summary of the 12-month frequency-of-use models.
- **Predicted Means:** The predicted mean that resulted from the 12-month frequency-of-use model was a logit of the proportion of the year used. This logit was back-transformed into a proportion for use as the variable from which the neighborhoods were created. This proportion could be treated as a probability, which, in turn, could be multiplied by the probability of past year use to make the predicted mean conditional on lifetime use of the drug in question. When calculating predicted means for some item nonrespondents, sometimes it was not known whether they were past year users. Hence, to make the predicted means conditional on the same recency of use, the variables were transformed to make them conditional on what was known.

6.3.5.2.3 Provisional Hot-Deck Step

For imputation sets that included both 12-month frequency and 30-day frequency—alcohol (Set 11), inhalants (13), marijuana (15), hallucinogens (17), cocaine and crack (27), and heroin (29)—it was necessary to provisionally impute the 12-month frequency-of-use variable so that it could be used as a covariate in the 30-day frequency-of-use imputations.

The logical constraints involved the interview date, incidence (AFU, MFU, and YFU), birthday, recency of use, and 30-day frequency of use. The likeness constraints used in the assignment of values for 12-month frequency of use were similar to those used for recency of use. State-rank groups were again based on level of past month usage. Recipients and donors were also required to have the same recency of use (past month vs. past year but not past month),

[102] If the respondent was a daily user of the substance, then $\log[(Y+0.5)/(N-Y+0.5)] \approx \log[(N+0.5)/0.5]$ with N = 365.25 so that it was defined for all respondents. See Cox and Snell (1989) for a discussion of the empirical logit transformation.

[103] For item nonrespondents, where parameter estimates were used to determine predicted means, past year use was defined based on a provisional imputation.

[104] As with the recency-of-use models, for a few cases, the state-rank variable could not be included in the model. Usually, but not always, the age group/drug combination that had problems was the same for recency of use and 12-month frequency of use.

whether that recency of use was reported or imputed.[105] If no donors were available within these constraints, then they were loosened in the following order: (1) the delta constraint was removed, (2) donors and recipients were no longer required to be from states with similar usage levels, and (3) donors and recipients were no longer required to have the same recency of use.

6.3.5.2.4 Assignment of Provisional Imputed Values

For all drug use measures except 12-month frequency, the observed value of interest was donated directly to the recipient. However, because donors and recipients could potentially have had a different maximum possible number of days in the year that they could have used a substance, the observed proportion of the total period was donated rather than the observed 12-month frequency. In the assignment step, the donor's proportion of the total period was multiplied by the recipient's maximum possible number of days in the year on which he or she could have used the substance in order to arrive at a 12-month frequency-of-use value for the recipient.

Occasionally, more than one substance was associated with a single predicted mean. However, for the provisionally imputed values, only the parent drug was used as a covariate in later models. Therefore, multivariate assignments were not needed in the provisional hot-deck step, but they did occur in the final hot-deck step for recency and frequency. For example, the recency and frequency variables for cocaine and crack formed a single imputation set (27). Although 12-month frequency questions were asked for both cocaine and crack, only the 12-month frequency for cocaine was modeled, and only the 12-month frequency for cocaine was used as a covariate in the subsequent PRD model (30-day cocaine frequency). This means that there was no need to impute a provisional value for 12-month frequency for crack.

For pain relievers, tranquilizers, stimulants, and sedatives, no provisional assignment of imputed values was necessary, because these drugs did not include a measure for 30-day frequency (Table 6.5).

6.3.5.3 30-Day Frequency of Use

6.3.5.3.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the 30-day drug frequency-of-use variables was the preliminary analysis weight, PANALWT. As with cigarettes, the file was first reduced to the domain of past month users, as defined by the provisional recency variable. Next, item respondents and nonrespondents were defined according to the same criterion used for the recency and 12-month frequency imputations. To be an item respondent, the individual had to have provided valid responses to all applicable measures for the drug of interest. The item response propensity adjustment was then computed so that the respondents' weights accurately represented all past month users of the drug. In contrast with the RP model for cigarettes, the provisional 12-month frequency was included as a covariate for

[105] Because all respondents in the 12-month frequency-of-use imputation were past year users by definition, item nonrespondents who were past month users required donors who were past month users, and item nonrespondents who were past year but not past month users required donors who matched that specific recency category.

those drug modules that asked the respondent to report this measure (Table 6.5). In addition, recencies of use for cigarettes, smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives were included if available.[106] See Appendix D for details of the covariates used in the RP model for these variables.

6.3.5.3.2 Prediction Step

As with cigarettes, the empirical distribution for 30-day frequency of use for chewing tobacco and snuff was a mixture distribution, with a positively skewed distribution from 1 to 29 and a spike at 30. For both chewing tobacco and snuff, two models were fit. The first model determined daily versus nondaily use among past month users and the second model was used for nondaily past month users. For this second model, the response variable being modeled was the logit of the proportion of the period (30 days) during which the respondent used the substance. All other drugs that included a measure for 30-day frequency used a single model for all past month users.

6.3.5.3.3 Provisional Hot-Deck Step

The only drug for which provisional 30-day frequency values were required was alcohol, because provisional 30-day frequencies were required to calculate 30-day binge drinking provisional values. Neighborhoods were created for each alcohol item nonrespondent using univariate matching. The predicted means used to create the neighborhoods were given by the product of the predicted proportion of the month used (conditioned on past month use) and the probability of past month use given lifetime use (taken from the recency-of-use models).

Logical constraints involved lower and upper bounds for the donated 30-day frequency based on the respondent's preexisting 12-month frequency, 30-day binge drinking frequency, birthday, interview date, and incidence (AFU, MFU, and YFU). The likeness constraints were similar to those used in the provisional hot-deck step for 12-month frequency and were loosened in the following order: (1) the delta constraint was removed, and (2) donors and recipients were no longer required to be from states with similar usage levels.

6.3.5.3.4 Assignment of Provisional Imputed Values

Although more than one substance was occasionally associated with a single predicted mean, the provisionally imputed 30-day frequencies were required only if they were needed as covariates in a subsequent model. Of the substances within the multivariate set of recency of use and frequencies of use, only alcohol contained a measure (30-day binge drinking frequency) that was lower in the sequence than 30-day frequency of use. Because alcohol is not a parent/child drug, no multivariate assignments were required for provisionally imputed 30-day frequency.

[106] If the recency of use for a particular drug was not yet defined, the lifetime indication of use was used instead. The recency of use of the drug being modeled was not used, because all respondents in the model were past month users.

6.3.5.4 30-Day Binge Drinking Frequency

For alcohol, an additional variable was defined that measured level of usage. In particular, the variable DR5DAY measured the binge drinking frequency or the number of days in the past month during which the respondent had five or more drinks. The imputation of the 30-day binge drinking frequency was similar to the imputation of 30-day frequency of alcohol use.

6.3.5.4.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the 30-day binge drinking frequency variable was the preliminary analysis weight, PANALWT. The response propensity model was built using all past month users of alcohol, whether they were binge drinkers or not. Item respondents for alcohol were defined across recency, 12-month frequency, 30-day frequency, and the 30-day binge drinking frequency measures. Therefore, the weight adjustment used in the modeling of the 30-day binge drinking frequency was the same as was used for the 30-day frequency model. See Tables D.16, D.17, and D.18 in Appendix D for details of the covariates used in the RP model for this variable.

6.3.5.4.2 Prediction Step

The response variable of interest in the 30-day binge drinking frequency model, prior to a normalizing transformation, was the proportion of the days in a month (30) on which a respondent had five or more drinks. The range of values for the proportion was from 0 to 1, inclusive. Hence, to model 30-day binge drinking frequency of use, the following empirical logit transformation was computed for all respondents:

$$\log\left[\left(Y_i+0.5\right)/\left(N-Y_i+0.5\right)\right],$$

where Y_i was the observed 30-day binge drinking frequency for respondent i and N was 30, the total number of days in the month that the respondent could have binge drunk. This transformation was nearly equivalent to the standard logit transformation:

$$Y_i^* = \log\left[P_i/\left(1-P_i\right)\right],$$

where P_i was defined as the proportion of days in the past month during which respondent i had five or more drinks. The standard logit transformation was not used, because it was not defined for daily binge drinkers nor was it defined for nonbinge drinkers among past month users.[107] Using the adjusted weights, a linear univariate regression model was then fitted for the log-transformed variable Y_i within each age group.

[107] If the respondent was a daily binge drinker of alcohol, then $\log[(Y+0.5)/(N-Y+0.5)] \approx \log[(N+0.5)/0.5]$, where Y was the observed 30-day binge drinking frequency and N was the total number of days that the respondent could have used (usually 30). If the proportion was zero, then $\log[(Y+0.5)/(N-Y+0.5)] \approx \log[0.5/(N+0.5)]$. See Cox and Snell (1989) for a discussion of the empirical logit transformation.

The predicted means from this model were used solely in the multivariate predictive mean vector used in the final hot-deck step. No provisional imputed values were determined.

6.3.5.5 Final Hot-Deck Step

The same principles that applied to the final hot-deck step for cigarettes applied to other drugs. However, for substances with child drugs and substances with both 12-month and 30-day frequencies, the logical constraints were considerably more complicated, the predictive mean vectors were larger, and the number of missingness patterns was greater. Appendix E provides detailed information on these hot-deck steps, and Table 6.8 provides a listing of the full predictive mean vector as applied to all final hot-deck programs for recency and frequency.

The construction of the predictive mean vectors for certain drugs was often complex. The main reason for the complexity is that recency and frequency models were not fit for all child drugs. In fact, the predicted means from the models for the parent drug were often used as surrogates for the child drug predicted means to reduce the number of models that needed to be fit and to avoid convergence problems with small sample sizes for some of the rarer child drugs. For example, if the individual requiring imputation is a past year user of cocaine but he or she has a missing crack recency, then the predictive mean vector includes the probability of past month cocaine use, given that the individual is a past year user of cocaine. When constructing the predictive mean vectors, the following general principles were followed:

1. If both the parent drug recency and the child drug recency(ies) were missing, condition on the general recency category of the parent drug.
2. For smokeless tobacco, if both the chewing tobacco recency and the snuff recency were missing, condition on the union of the two sets of possible specific recency categories. For example, if chewing tobacco recency was "past year" and snuff recency was "past 3 years but not past month," condition on use in the past 3 years.
3. Condition all elements of the predictive mean vector on the same general recency level.

Table 6.8 Elements of Full Predictive Mean Vector

Drug Use Measure and Category of Interest	Predicted Mean	Substance
Recency of Use, Past Month Use[1]	*P*(past month user | lifetime user)	All substances
Recency of Use, Past Year but Not Past Month Use[1]	*P*(past year but not past month user | lifetime user)	All substances except pipes
Recency of Use, Past 3 Years but Not Past Year Use[1]	*P*(past 3 years but not past year user | lifetime user)	Tobacco products[2] only
12-Month Frequency of Use	*P*(use on a given day in the year | past year user) × *P*(past year user | lifetime user)[3]	All substances except tobacco
30-Day Frequency of Use for Alcohol and Substances with Few Daily Users[4]	*P*(use on a given day in the month | past month user) × *P*(past month user | lifetime user)[5]	All substances except cigarettes, chew,[6] snuff, pipes, and pills[7]

Drug Use Measure and Category of Interest	Predicted Mean	Substance
30-Day Frequency of Use for Substances with Many Daily Users (excluding Alcohol)	*P*(use on a given day in the month | past month user, not a daily user) × *P*(not a daily user | lifetime user) × *P*(past month user | lifetime user)[5]	Cigarettes, chewing tobacco, snuff
Daily User over Past 30 Days	*P*(daily user | past month user) × *P*(past month user | lifetime user)[5]	Cigarettes, chewing tobacco, snuff
30-Day Binge Drinking Frequency	*P*(drank 5 or more drinks on a given day in the past month | past month user) × *P*(past month user | lifetime user)[5]	Alcohol only

[1] The final category for recency (lifetime but not past year or lifetime but not past 3 years) was not needed in the predictive mean vector, because the multinomial probabilities summed to 1, and this probability was determined by the other probabilities.
[2] "Tobacco products" includes cigarettes, cigars, chewing tobacco, and snuff.
[3] Interpreting the proportion of the year used as a probability of use on a given day in the year assumed that the probability of use on each day in the year was equal. However, this was not true. The violation of this assumption did not seriously affect the ability to find a reasonable variable to use for finding a neighborhood, and it did allow the predicted mean to be made conditional on what was known.
[4] Alcohol, with many daily users, was included in this group because the distribution did not show a severe drop-off from 30 days a month to 29 days a month, as was apparent with cigarettes, chewing tobacco, and snuff.
[5] Interpreting the proportion of the month used as a probability of use on a given day in the month assumed that the probability of use on each day in the month was equal, which was not true, in the same manner as the 12-month frequency of use (see note #3 for this table).
[6] "Chew" is short for "chewing tobacco."
[7] "Pills" includes pain relievers, tranquilizers, stimulants, and sedatives.

6.3.5.6 Final Variables

Similar to the final imputation-revised recency-of-use and 30-day frequency variables for cigarettes (IRCIGRC and IRCIGFM), the final imputation-revised recency and frequency variables for other drugs were identified with the prefix IR, followed by a five-letter identifier, where a three-letter code identified the drug and the final two letters identified the measure. In addition to the RC and FM identifiers used for cigarettes, the identifier FY was used for the 12-month frequency variable. Again, each IR variable was accompanied by an imputation indicator with the requisite II prefix.

6.3.5.7 Recodes for Additional Analyses

Section 6.3.2.9 lists three dichotomous indicator variables that were created to indicate cigarette use in the lifetime (CIGFLAG), past year (CIGYEAR), or past month (CIGMON). Analogous variables were also created for each drug for which an imputation-revised recency was created.

Several other prevalence recodes, which covered the same three measures, were created to incorporate information from several different drugs. Table 6.9 lists these recodes and the recency variables that were used to create them. The creation of these variables was also straightforward. If the respondent was a lifetime user of any of the drugs, then the FLAG variable was set to 1; otherwise, it was set to 0. The YR and MON variables were processed in a similar manner.

Table 6.9 Prevalence Recodes Incorporating More than One Recency Variable

General Drug Category	Variable Names	Source Recency Variables
Tobacco	TOBFLAG, TOBYR, TOBMON	Cigarettes, smokeless tobacco, cigars, pipes
Psychotherapeutics	PSYFLAG2, PSYYR2, PSYMON2[1]	Pain relievers, tranquilizers, stimulants, sedatives
Illicit Drugs Other than Marijuana	IEMFLAG, IEMYR, IEMMON	Psychotherapeutics, plus inhalants, hallucinogens, cocaine, and heroin
Illicit Drugs, but Only Marijuana	MJOFLAG, MJOYR, MJOMON	Same as MRJFLAG, MRJYR, and MRJMON, except set to 0 if the corresponding IEM variable is equal to 1
Illicit Drugs	SUMFLAG, SUMYR, SUMMON	Illicit drugs other than marijuana, plus marijuana

[1] These variable names include a suffix of "2" to distinguish them from earlier versions of psychotherapeutics recodes.

6.3.6 "Other" Drugs Age at First Use

The AFU variables for imputation sets 7, 9, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30 in Table 6.5 (smokeless tobacco, cigars, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, cocaine, and heroin, respectively) were imputed in a manner similar to that described for cigarettes. However, some deviations from the process described for cigarettes applied to the AFU for these "other" drugs as described below.

6.3.6.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the AFU variables for other drugs was the preliminary analysis weight, PANALWT. The RP step for AFU for other drugs was very similar to the RP step for cigarette AFU (Section 6.3.3.1). For substances that included child drugs, a response had to be provided for each drug for an individual to be considered a respondent for imputation purposes. Appendix D provides a complete list of covariates used in each model to properly adjust the weights.

6.3.6.2 Prediction Step

The PRD step for AFU for other drugs was also very similar to the analogous step for cigarettes (Section 6.3.3.2). For substances with child drugs, only the parent drug was modeled. Modified versions of the 12-month frequency of use (where applicable) and AFU of previously imputed drugs were used as covariates and were defined as follows:

$\mathbf{new12_i}$ = 0		if respondent *i* did **not** use the drug of interest in the past 12 months
	= 12-month frequency	if respondent *i* used the drug of interest in the past 12 months
$\mathbf{AFU_i}$ = 0		if respondent *i* is **not** a lifetime drug user of the drug of interest

= age at first use   if respondent i is a lifetime drug user of the drug of interest

6.3.6.3 Hot-Deck Step

For smokeless tobacco (chewing tobacco and snuff), cocaine (crack), hallucinogens (LSD, PCP, and Ecstasy), pain relievers (OxyContin®), and stimulants (methamphetamine), more than one AFU variable was associated with a single predicted AFU. This led to a multivariate assignment of the imputed values.

- One model for smokeless tobacco was fitted rather than individual models for chewing tobacco and snuff. The item nonrespondent received values from the donor for both chewing tobacco (if missing) and snuff (if missing), and the AFU for smokeless tobacco was obtained by taking the minimum AFU from chewing tobacco and snuff. Respondents were never asked directly for their AFU for smokeless tobacco.
- For other substances with child drugs, respondents were asked for the AFU for the parent drug and were also asked for their AFU for each child drug. This often led to complex and numerous logical constraints. These constraints used not only parent and child ages at first use but also imputation-revised recencies and frequencies.

6.3.6.4 Month-of-First-Use, Year-of-First-Use, and Day-of-First-Use Assignments

The general principles described in Section 6.3.3.4 applied to the remaining drugs with the following exceptions.

- For smokeless tobacco, the minimum of the chewing tobacco and snuff dates was used to produce the smokeless tobacco date of first use.
- For all child drugs (daily cigarettes, LSD, PCP, ecstasy, OxyContin®, methamphetamine, and crack), the corresponding parent drug's date of first use was assigned first. Then, in the setting of the *earliest possible date* for the child drug, the parent drug's date of first use was used as an additional bound. This was done to ensure that the child drug's date of first use was never earlier than the parent drug's date of first use.
- For all parent drugs whose child drugs had recency and frequency information (hallucinogens, pain relievers, stimulants, and cocaine), the child drug recency and frequency information was used to bound the *latest possible date* for the date of first use. For example, respondents with LSD recency = 3 (i.e., lifetime but not past year user of LSD) could not have first used hallucinogens within the past year, regardless of the hallucinogens recency value. The bound created using the child drug recency and frequency was calculated in exactly the same way as for the parent recency and frequency information.
- For hallucinogens, pain relievers, and stimulants, an indicator of lifetime use of drugs other than the child drugs was created (Table 6.1). For pain relievers and stimulants,

if the respondent was not a lifetime user of the "other" drugs, then the child drug's date of first use was logically assigned to the parent drug's date of first use. The handling of the child drugs for hallucinogens was more complex, because there was more than one of them. The algorithm follows:

1. The date of first use was assigned for overall hallucinogens.
2. The *earliest possible date*, *latest possible date*, and the final date of first use for each child drug for which the respondent was a lifetime user were assigned.
3. For respondents who were lifetime nonusers of other hallucinogens, it was determined which, if any, child drug could have had the same date of first use as hallucinogens. Specifically, it was determined whether the date of first use for hallucinogens was between *earliest possible date* and *latest possible date* for each child drug. If none of the child drugs were eligible to receive the hallucinogens date of first use, nothing was done. Otherwise, one of the eligible child drugs was chosen at random, and its date of first use was overwritten with the hallucinogens date.

6.3.6.5 Final Age and Date-of-First-Use Variables

The final imputation-revised date-of-first-use variables for "other" variables were named in the same manner as those for cigarettes: with the prefix IR, followed by a three-letter code identifying the drug and the final three letters identifying the measure (AGE = age at first use, MFU = month of first use, YFU = year of first use, DFU = day of first use). Again, each IR variable was accompanied by an imputation indicator with the requisite II prefix.

6.3.6.6 Recodes for Additional Analyses

Section 6.3.5.7 discusses some prevalence recodes that incorporated information from several different drugs. Incidence recodes were also created that incorporated information from multiple drugs. These incidence recodes were created for only PSY, IEM, and SUM (Table 6.9).

The AFU recodes were simply set to the minimum of the source AFU variables, and they were named with the suffix AGE: PSYAGE2, IEMAGE, and SUMAGE. For example, PSYAGE2 = minimum of IRANLAGE, IRTRNAGE, IRSTMAGE, and IRSEDAGE.

To set the date-of-first-use variables, the earliest date of first use was found among the source variables for which the respondent was a lifetime user, and the new YFU, MFU, and DFU variables were determined using the YEAR, MONTH, and DAY functions in SAS. For example, PSYYFU2 = YEAR (minimum of dates of first use of pain relievers, tranquilizers, stimulants, and sedatives).

6.3.7 Special Section: Core-Plus-Noncore Methamphetamine and Stimulants Lifetime Use and Recency of Use (Imputation Sets 31 and 32)

New questions were added to the noncore special drugs module in the 2005 NSDUH to capture information from respondents who may have used methamphetamine but did not recognize it as a prescription drug and therefore did not report use in the core stimulants module. Additional follow-up items were added in the 2006 NSDUH to resolve inconsistencies between responses regarding methamphetamine in the core stimulants module and the noncore special drugs module. These additional methamphetamine questions asked about 12-month frequency, AFU, and date of first use.

Findings from the methamphetamine analysis report (Ruppenkamp et al., 2007) showed that it would be important to use responses from the noncore special drugs module in order to determine the best estimate of the prevalence of methamphetamine use from the NSDUH. Therefore, after the normal imputation processing of the drug variables was complete, new imputation-revised versions of lifetime use and recency-of-use variables for both methamphetamine and stimulants were created, which incorporated responses from the noncore special drugs module as well as the core module. Estimates of nonmedical use of psychotherapeutics, stimulants, and methamphetamine in the detailed tables include data from the new methamphetamine items added in 2005 and 2006 and are not comparable with estimates presented in NSDUH reports prior to the 2007 national findings report (Office of Applied Studies, 2008). This situation is unique for lifetime use and recency of use in that it incorporates information from the noncore section of the questionnaire. For more information on the reporting of methamphetamine prevalence in the 2014 NSDUH, see Section B.4.8 of Appendix B of the 2008 national findings (Office of Applied Studies, 2009).

A detailed description of the creation of these imputation-revised variables follows. The approach used was similar to the process used in normal processing with the following exceptions:

- The provisionally imputed values for lifetime use and recency of use for CPN stimulants and methamphetamine were used as the final imputation-revised variables.
- A different set of edited variables was used as the base for imputation.

The use of provisionally imputed values as the final imputation-revised variables for lifetime use and recency of use was implemented beginning in 2011 to eliminate the need to reimpute variables that were not used in subsequent analyses. Before 2011, lifetime drug use models were refit for stimulants, sedatives, cocaine, crack, and heroin, and provisional imputations were performed. After these models were refit, the lifetime use indicators for all drugs were reimputed using the PMN Type 3 methodology outlined in Section 3.4.3 to incorporate the noncore methamphetamine and stimulants questions. However, the only imputation-revised lifetime use questions used in further processing were the ones for stimulants and methamphetamine. Similarly, the imputation of CPN recency variables for stimulants and methamphetamine proceeded in the same manner as the core-only variables. This process included an RP and PRD step for recency of use, followed by a provisional imputation. A response-propensity adjustment and prediction model were then fit for 12-month frequency of use, and the final CPN recency and 12-month frequency variables were imputed in a final hot-deck step that incorporated additional noncore variables as logical constraints.

The simpler imputation methods were implemented after an impact assessment was conducted using 2010 data. For the results of this analysis, see Section 5.3.7 of the 2011 NSDUH MRB imputation report (Frechtel et al., 2013).

6.3.7.1 Final Creation of Base Variables for Imputation

The edited recency-of-use variables MTHREC06 and STMREC06, which were created during editing, were used as a starting point for the final creation of the base variables for imputation. These variables are described in Section 6.2.6.2. They are similar to METHREC and STIMREC, the edited recency-of-use variables used in normal processing, except that they incorporate responses from the noncore special drugs module and the core module.

The final base variable for imputation of lifetime use of methamphetamine was called EDMTHLIFE. It was created as follows:

EDMTHLIFE =

- 1 (lifetime user), if MTHREC06 was 1, 2, 3, 8, 9, 11, 12, or 13; else
- 2 (lifetime nonuser), if MTHREC06 was 81 or 91; else
- missing.

The final base variable for imputation of lifetime use of stimulants, EDSTMLIFE, was created in a similar manner.

The final base variable for imputation of recency of use of methamphetamine was called EDMTHREC. It was created as follows:

EDMTHREC =

- 1 (past month user), if MTHREC06 was 1, or if MTHREC06 was 11 and METHREC was not equal to 11; else
- 2 (past year but not past month user), if MTHREC06 was 2, or if MTHREC06 was 12 and METHREC was not equal to 12; else
- 3 (lifetime but not past year user), if MTHREC06 was 3 or 13; else
- MTHREC06.

Note that respondents who responded to the noncore recency question (most of those with MTHREC06 values of 11, 12, and 13) were treated identically to respondents who gave the same response to the core recency question (those with MTHREC06 values of 1, 2, and 3). This was done based on the decision to treat respondents to the noncore questions as item respondents eligible to be donors and therefore used to fit the models. This is an exception to the general rule that respondents with logically assigned responses were treated as item *non*respondents.

The final base variable for imputation of recency of use of stimulants, called EDSTMREC, was created in a similar manner.

6.3.7.2 Reimputation of Lifetime Use Indicators (Imputation Set 31)

Using EDMTHLIFE and EDSTMLIFE, the processing of the lifetime use indicators proceeded, as described in Section 6.3.1. The set of item respondents did not change between the original imputation of the lifetime indicators and the reimputation of the lifetime indicators. Therefore, it was not necessary to readjust the weights for item nonresponse. As shown in Table 6.5, the stimulants lifetime drug use indicator was modeled toward the end of the hierarchy. Rather than reimputing stimulants and all variables that came after it, lifetime models were refit for stimulants and methamphetamine only, and missing values were imputed in one univariate hot-deck step.

6.3.7.3 Reimputation of Recency of Use

Using EDMTHREC and EDSTMREC instead of METHREC and STIMREC, the processing of the recency data proceeded, as described previously.[108] Final recency-of-use variables for methamphetamine and stimulants were created.

6.3.7.4 Recodes for Additional Analyses

In the manner described in Section 6.3.5.7, some prevalence recodes were created that incorporated information from the noncore special drugs module. The CPN methamphetamine recodes were CPNMTHFG, CPNMTHYR, and CPNMTHMN. The CPN stimulants recodes were CPNSTMFG, CPNSTMYR, and CPNSTMMN. The CPN psychotherapeutic recodes were CPNPSYFG, CPNPSYYR, and CPNPSYMN. No CPN versions of the IEM or SUM recodes described in Table 6.9 were created for use in the detailed tables, even though the prevalence estimates would likely increase slightly if the noncore methamphetamine data were incorporated.

[108] Other core-plus-noncore edited variables also were used in these reimputation steps, in logical constraints of hot-deck steps.

7. Editing and Imputation for Variables from the NSDUH Noncore ACASI Modules

7.1 Introduction

The 2014 computer-assisted interviewing (CAI) instrument contained the following noncore self-administered modules, which used audio computer-assisted self-interviewing (ACASI; see Chapter 1):

- special drugs,
- risk/availability,
- blunts,
- substance dependence and abuse,
- special topics,
- marijuana purchases,
- prior substance use,
- substance treatment,
- health care,
- adult mental health service utilization (administered only to adults aged 18 or older),
- social environment (administered only to adults),
- parenting experiences (administered only to parent/legal guardian in dwelling units where a 12- to 17-year-old also was selected for an interview),
- youth experiences (administered only to youths aged 12 to 17),
- mental health (administered only to adults),
- adult depression and adolescent depression (the former administered to adults and the latter administered to youths),
- youth mental health service utilization (administered only to youths), and
- consumption of alcohol.

This chapter describes the content of the individual noncore ACASI modules. Except for the substance dependence and abuse module, no imputation for missing data was performed for variables in these modules. Therefore, the focus of the discussion in this chapter is on processing of the edited variables for these modules, along with any specific edits that were relevant to the data in a given module. The structure of this chapter also differs from the structure of other chapters beginning with Chapter 4 because only one set of noncore ACASI variables was imputed. In this chapter, subheadings in Section 7.4 focus on the specific noncore ACASI modules. Section 7.4.4 for the substance dependence and abuse module also discusses imputation procedures for the cigarette (i.e., nicotine) dependence items in this module.

7.2 Editing of Noncore ACASI Data Because of Patterned Responses

Two types of edits could be made to noncore ACASI data because of patterned responses in the interview. The first type involved edits because of patterned responses in corresponding variables from core modules (see Section 2.3.2 in Chapter 2). When a case was retained as a final respondent but with variables from one or more core modules being assigned the appropriate codes for bad data (Section 2.4.3), codes for bad data also were assigned to the corresponding variables in noncore modules. This edit was implemented regardless of whether the corresponding noncore variables had patterned responses. For example, if a respondent's variables in the pain relievers module were assigned codes for bad data because of patterned responses in that module and the respondent was routed to questions pertaining to pain relievers in the substance dependence and abuse module, codes for bad data were assigned to the pain relievers data in the substance dependence and abuse module.

The second type of edit involved edits because of other occurrences of patterned responses specifically in the noncore ACASI data. These edits were implemented for a small number of cases in 2014 (i.e., fewer than 20). Responses were replaced with bad data codes when these patterns occurred. In particular, answers were replaced with bad data codes if respondents keyed answers of "1" (or multiple-digit responses of "1," where applicable) to every question they were asked within a module. In the youth experiences module, adolescents who keyed answers of "2" wherever possible also had their answers replaced with bad data codes. For respondents to key all possible items as "2" in the youth experiences module would mean that they keyed this response to 42 items in a row, including that they had engaged in all "problem" activities (such as getting into fights, carrying a handgun, or selling illegal drugs) exactly "1 or 2 times" in the past year. These respondents who keyed "2" wherever possible also would have participated in exactly two school-based activities, two community-based activities, two church- or faith-based activities, and two other kinds of activities in the past year.

7.3 "Lockout" Following Completion of ACASI Modules

Since 2003, once the ACASI portion of the interview had been completed, interviewers and respondents are locked out of the ACASI section of the interview to provide additional protection of the privacy of respondents' answers. Based on answers that respondents gave in the interviewer-administered questions following the ACASI section, however, interviewers could go back and change information in core demographics questions that were in the interviewer-administered section that preceded the ACASI questions. For example, interviewers could discover when they were completing the household roster section of the interview that a female respondent's gender had been incorrectly keyed as male, or vice versa (Section 4.2.4). The interviewer was allowed to change the information about the respondent's gender at the beginning of the interview.

If interviewers went back and changed answers to core demographic questions, however, the CAI program did not reroute respondents back into the ACASI section to answer questions that previously had been skipped. The CAI program also did not update the answers in the

ACASI section to remove data from questions that logically should have been skipped based on any new information that the interviewers entered. This "lockout" feature was particularly relevant to editing of data in the modules for health care (Section 7.4.9), parenting experiences (Section 7.4.12), and consumption of alcohol (Section 7.4.17).

7.4 Editing and Imputation Procedures for Specific Noncore ACASI Modules

7.4.1 Special Drugs Module

The special drugs module asked about the smoking and sniffing of heroin; use of heroin, methamphetamine, other stimulants, cocaine, or other drugs with a needle for nonmedical reasons; general needle use behaviors (e.g., needle sharing); and where respondents got the last needle that they used. Questions also have been included in the module since 2006 for additional drugs that did not have specific questions in any of the core drug modules.

As noted in Sections 6.1.3 and 6.2.6 in Chapter 6, new questions were added to the special drugs module in 2005 and were updated in 2006 to capture information about methamphetamine use from respondents who did not report methamphetamine use in the core stimulants module (e.g., if they may not have recognized it as a stimulant in the context of questions about prescription stimulants). Details about the logic for asking these follow-up questions about methamphetamine use are discussed in Section 6.2.6.

Additional questions have been included in the special drugs module since 2007 that ask about first use of methamphetamine and frequency of use in the past 12 months. However, these items on first use of methamphetamine and 12-month frequency were not used in editing the methamphetamine recency variable MTHAREC that was described in Section 6.2.6. Thus, MTHAREC was created in a manner comparable with how it was created since 2006 to maintain data on trends in methamphetamine use. However, inconsistencies could exist between MTHAREC and these new variables, such as if respondents reported that they initiated methamphetamine use in the past 12 months but they reported last using methamphetamine more than 12 months ago.

7.4.1.1 Noncore Methamphetamine Initiation and 12-Month Frequency

A key aspect of processing the noncore variables about first use of methamphetamine and frequency of use in the past 12 months involved assignment of an appropriate code of 91 (or 991, etc.) if respondents never used methamphetamine (Section 2.4.2). In addition, a code of 93 (or 993) was assigned to the 12-month frequency variables if respondents last used methamphetamine more than 12 months ago.

Respondents who confirmed methamphetamine use in the consistency check questions (SD17A1 or SD17ALT) were asked to report their age when they first used methamphetamine (edited variable MTHAAGE). Respondents who first used within 1 year of their current age were asked to report the year and month when they first used methamphetamine (edited variables MTHAYFU and MTHAMFU, respectively). The CAI logic triggered consistency checks if respondents reported a month of first use (MFU) and year of first use (YFU) that was inconsistent with their reported age at first use (AFU). Thus, these items were analogous to those that asked about first use of drugs in the core modules (Section 6.2.5). Consequently, the same edits were applied to these methamphetamine variables as were applied to the corresponding questions in the core modules that were described in Section 6.2.5.

Additionally, respondents who reported using methamphetamine in the past 12 months in questions SD17B or SD18B and who did not report that an earlier answer that they never used methamphetamine was correct (as reported in the core stimulants module) were asked to report on their frequency of use of methamphetamine in the past 12 months. As in the core drug modules, respondents could report their methamphetamine frequency of use in one of three ways:

- use on an average number of days per week in the past 12 months,
- use on an average number of days per month in the past 12 months, or
- the total number of days they used in the past 12 months.

As for the 12-month frequency variables in the core modules (Section 6.2.4), a single composite edited variable MTHAYTOT was created to translate these answers to a total number of days that respondents used methamphetamine in the past 12 months. In the same manner as for 12-month frequency variables in the core drug modules, the value in MTHAYTOT was prorated as needed for respondents who initiated use in the past 12 months and therefore were not users over the entire 12-month period (see Table B.5 in Appendix B).

In addition, the CAI logic for asking about first use of methamphetamine and the 12-month frequency of use differed slightly with respect to the consistency checks SD17A1 and SD17ALT that were triggered when respondents reported in the core stimulants module that they never used methamphetamine but reported use in the special drugs module. Specifically, the only respondents who were eligible to be asked about first use of methamphetamine in the special drugs module were those who specifically reported in the core stimulants module that they never used methamphetamine and then explicitly confirmed their methamphetamine use in SD17A1 or SD17ALT. However, respondents who reported last using methamphetamine in the past 12 months were eligible to be asked about their 12-month frequency, as long as they had not reported in SD17A1 or SD17ALT that an earlier report of having never used stimulants was correct (Section 6.2.6); this included situations in which SD17A1 and SD17ALT had been skipped because respondents reported in the core stimulants module that they did not know whether they ever used methamphetamine (i.e., and therefore had not previously denied ever using the drug).

7.4.1.2 Heroin Use and Needle Use Variables

An important aspect of the processing of variables in this module pertaining to the smoking of heroin, the sniffing of heroin, or needle use consisted of assigning codes of 91, 93, or 99 (Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, respondents who never used heroin, stimulants other than methamphetamine, or cocaine were not asked questions in the special drugs module that pertained to these drugs. Similarly, respondents who did not indicate use of methamphetamine in the core stimulants module and who continued to indicate in the follow-up questions in the special drugs module that they never used methamphetamine did not need to be asked further questions pertaining to the use of methamphetamine with a needle. In addition, respondents who indicated that they never used heroin, methamphetamine (in both the core stimulants module and in the noncore special drugs module), stimulants other than methamphetamine, cocaine, or any other drug with a needle for nonmedical reasons did not need to be asked questions about general needle use behaviors or the source of the last needle they used. Table B.13 in Appendix B describes specific edits that were implemented for these variables when the corresponding items were skipped.

Respondents who reported in the core heroin module that they were lifetime heroin users but who answered "no" to all questions about smoking heroin (question SD01), sniffing heroin (question SD03), or using it with a needle (question SD08) were asked a follow-up question SDHEUSE to determine how these respondents administered the heroin they had reported using. SDHEUSE was an "enter all that apply" type of question that allowed respondents to report multiple ways that they used heroin. SDHEUSE included response options for smoking heroin, sniffing heroin, using heroin with a needle, or use of heroin "some other way." Respondents who reported using heroin "some other way" were asked to specify in question SDHEUSE2 what this "other" mode of heroin administration was. If respondents specified using heroin in a way that corresponded to one of the ways that they had been asked about, it was logically inferred that the response in the edited SDHEUSE question should have been chosen. For example, if respondents did not report in SDHEUSE that they smoked heroin but they specified this as "some other way," the edited variable HEOTSMK was assigned a code of 3 (Response entered LOGICALLY ASSIGNED).

In turn, HEOTSMK, HEOTSNF, and HEOTNDL were used to edit the corresponding lifetime heroin variables HERSMOKE, HERSNIFF, and HERNEEDL, respectively. For example, if HEOTSMK indicated that the respondent had smoked heroin (HEOTSMK = 1 or 3), HERSMOKE was edited to indicate that the respondent was logically inferred to have smoked heroin at least once in his or her lifetime.

In addition, respondents were routed to follow-up recency questions for smoking, sniffing, or using heroin with a needle if they reported using heroin in any of these ways in question SDHEUSE. Information from these follow-up questions was used in the creation of the heroin smoking, sniffing, or needle recency variables HRSMKREC, HRSNFREC, and HRNDLREC. However, if respondents initially did not report using heroin in these ways in SDHEUSE, they were skipped out of these follow-up recency questions for smoking, sniffing, or using heroin with a needle. Therefore, if respondents' only indication of smoking, sniffing, or using heroin with a needle came from the "OTHER, Specify" response associated with SDHEUSE, the corresponding variables HRSMKREC, HRSNFREC, or HRNDLREC were set to 9 (Used at some point in the lifetime LOGICALLY ASSIGNED).

Table B.14 describes specific edits other than those pertaining to skip logic that were implemented in the special drugs module for heroin use and use of drugs with a needle.[109] For

[109] Issues related to the editing of the needle use variables for methamphetamine and stimulants are discussed in Table B.11 in the context of other issues for editing the noncore special drugs variables that are predecessors to the edited core-plus-noncore recency variables for these substances.

example, lifetime users of heroin could report that they smoked heroin at least once but not indicate when they last smoked it. The general edit was to assign a nonspecific value to the edited recency variable (i.e., HRSMKREC) to indicate that the respondent smoked heroin at some point in his or her lifetime. In some special situations, however, it was possible to infer that respondents could not have smoked heroin in the past 12 months. In these situations, respondents reported last using heroin more than 12 months ago, and there were no responses for other heroin-related questions in the special drugs module to indicate that these respondents had used heroin in the past 12 months. Since 2001, respondents have been asked in question SD05 (edited variable OTDGNEDL) whether they ever used a needle to inject "some other drug" with a needle (if respondents previously reported lifetime use of heroin, methamphetamine, other stimulants, or cocaine with a needle) or "any drug" with a needle (if respondents had not previously indicated use of any of the above drugs with a needle). If question SD05 was answered as "yes" (OTDGNEDL = 1), respondents then were asked to specify what (other) drug(s) they used with a needle. Respondents could specify up to five drugs that they had injected (edited variables OTDGNDLA through OTDGNDLE).

Consequently, it was possible for respondents to have reported in a core drug module that they never used a particular drug that was covered in the special drugs module but then specify use of that drug with a needle in OTDGNDLA through OTDGNDLE. For example, respondents could indicate in the core heroin module that they never used heroin but then specify lifetime injection of heroin in OTDGNDLA through OTDGNDLE. In this situation, no editing was done to the core drug data. However, these respondents were logically inferred in the relevant special drugs variables to be users of that particular drug at some point in the lifetime. Thus, for example, if respondents reported in the core heroin module that they never used heroin, but then they specified heroin as "some other drug" that they used with a needle, the edited lifetime heroin needle use variable HERNEEDL was assigned a code of 3 (Yes LOGICALLY ASSIGNED) and the heroin needle recency variable HRNDLREC was assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED).

Respondents also could report in the special drugs module that they used a needle to inject a drug for nonmedical reasons, even though they previously reported that they never used marijuana, cocaine, heroin, hallucinogens, inhalants, prescription pain relievers, prescription tranquilizers, prescription stimulants, or prescription sedatives. Beginning in 2001, however, respondents were asked to specify what "other" drug(s) they had injected. Thus, it was possible to identify respondents who corroborated their report of lifetime injection drug use (e.g., if injection of anabolic steroids was reported) despite having previously reported nonuse of all drugs covered in the core modules. Similarly, it was possible from "OTHER, Specify" data on other drugs that respondents had injected to identify those whose needle use probably was limited to legitimate, medical uses (e.g., injection of antibiotics). Therefore, it was logically inferred that respondents had never used needles for nonmedical purposes if (1) they were lifetime nonusers of all drugs covered in the core; (2) they indicated that they never engaged in behaviors that would be indicative of nonmedical needle use, such as needle sharing, use of bleach to clean needles, or reusing of needles; and (3) all of the "other" drugs they reported using with a needle typically were not drugs of abuse. In this situation, the variable OTDGNEDL corresponding to question SD05 was set to a value of 4 (No LOGICALLY ASSIGNED). Any data in the general needle use variables GNNDREUS through GNNDGET were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned) to indicate that respondents logically should have skipped these items because they appeared to be reporting about legitimate use of drugs with a needle.

7.4.1.3 Recoded Needle Use Variables

Recoded needle use variables (STNEELDR, STNLRECR, and NEDLRECR) were created from variables pertaining to use of methamphetamine, other stimulants, heroin, or cocaine with a needle. STNEEDLR and STNLRECR were analogous to the variables STNEEDLE and STNDLREC, respectively, that existed in 1999 and 2000. Similarly, NEDLRECR was analogous to the recoded needle recency variable NEDLRECC that existed in 1999 and 2000. These variables did not take into account data from the methamphetamine variables that have been present in the special drugs module since 2005. Consequently, STNEEDLR, STNLRECR, and NEDLRECR were designed to be comparable with data in these variables prior to 2005 for use in analysis of trends in needle use.

STNEEDLR was created from the lifetime methamphetamine needle use variable MTHNEEDL and the lifetime other stimulant needle use variable OSTNEEDL. The logic for creating STNEEDLR is documented below.

- If respondents reported use of methamphetamine or other stimulants with a needle, STNEEDLR was coded as 1 ("yes").
- If respondents were users of methamphetamine or other stimulants but reported never injecting stimulants with a needle, STNEEDLR was coded as 2 ("no").
- If respondents had never used stimulants, STNEEDLR was coded as 81 or 91 (depending on the value in the core stimulant recency variable STIMREC). Missing data from MTHNEEDL or OSTNEEDL were carried over to STNEEDLR.

STNLRECR was derived from the needle recency variables MTNDLREC (most recent use of methamphetamine with a needle) and OSTNLREC (most recent use of other stimulants with a needle). The logic for creating STNLRECR is documented below.

- If respondents had never used stimulants, STNLRECR was assigned a code of 81 or 91, depending on the value in STIMREC.
- Similarly, if STNEEDLR was coded as 2 ("no"), STNLRECR was coded as 93 (used stimulants but never with a needle).
- If respondents had injected methamphetamine or some other stimulant, the general principle in assigning a value to STNLRECR was to pick the most recent use that the respondent reported. For example, if respondents indicated in either MTNDLREC or OSTNLREC that they used methamphetamine or other stimulants with a needle within the past 30 days, then STNLRECR indicated use of stimulants with a needle in the past 30 days.
- However, if respondents reported last using methamphetamine with a needle more than 30 days ago but within the past 12 months but all that was known was that they used other stimulants with a needle at some point in their lifetime, it still could be inferred that they had used a needle to inject any stimulant at some point in the past

12 months (potentially including the past 30 days). The nonspecific value for past year use was assigned (i.e., a code of 8) because the respondents could have been more recent users of other stimulants with a needle.

- Similarly, if respondents indicated use of one of these stimulants (i.e., methamphetamine or other stimulants) with a needle in a definite period more than 30 days ago but the respondents did not know or refused to indicate whether they had ever used the other type of stimulant with a needle, a nonspecific value of lifetime use (i.e., a code of 9) was assigned to STNLRECR because the respondents may have been more recent users of stimulants with a needle than what they had reported.

The needle recency variable NEDLRECR was created from the variables HRNDLREC (most recent use of heroin with a needle), CONDLREC (most recent use of cocaine with a needle), and STNLRECR (most recent use of any stimulant with a needle, as noted above). The logic for creating NEDLRECR is documented below.

- If respondents never used a needle to inject any of these drugs nonmedically (including situations in which respondents never used cocaine, heroin, or stimulants), NEDLRECR indicated that the respondents had never used cocaine, heroin, or stimulants with a needle.
- If respondents reported using one or more of these drugs with a needle, the general principle in assigning a value to NEDLRECR was to identify the most recent use reported by the respondents. In particular, if respondents reported using one or more of these drugs with a needle in the past 30 days, it could be determined unambiguously that the respondents were past month needle users.
- In other situations, however, if one or more of the cocaine, heroin, or stimulant needle recency variables indicated nonspecific use at some point in the respondents' lifetime, NEDLRECR was assigned a value to indicate nonspecific past year or lifetime use. For example, if respondents reported last using cocaine with a needle more than 30 days ago but within the past 12 months, yet all that was known was that they used heroin with a needle at some point in their lifetime, it still could be inferred that they had used some drug with a needle in the past 12 months. The nonspecific value for past year use was assigned (i.e., a code of 8) because the respondents could have been more recent users of heroin with a needle.
- Similarly, if respondents indicated use of one of these drugs with a needle more than 12 months ago and they did not know or refused to indicate when they last used one of the other drugs with a needle, a nonspecific value of lifetime use (i.e., a code of 9) was assigned to NEDLRECR because the respondents may have been more recent needle users than what they had reported elsewhere.

In addition, recoded needle use variables (STNEDL06, STNLRC06, and NDLREC06) were created to incorporate the methamphetamine needle use questions that were added to the module in 2005 and the follow-up items on methamphetamine use that were added to the module in 2006 (and retained since then). These new needle use variables were based on the variables STNEELDR, STNLRECR, and NEDLRECR (documented above). Consequently, data in

STNEDL06, STNLRC06, and NDLREC06 may not be comparable with data in STNEEDLR, STNLRECR, and NEDLRECR from prior survey years.

The recoded lifetime stimulant needle variable STNEDL06 was derived from the variable STNEEDLR. STNEDL06 was updated to reflect data from the new variable MTHANEDL (ever used a needle to inject methamphetamine), after MTHANEDL had been edited (see Section 6.2.6 and Table B.11 in Appendix B). The logic for creating STNEDL06 is documented below. This logic for creating STNEDL06 was implemented only if (1) MTHEVCK had been set to bad data (see Table B.11), (2) respondents had specified use of methamphetamine as "some other drug" that they used with a needle, or (3) respondents had not previously reported methamphetamine use in the core stimulants module because they had not thought of methamphetamine as a prescription drug or for some similar reason (e.g., knowing methamphetamine by a slang term other than the examples given in the core).

- If MTHANEDL indicated use of methamphetamine with a needle and STNEEDLR did not already indicate use of any stimulant with a needle, then STNEDL06 was recoded as 1 ("yes").
- If STNEEDLR had been coded as 81 or 91 (i.e., never used stimulants) and MTHAMP indicated lifetime methamphetamine use but MTHANEDL did not indicate use with a needle, then STNEDL06 was recoded as 2 ("no"), instead of 81 or 91.
- Otherwise, if MTHANEDL had a missing value and STNEEDLR did not indicate that respondents used stimulants with a needle, STNEDL06 was assigned a missing value.

The recoded stimulant needle recency variable STNLRC06 was derived from the variable STNLRECR and was updated to reflect data from MTANDLRC (most recent use of methamphetamine with a needle), after MTANDLRC had been edited. The logic for creating STNLRC06 is documented below. The same conditions described above for implementing the logic for STNEDL06 also applied to STNLRC06.

- If respondents indicated last using methamphetamine with a needle within the past 30 days in MTANDLRC and STNLRECR did not already indicate use of stimulants with a needle in the past 30 days, then STNLRC06 was recoded to indicate use of stimulants with a needle in the past 30 days.
- If respondents indicated needle use in MTANDLRC in some period other than the past 30 days but STNLRECR indicated that respondents either had never used stimulants or had never used stimulants with a needle (i.e., based on core stimulants data and special drugs variables that existed prior to 2005), then STNLREC06 was updated with the period of most recent use that the respondents indicated in MTANDLRC.
- If respondents reported last using methamphetamine with a needle more than 30 days ago but within the past 12 months (in MTANDLRC) but all that was known was that the respondents used other stimulants with a needle at some point in their lifetime, it still could be inferred that the respondents had used a needle to inject any stimulant at

some point in the past 12 months (potentially including the past 30 days). The nonspecific value for past year use was assigned (i.e., a code of 8) because respondents could have been more recent users of other stimulants with a needle.

- Similarly, if respondents indicated use of methamphetamine (from MTANDLRC) or other stimulants with a needle in a definite period more than 30 days ago and the respondents did not know or refused to indicate whether they had ever used the other type of stimulant, a nonspecific value of lifetime use (i.e., a code of 9) was assigned to STNLRC06 because the respondents may have been more recent users of stimulants with a needle than what they had reported.
- If MTANDLRC had a definite value indicating most recent use more than 30 days ago, STNLRECR was coded as 9 (i.e., used at some point in the lifetime), but OSTNEEDL indicated that the respondents had not used other stimulants with a needle, then the STNLRC06 was assigned the value from MTANDLRC plus a value of 10. For example, if MTANDLRC was coded as 3 (last used methamphetamine with a needle more than 12 months ago) and STNLRECR was coded as 9, then STNLREC06 was assigned a value of 13. In this situation, the code of 9 in STNLRECR would have come from the methamphetamine data that existed in the special drugs module prior to 2006, and not from data for other stimulants.
- If STNLRECR indicated that that respondents had never used stimulants (i.e., STNLRECR = 91) but MTANDLRC indicated that they had used methamphetamine but never with a needle, then STNLREC06 was recoded to a value of 93 (USED STIMULANTS BUT NEVER WITH A NEEDLE).
- If STNLRECR was coded as 91 (see above) and MTANDLRC also indicated that respondents had never used methamphetamine, then STNLREC06 continued to be coded as 91.

Because of the addition of the new methamphetamine variables, the needle recency variable NDLREC06 was derived from the variable NEDLRECR (see above). NDLREC06 was updated based on values in STNLRC06. The logic for creating NDLREC06 is documented below.

- If respondents indicated last using stimulants with a needle within the past 30 days based on STNLRC06, and NDLREC06 did not already indicate use in that period, then NDLREC06 was recoded to indicate use of cocaine, heroin, or stimulants with a needle in the past 30 days.
- If STNLRC06 indicated that respondents used stimulants with a needle at some point in the past 12 months (STNLRC06 = 8) and NEDLRECR did not indicate use in the past 30 days or NEDLRECR was not already coded as 8 (i.e., used at some point in the past 12 months), then NDLREC06 was recoded as 8.
- Similarly, if STNLRC06 indicated that respondents used stimulants with a needle at some point in their lifetime (STNLRC06 = 9) and NEDLRECR did not indicate use in the past 12 months or NEDLRECR was not already coded as 9 (i.e., use at some point in the lifetime), then NEDLREC06 was recoded as 9.

- If respondents reported using stimulants with a needle (from STNLRC06) in some definite period other than the past 30 days, the general principle in assigning a value to NDLREC06 was to identify the most recent use reported by the respondents. For example, STNLRC06 indicated that respondents last used stimulants with a needle more than 30 days ago but within the past 12 months but NEDLRECR indicated that respondents last used a needle to inject cocaine, heroin, or stimulants more than 30 days ago but within the past 12 months, then NDLREC06 was recoded to reflect the value from STNLRC06.
- If respondents were logically inferred to have last used stimulants with a needle more than 12 months ago (STNLRC06 = 13) and NEDLRECR was coded as 9 (i.e., used at some point in the lifetime) but respondents had never used cocaine or heroin with a needle, then NEDLREC06 was recoded as 13.

If STNLRC06 had a missing value and NEDLRECR indicated needle use at some point in the past 12 months (but not in the past 30 days), then NDLREC06 was recoded as 8 (i.e., used at some point in the past 12 months). Similarly, if STNLRC06 had a missing value and NEDLRECR indicated needle use but not in the past 12 months, then NDLREC06 was recoded as 9 (i.e., used at some point in the lifetime).

7.4.1.4 Additional Drug Use Variables

Questions have been included in the special drugs module since 2006 for the following drugs that did not have specific questions in any of the core drug modules:

- GHB (gamma hydroxybutyrate);
- Adderall® (a prescription stimulant) used without a prescription or for the experience or feeling it caused (i.e., "nonmedical" use);
- Ambien® (a prescription sedative) used without a prescription or for the experience or feeling it caused;
- nonprescription cough or cold medicines, also known as over-the-counter (OTC) medicines, to get high;
- ketamine (a hallucinogen);
- the hallucinogens DMT (dimethyltryptamine), AMT (alpha-methyltryptamine), and 5-MeO-DIPT (5-methoxy-diisopropyltryptamine); and
- *Salvia divinorum* (a hallucinogen).

Respondents were shown images on-screen of Adderall® and Ambien® to aid them in recalling whether they had ever used these drugs nonmedically.

If respondents reported that they ever used a particular drug mentioned above, they were asked to report when they last used that drug. In addition, respondents who reported that they last used OTC cough or cold medicines to get high in the past 12 months were asked to specify what OTC drugs they used to get high in that period. Respondents could specify up to five drugs that they used to get high. Therefore, the edited variables COLDYR1 through COLDYR5 (corresponding to the "OTHER, Specify" items SD27a through SD27e) captured up to five medications that respondents reported using to get high.

The "OTHER, Specify" variables for OTC cough and cold medicines were coded using a decimal system to preserve relationships between similar responses and to capture additional details that respondents provided. For example, there are multiple formulations of OTC cold medicines under the brand name of Robitussin® that vary in terms of the specific ingredients they contain, such as the OTC cough suppressant dextromethorphan (DXM). Therefore, if respondents specified that they used a Robitussin® product to get high in the past 12 months but the only detail they provided was that they used "Robitussin," that response was assigned a default code of 829.0 (Robitussin®, no other information). However, if respondents specifically reported that they used Robitussin® DM to get high (a product that contains DXM), that response was coded as 829.1 and was documented as Robitussin® DM. Thus, the main code of 829 indicated use of a Robitussin® product to get high in the past 12 months, and the decimal value indicated whether respondents provided additional detail about their use of specific Robitussin® products to get high.

An important aspect of the processing of these drug variables consisted of assigning a code of 91 (or 9991) to variables that had been skipped because the questions did not apply. For example, if respondents reported in question SD19 (edited variable GHB) that they never used GHB, the edited recency-of-use variable GHBREC (corresponding to question SD24) was assigned a code of 91. Similarly, if respondents reported that they last used an OTC cough or cold medicine to get high more than 12 months ago, COLDYR1 through COLDYR5 were assigned a code of 9993.0 to indicate that respondents had used OTC cough or cold medicines to get high, but not in the past 12 months.

In addition, if respondents reported that they had ever used a drug (e.g., GHB) but they did not know or refused to report when they last used it, the edited recency-of-use variable (e.g., GHBREC) was assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). If the most recent use of OTC cough or cold medicines (edited variable COLDREC, corresponding to question SD27) had been assigned a code of 9, COLDYR1 through COLDYR5 were left as blank (code of 9998.0) because some of these respondents may have used OTC cough or cold medicines to get high in the past 12 months.

Data from the "OTHER, Specify" needle use variables OTDGNDLA through OTDGNDLE and the cough and cold medicine variables COLDYR1 through COLDYR5 were used in editing these additional drug use variables, except for editing of the OTC cough and cold variables COLDMEDS (corresponding to question SD22A) and COLDREC. Specifically, OTDGNDLA through OTDGNEDLE were not used to edit COLDMEDS and COLDREC because there could be too much room for interpretation regarding what might constitute a nonprescription cough or cold medicine. In contrast, the questions for the other additional drugs in the special drugs module pertained to specific drugs.

Thus, for example, if a respondent had not reported lifetime use of ketamine but the respondent specified using ketamine with a needle in OTDGNDLA through OTDGNDLE, then the edited variable KETAMINE (corresponding to question SD23a) would be assigned a code of 3 (Yes LOGICALLY ASSIGNED). In turn, the recency-of-use variable KETAREC (corresponding to question SD28) would be assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). However, there were no situations in 2014 in which data from the "OTHER, Specify" needle use variables OTDGNDLA through OTDGNDLE resulted in any responses being logically inferred in these additional drug use variables. Nevertheless, this logic is in place for handling these types of patterns for GHB in future survey years.[110]

In addition, if a respondent specified use in the past year of GHB, Adderall®, Ambien®, ketamine, DMT, AMT, 5-MeO-DIPT, or *Salvia divinorum* as a nonprescription cough or cold medicine in COLDYR1 through COLDYR5, that response was used to edit the other drug data. Using Ambien® as an example, if question SD21 pertaining to lifetime nonmedical use of Ambien® was not answered as "yes" (or had not logically been inferred as "yes" from the above edits based on OTDGNDLA through OTDGNDLE), the corresponding edited variable AMBIEN was assigned a code of 3 (Yes LOGICALLY ASSIGNED). In turn, the corresponding edited recency-of-use variable (e.g., AMBIREC, corresponding to question SD26) was assigned a code of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) because COLDYR1 through COLDYR5 pertain to use in the past 12 months. If a respondent had answered question SD21 for Ambien® as "yes" and reported last using it more than 12 months ago but also specified using it in the past 12 months in COLDYR1 through COLDYR5, AMBIREC also was assigned a code of 8.

Because the questions corresponding to COLDMEDS and COLDREC asked specifically about use of OTC cough or cold medicines, respondents also would be expected to report at least one OTC medicine in COLDYR1 through COLDYR5 or to have some response of "don't know" or "refused" in the corresponding items (i.e., if respondents did not know the name of the drug or did not want to report it). Therefore, a variable called OTCFLAG was created to indicate when respondents' reports of use of OTC cough or cold medicines in the past year were not substantiated by the "OTHER, Specify" data. OTCFLAG had the values shown below when these situations occurred.

- OTCFLAG had a value of 1 (i.e., "yes") when respondents specified only the use of drugs that require a prescription in the United States[111] or drugs that are not available as OTC medicines in the United States, even if they may be available as OTC medicines elsewhere.
- OTCFLAG had a value of 2 (i.e., "no") when COLDYR1 through COLDYR5 had values only of 9985.0 (bad data) or 9998.0 (blank).

No further editing was done to the OTC variables when OTCFLAG indicated that respondents' reports of use of OTC cough or cold medicines to get high in the past year were not substantiated in COLDYR1 through COLDYR5. However, OTCFLAG would alert analysts when this issue occurred.

---

[110] Questions about use of GHB will remain in the noncore special drugs module for 2015. Because questions about nonmedical use of Adderall® or Ambien® or use of ketamine, DMT, AMT, 5-MeO-DIPT, or *Salvia divinorum* will be asked in core modules for 2015, this edit will no longer apply to these other drugs in future years.

[111] If respondents reported the use of drugs that may require a prescription in some locations in the United States but may be available without a prescription elsewhere in the United States (e.g., cough medicines containing low doses of codeine), the medication was assumed to have been obtained without a prescription.

7.4.2 Risk/Availability Module

The risk/availability module asked about the perceived risk of harm associated with use of alcohol or specific illegal drugs, perceived ease of obtaining different illegal drugs, whether respondents were approached by someone in the past 30 days who was trying to sell an illegal drug, and general risk-taking types of behaviors. The latter included questions on the frequency with which respondents got a "kick out of doing things that are a little dangerous," how often they tried to test themselves "by doing something a little risky," and their frequency of seatbelt use.

Minimal processing of data was done to variables in this section. Specifically, the unedited variables were assigned final, mnemonic variable names (e.g., RSKPKCIG corresponding to question RK01A, which asked about the perceived risk of harm associated with smoking one or more packs of cigarettes per day). No further editing or processing was done to the variables in this module.

7.4.3 Blunts Module

Since 2004, all respondents have been asked questions about smoking part or all of a cigar with marijuana in it, commonly referred to as a "blunt." Two new questions also were added to the noncore blunts module in 2013 about use of marijuana that had been recommended by a doctor or other health care professional (subsequently referred to as medical marijuana use). In 2014, the logic for asking the questions about medical marijuana use was updated to include respondents who confirmed that they smoked blunts in the past 30 days despite reporting in the core marijuana module that they last used marijuana more than 12 months ago.

7.4.3.1 Use of Blunts

Respondents first were asked if they ever smoked part or all of a cigar with marijuana in it, sometimes called a "blunt." If respondents were lifetime users of blunts, they were asked how long it had been since they had smoked a blunt. If respondents indicated blunt use in the past 30 days, they were asked to report the number of days they smoked blunts in the past 30 days. If respondents reported use of both cigars (from the tobacco module) and blunts in the past 30 days, they also were asked if they had smoked a cigar *without* marijuana in it during the past 30 days.

In addition, the module included consistency checks in selected situations where respondents gave reports of blunt use that were inconsistent with what they had reported in the core marijuana module. Specifically, respondents could have reported earlier in the interview that they never used marijuana or hashish and then indicate that they had smoked a blunt in the past 30 days. Similarly, respondents could have reported that their last use of marijuana was more than 30 days ago and then indicate use of blunts in the past 30 days. In these situations, respondents were asked why they had reported earlier that they had never used marijuana or hashish or that they had not used it in the past 30 days, respectively.

An important aspect of the processing of variables in this module pertaining to use of blunts consisted of assigning codes of 91 and 99 (Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, if respondents said they never smoked part or all of a blunt, they were skipped out of the remaining blunts questions. Thus, respondents were assigned a code of 91 (NEVER USED BLUNTS) to the variables BLNTREC (most recent use of a blunt, corresponding to question BL02) and BLNT30DY (frequency of use of blunts in the past 30 days, corresponding to question BL02A) if they reported that they had never smoked blunts. Similarly, if respondents refused to report whether they had ever smoked part or all of a blunt, that refusal code was assigned to BLNTREC and BLNT30DY as well. In addition, if respondents reported being lifetime users of blunts but reported that they last used blunts more than 30 days ago, question BL02a would have been skipped. Therefore, BLNT30DY was assigned a code of 93 (DID NOT USE BLUNTS IN THE PAST 30 DAYS) in this situation.

Minimal editing was done to the variables BLNT30C1, BLNT30C2, RSNOMRJ, and RSNMRJMO, corresponding to questions BL03, BL04, BL05, and BL06, respectively; these variables pertained to the above-mentioned consistency checks between respondents' answers in the blunts and marijuana modules. If respondents reported not using blunts in the past month or reported that they were a lifetime marijuana user, questions BL03 and BL05 would have been skipped. Therefore, BLNT30C1 and RSNOMRJ were assigned a code of 99 in this situation. Similarly, if respondents were not past month users of blunts or reported last using marijuana in the past 30 days in the core marijuana module (i.e., a response in the marijuana module that was consistent with reported use of blunts in the past 30 days), then questions BL04 and BL06 would have been skipped. Therefore, BLNT30C2 and RSNMRJMO were assigned a code of 99 in these situations.

If respondents were routed to question BL03 (because they reported past month use of blunts but previously reported never using marijuana) and then they refused to answer question BL03, a refusal was propagated from BLNT30C1 to RSNOMRJ. If respondents answered BL03 as "don't know," RSNOMRJ retained a code of 98. Similarly, if respondents were routed to BL04 (because they reported past month use of blunts but previously reported using marijuana but not in the past 30 days) and they refused to answer, a refusal was propagated from BLNT30C2 to RSNMRJMO. Respondents answering "don't know" to question BL04 retained a code of 98 in RSNMRJMO.

In addition, since 2005, respondents who were lifetime users of blunts have been asked how old they were when they first used blunts (BL08). If respondents reported first using blunts within 1 year of their current age, they were asked to report the specific year and month when they first used, with the allowable years ranging from 2012 to 2014. If respondents reported first using blunts at their current age, their last birthday was in the current year, and they were interviewed after their last birthday, the CAI program assumed that the first use of blunts occurred in the current year (i.e., 2014). These respondents were asked only for the month when they first used in the current year. The remaining respondents who first used blunts within 1 year of their current age could be routed to one of two possible questions on the specific year they first used. They then were routed to a question to report on the specific month they first used blunts in the year they had reported previously.

Because the routing logic to the different versions of the MFU and YFU questions was mutually exclusive, composite sets of MFU and YFU variables (BLNTYFU and BLNTMFU, respectively) were created from the individual unedited variables. In addition, the final YFU variables were recoded to replace codes of 1 or 2 from the questionnaire that denoted specific years with values for the years (i.e., 2012 through 2014). In situations in which the CAI program skipped the YFU questions because it could be logically inferred that first use of blunts occurred in the interview year (i.e., 2014), a code of "2014" was assigned to BLNTYFU; this was done even if respondents did not know what month they first used in the current year, or if they refused to report what month they first used in the current year. If the MFU and YFU questions had been skipped because respondents first used blunts at ages that were more than 1 year below their current ages, legitimate skip codes were assigned to the BLNTYFU and BLNTMFU variables.

A consistency check also was included for situations in which the apparent age when respondents first used blunts that was calculated from their MFU and YFU and their date of birth (CAI variable MYR1STBL) was inconsistent with the age that respondents directly reported for when they first used blunts (CAI variable AGE1STBL). For example, the consistency check was triggered if a 16-year-old respondent reported first using blunts at age 16 but then reported first using blunts in a year and month that would have meant the respondent was 15 years old when he or she first used blunts. No further editing needed to be done if respondents indicated twice in a row that the AFU that was calculated from the MFU and YFU was correct. The CAI program updated the value for AGE1STBL (corresponding to the edited variable BLNTAGE) to agree with the values for the MFU and YFU.

A calculation of an AFU based on respondents' reported MFU and YFU of blunts was not done if respondents reported that they first used blunts in the same month as their birth dates; in this situation, a unique AFU could not be determined. Similarly, this consistency check was not triggered if respondents had missing data in either of the year or month questions, such as if respondents knew the year when they first used blunts but they did not know the MFU.

If respondents indicated at some point in the consistency check sequence that the value they had originally reported for their AFU was correct (and by extension, that the MFU and YFU was not correct), they had an opportunity to revise the values for their MFU and YFU. These revised reports for MFU and YFU were used in subsequent editing steps. Otherwise, respondents' answers to the original MFU and YFU questions were used in subsequent editing.

Table B.15 in Appendix B discusses edits pertaining to the consistency checks for these incidence variables for blunts. For consistency with how the core incidence data have been edited, the default when a respondent did not resolve an inconsistency between the AFU and the MFU and YFU was to favor the AFU in subsequent editing decisions. Unlike the core incidence data, however, the incidence variables for blunts did not undergo subsequent statistical imputation.

Because these incidence variables were not present in 2004, editing of BLNTREC beyond 2004 did not take into account these incidence data for consistency with how BLNTREC was edited in 2004. Keeping the editing procedures comparable with those in 2004 would permit analysts to examine trends since 2004. However, a flag variable was created that indicated when incidence data suggested more recent use than what the respondent indicated in BLNTREC. The flag was coded as 98 (blank) if no inconsistency existed between BLNTREC and the incidence data.

Table B.16 presents additional edits pertaining to variables in the blunts module. Many of these edits were developed when the module was added to the survey in 2004, particularly with regard to the editing of BLNTREC. For example, if respondents reported using blunts but they also reported never using marijuana, then question BL03 was triggered only when respondents reported using blunts in the past 30 days. Respondents who verified that they had never used marijuana would then have their BLNTEVER answer assigned to 4 (No LOGICALLY ASSIGNED) and BNLTREC assigned to 81 (NEVER USED BLUNTS, Logically assigned). Table B.16 also discusses miscellaneous edits for the incidence variables that have been present since 2005, other than those described in Table B.15 that were related to consistency checks.

No editing was done to make the incidence data for blunts consistent with incidence data for cigars or marijuana from their respective core sections of the interview, or vice versa. Thus, for example, incidence data in the blunts module could indicate that respondents first smoked a cigar with marijuana in it at an earlier age than they reported for when they first used cigars or marijuana. However, variables in the blunts section of the National Survey on Drug Use and Health (NSDUH) codebooks included a standard footnote to indicate that these noncore data may be inconsistent with data from core modules.

7.4.3.2 Medical Marijuana Use

In the 2014 NSDUH, if respondents reported use of marijuana in the past 12 months in the core marijuana module, or they reported smoking a blunt in the past 12 months, or they confirmed smoking a blunt in the past 30 days (i.e., despite reporting in the core marijuana module that they last used marijuana more than 12 months ago), they were asked question MJMM01, regarding whether any of their marijuana use in the past 12 months was recommended by a doctor or other health care professional (i.e., medical marijuana use). If respondents reported any medical marijuana use in the past 12 months (MJMM01 = 1), they were asked question MJMM02 regarding whether *all* of their marijuana use in the past 12 months was recommended by a doctor or other health care professional.

An important aspect of the processing of variables pertaining to medical marijuana use consisted of assigning codes of 91, 93, and 99 (Section 2.4.2) to variables that had been skipped because the medical marijuana questions did not apply to respondents. For example, if respondents never used marijuana and they never smoked part or all of a blunt, they were assigned a code of 91 (NEVER USED MARIJUANA OR BLUNTS) to the variables MEDMJYR (corresponding to MJMM01) and MEDMJALL (corresponding to MJMM02). This edit also was implemented when MJMM01 and MJMM02 had been skipped if respondents reported in the core marijuana module that they never used marijuana but data on lifetime use of blunts was unknown. Similarly, if MJMM01 and MJMM02 had been skipped because respondents last used marijuana more than 12 months ago and there were no reports of use of blunts in the past 12 months, then MEDMJYR and MEDMJALL were assigned a code of 93 (DID NOT USE MARIJUANA OR BLUNTS IN THE PAST 30 DAYS). If respondents were past year users of marijuana or blunts but they answered question MJMM01 as "no," then MEDMJALL was assigned a code of 99 (LEGITIMATE SKIP).

If respondents in 2014 had not been asked MJMM01 and MJMM02, the respective edited variables MEDMJYR and MEDMJALL retained missing values (i.e., blank; see Section 2.2.1) in the following situations:

- The edited variable from the core marijuana module for most recent use of marijuana (MJREC) had a missing value; or
- MJREC indicated use of marijuana at some point in the respondent's lifetime but the exact period of most recent use was unknown (MJREC = 9).

Respondents who previously reported marijuana use in the past 12 months could answer question MJMM01 as "yes" (MEDMJYR=1), even if they were living in a state where a law approving marijuana for medical use was not in effect at the time they were interviewed. Table 7.1 lists the 23 states and the District of Columbia where a law approving marijuana for medical use was enacted by the end of 2014, along with the dates when the laws took effect in these states. Laws approving marijuana for medical use were in effect (i.e., not just enacted) in all but three of these states (Maryland, Minnesota, and New York) in all months of 2014.

In order to better understand the quality of the medical marijuana data, three recoded variables were created that examined (1) reported medical marijuana use, (2) state of residence at the time of interview and up to 2 years before that, (3) dates when medical marijuana laws took effect (where applicable), and (4) the interview dates. The three recoded variables were as follows:

- MEDMJSTA: This variable identifies the medical marijuana law status for all respondents' states of residence at the time of the interview without taking into account their responses to MJMM01 or MJMM02.
- MEDMJCST: This variable examines responses to MJMM01 to determine whether they were consistent with the status of any medical marijuana law being effective in the respondent's current state of residence at the time of the interview.
- MEDMJFLG: This variable is an extension of MEDMJCST, where all the states of residence up to 2 years ago are considered.

The variable MEDMJSTA was created based on the information in Table 7.1 on the status of state medical marijuana laws, along with information on the date of interview and the state where respondents were living on the date they were interviewed. MEDMJSTA was created for *all* respondents in the 2014 NSDUH, regardless of whether they reported any lifetime or past year marijuana use or past year medical marijuana use. MEDMJSTA was coded as follows:

- MEDMJSTA = 1: The respondent was in a state where a law allowing the use of marijuana for medical reasons was in effect on or before the interview date.
- MEDMJSTA = 2: The respondent was in a state that did not have a law in effect for any period of 2014 that allowed the use of marijuana for medical reasons.

- MEDMJSTA = 3: The respondent in 2014 was living in Maryland, Minnesota, or New York but was interviewed before that state's medical marijuana law took effect.[112]

Table 7.1 States Where Marijuana Was Approved for Medical Use as of December 31, 2014, and Dates When the Laws Took Effect

State	Date Enacted (Approved, Passed, or Signed into Law)	Effective Date
California	November 5, 1996	November 6, 1996
Alaska	November 3, 1998	March 4, 1999
Oregon	November 3, 1998	December 3, 1998
Washington	November 3, 1998	November 3, 1998
Maine	November 2, 1999	December 22, 1999
Hawaii	June 14, 2000	December 28, 2000
Colorado	November 7, 2000	June 1, 2001
Nevada	November 7, 2000	October 1, 2001
Vermont	May 26, 2004	July 1, 2004
Montana	November 2, 2004	November 2, 2004
Rhode Island	January 3, 2006	January 3, 2006
New Mexico	March 13, 2007	July 1, 2007
Michigan	November 4, 2008	December 4, 2008
New Jersey	January 18, 2010	July 1, 2011
District of Columbia	May 21, 2010	July 27, 2010
Arizona	November 2, 2010	November 2, 2010
Delaware	May 13, 2011	July 1, 2011
Connecticut	June 1, 2012	October 1, 2012
Massachusetts	November 6, 2012	January 1, 2013
New Hampshire	July 23, 2013	July 23, 2013
Illinois	August 1, 2013	January 1, 2014
Maryland	April 14, 2014	June 1, 2014
Minnesota	May 29, 2014	May 30, 2014
New York	July 5, 2014	July 5, 2014

As discussed, two additional related variables, MEDMJCST and MEDMJFLG, were created that identified reports of medical marijuana use in the past 12 months (MEDMJYR = 1) when respondents were not living in a state where marijuana was approved for medical use at the time they were interviewed (Table 7.1).[113] MEDMJCST compares reports of medical marijuana use in the past 12 months with the states where respondents were living at the time they were

[112] To protect respondent confidentiality, these respondents on the 2014 NSDUH public use file were treated as being equivalent to those for whom MEDMJSTA = 2 (i.e., a medical marijuana law was not in effect in the state at the time the respondents were interviewed).

[113] These variables are not available on the 2014 NSDUH public use file.

interviewed. MEDMJFLG compares reports of medical marijuana use in the past 12 months with the states where respondents had lived within the past 2 years.

Values of 1 through 5 in MEDMJCST have the following meanings:

- MEDMJCST = 1: The respondent reported medical marijuana use in the past 12 months and was living in a state that had a medical marijuana law in effect at the time the respondent was interviewed (corresponds to MEDMJFLG = 0, described below). This includes respondents in 2014 in Maryland, Minnesota, or New York who reported medical marijuana use and were interviewed after their state's law took effect.
- MEDMJCST = 2: The respondent reported medical marijuana use in the past 12 months and was living in a state that did not have a medical marijuana law in effect at any time in 2014.
- MEDMJCST = 3: The respondent reported medical marijuana use in the past 12 months and was in a state (i.e., Maryland, Minnesota, or New York) where a medical marijuana law went into effect in 2014. However, the respondent was interviewed before their state's law took effect. Stated another way, the report of medical marijuana use in the past 12 months occurred before the law was in effect.
- MEDMJCST = 4: The respondent used marijuana in the past 12 months but answered question MJMM01 about use of marijuana that had been recommended by a doctor or other health care professional as "no" (MEDMJYR = 2). This corresponds to MEDMJFLG = 4 (see below).
- MEDMJCST = 5: Whether or not any of the respondent's past year use of marijuana was recommended by a doctor or other health professional was unknown. This corresponds to MEDMJFLG = 98 (see below).

In addition, codes of 91 and 93 were used in MEDMJCST to identify respondents who reported never using marijuana or blunts and not using marijuana or blunts in the past year, respectively.

MEDMJFLG is similar to MEDMJCST, but as noted previously, MEDMJCST compares reports of medical marijuana use in the past 12 months only with the states where respondents were living at the time they were interviewed. For MEDMJFLG, reports of medical marijuana use in the past 12 months were compared with the states where respondents lived in the past 2 years (using LIVSTYAR for the state where respondents were living 1 year ago and LIVST2YR for the state where respondents were living 2 years ago; see Section 5.2.1), in case respondents who reported medical marijuana use in the past 12 months had recently moved from a state that allowed medical marijuana use to a state that did not have a medical marijuana law in effect at the time they were interviewed. Values of 0 through 4 in MEDMJFLG have the following meanings:

- MEDMJFLG = 0: The respondent reported medical marijuana use in the past 12 months and was living in a state that had a medical marijuana law in effect at the time the respondent was interviewed. These reports of use of marijuana in the past 12 months that had been recommended by a doctor or health professional are consistent with the respondents' current state residency (analogous to MEDMJCST = 1).

- MEDMJFLG = 1: The respondent reported medical marijuana use in the past 12 months, but based on LIVSTYAR and LIVST2YR, the respondent in the past 2 years had not lived in a state that had a medical marijuana law in effect. These reports of use of marijuana in the past 12 months that had been recommended by a doctor or health professional are most inconsistent with where respondents had recently been living because these respondents had not recently moved out of a state that allowed medical use of marijuana.
- MEDMJFLG = 2: The respondent reported medical marijuana use in the past 12 months and had lived in a state 2 years ago (from LIVST2YR) that had a medical marijuana law in effect. A year before the interview (from LIVSTYAR) or at the time of the interview, however, the respondent was not living in a state that had a medical marijuana law in effect. These reports of use of marijuana in the past 12 months that had been recommended by a doctor or health professional are not as inconsistent as the reports when MEDMJFLG = 1 because respondents for whom MEDMJFLG = 2 reported living in a state where marijuana was approved for medical use at some point within the past 2 years. However, these reports of medical marijuana use in the past 12 months are less consistent than data when MEDMJFLG = 3 (see below).
- MEDMJFLG = 3: The respondent reported medical marijuana use in the past 12 months and had lived in a state 1 year ago that had a medical marijuana law in effect (from LIVSTYAR). At the time of the interview, however, the respondent was not living in a state that had a medical marijuana law in effect. These reports of use of marijuana in the past 12 months that had been recommended by a doctor or health professional may not necessarily be inconsistent with where respondents currently were living because their 12-month reference period for MEDMJYR could have included a period of time when they were living in a state where marijuana was approved for medical use.
- MEDMJFLG = 4: The respondent used marijuana in the past 12 months but answered question MJMM01 about use of marijuana that had been recommended by a doctor or other health care professional as "no" (MEDMJYR = 2). For these respondents, it does not matter whether or when they lived in a state where marijuana was approved for medical use.

As for MEDMJCST, codes of 91 and 93 in MEDMJFLG were used to identify respondents who reported never using marijuana or blunts or who reported not using marijuana or blunts in the past year. In addition, a code of 98 in MEDMJFLG was used to denote missing data for lifetime marijuana use or for medical marijuana use, or if respondents were lifetime marijuana users but it was unknown whether they used it in the past 12 months.

7.4.4 Substance Dependence and Abuse Module

The substance dependence and abuse module asked about symptoms of dependence or abuse in the past 12 months that were associated with the use of alcohol, marijuana, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives). This section also included items to assess for dependence on cigarettes if respondents reported use of cigarettes in the past 30 days. Details on how respondents were classified as having dependence or abuse for alcohol or illicit drugs (i.e., not including cigarettes) are found in Section B.4.2 in the 2014 NSDUH methodological summary report (Center for Behavioral Health Statistics and Quality, 2015e).

7.4.4.1 Editing of Substance Dependence and Abuse Variables for Nonusers in the Period of Interest

An important aspect of the processing of variables in this module consisted of assigning codes of 91 or 93 (Section 2.4.2) to variables that had been skipped because respondents never used the substance or they used it but not in the period of interest (i.e., more than 30 days ago for cigarettes or more than 12 months ago for other substances).[114] For alcohol and marijuana, respondents who had used these substances in the past 12 months also were skipped out of the corresponding dependence and abuse questions if they were only infrequent users of these two drugs in the past 12 months. If recency-of-use variables for the psychotherapeutic drugs were assigned a code of 81 because they only drugs they ever used nonmedically were OTC drugs (Section 6.2.1.5), then any data in the substance dependence and abuse module for that psychotherapeutic drug were overwritten with a code of 81. For cocaine, heroin, and stimulants, however, respondents' answers in the substance dependence and abuse module were retained if they were routed into that respective section in the substance dependence and abuse module because they reported past year use in the special drugs module (Section 7.4.1.2). Special situations where data in the edited variables for stimulants or sedatives were overwritten with a code of 83 are discussed in Section 7.4.4.2.

For alcohol and marijuana, the final, edited 12-month frequency variables (ALCYRTOT and MJYRTOT, respectively) also were used in assigning codes of 93 or 83 to the substance dependence and abuse variables pertaining to these substances. For example, if the edited variable ALCYRTOT indicated that respondents had used alcohol in the past 12 months but on fewer than 6 days in that period, the edited substance dependence and abuse variables for alcohol were assigned a code of 93 if they had been skipped. If respondents answered one or more dependence or abuse questions for alcohol but the final value for ALCYRTOT indicated that they had used alcohol on fewer than 6 days in the past 12 months, the previous answers in the dependence and abuse questions were overwritten with a code of 83. Similar edits were done for marijuana if MJYRTOT indicated that respondents used marijuana on fewer than 6 days in the past 12 months.

[114] For cocaine, heroin, and stimulants, respondents were not asked the corresponding questions in the substance dependence and abuse module if there was no indication of use in the past 12 months either in the relevant core module (or modules, in the case of cocaine and crack) or in respondents' answers from the special drugs module. However, respondents who did not indicate past year use of cocaine, heroin, or stimulants in the relevant core sections but indicated past year use in the special drugs module *were* routed by the CAI instrument into the relevant drug dependence or abuse questions. For stimulants, this routing logic included situations in which respondents reported use of methamphetamine in the past year in the methamphetamine follow-up items that were added to the special drugs module in 2005 and the follow-up questions that were added in 2006 (see Section 6.2.6).

7.4.4.2 Editing of Substance Dependence and Abuse Variables for Alcohol through Sedatives

A second important aspect of the processing of the substance dependence and abuse variables for alcohol through sedatives involved assignment of a legitimate skip code of 99 (Section 2.4.2) when respondents qualified for being asked dependence or abuse questions about a given substance but they legitimately skipped out of one more questions about that substance. For example, the symptom of tolerance to the effects of alcohol was measured through two related questions, DRALC06 ("During the past 12 months, did you need to drink more alcohol than you used to in order to get the effect you wanted?") and DRALC07 ("During the past 12 months, did you notice that drinking the same amount of alcohol had less effect on you than it used to?"). An affirmative answer to either question would indicate tolerance. Thus, if respondents had already answered DRALC06 as "yes," there was no need to ask DRALC07. If the edited variable corresponding to question DRALC06 (ALCNDMOR) was coded as 1 (i.e., "yes"), the edited variable corresponding to question DRALC07 (ALCLSEFX) was assigned a legitimate skip code.

Aside from assignment of codes of 91, 93, or 99, minimal additional editing was done to the substance dependence and abuse variables. However, one notable change was that since 2005, the CAI logic for questions related to stimulant dependence or abuse have taken into account the new follow-up questions SD17A or SD18A that were added to the special drugs module for respondents who had not previously reported methamphetamine use in the core stimulants module (Section 6.2.6). Thus, if respondents in 2005 indicated past year use of methamphetamine in these noncore questions, they were asked questions in the substance dependence and abuse module about stimulant dependence or abuse. In addition, since 2006, this logic has skipped respondents out of the dependence and abuse questions for stimulants if they reported methamphetamine use in the past year in the special drugs module but then they reported in question SD17A1 that their earlier report of lifetime nonuse of methamphetamine from the core stimulants module was actually the correct response.

Since 2008, the instrument routing logic also has taken into account reports from the special drugs module of past year nonmedical use of Adderall® for asking the stimulant dependence and abuse questions and past year nonmedical use of Ambien® for asking the dependence and abuse questions for sedatives. In addition, respondents who reported past year use of ketamine, DMT, AMT, "Foxy," or *Salvia divinorum* have been asked the hallucinogen dependence and abuse questions.

In editing the stimulant dependence and abuse variables, stimulant dependence or abuse data were retained for these respondents who indicated past year methamphetamine use or past year nonmedical use of Adderall® in the noncore special drugs items. Consequently, if the stimulant dependence and abuse variables had been skipped because lifetime nonuse of methamphetamine had been logically inferred for the noncore methamphetamine recency MTHAREC and ADDEREC also indicated lifetime nonuse, the edited stimulant dependence and abuse variables were assigned a code of 91 (NEVER USED STIMULANTS). If MTHAREC or ADDEREC indicated past year use, data were retained in the stimulant dependence and abuse variables, even if the other stimulant variables that existed prior to 2005 (including the methamphetamine and other stimulant needle use variables that were in the special drugs module prior to 2005) would have indicated that the respondent was not a past year stimulant user. Similar principles since 2008 applied to the editing of the sedative dependence and abuse variables based on AMBIREC and the hallucinogen dependence and abuse variables based on KETAREC, TRYPREC, and SALVREC (Section 7.4.1.4). In the case of hallucinogens, for example, data were retained in the dependence and abuse variables if there was any indication of past year use in KETAREC, TRYPREC, or SALVREC, even if the core variable HALLREC did not indicate past year use.

As noted previously, substance dependence and abuse data for psychotherapeutic drugs were overwritten if the only drugs that respondents used nonmedically were OTC drugs. For stimulants and sedatives, the data were overwritten with a code of 83 rather than a code of 81 if ADDEREC or AMBIREC, respectively, indicated nonmedical use more than 12 months ago.

Therefore, in addition to the flag variable STMDAFLG that has been present since 2005 to indicate differences for analysts in the editing of the stimulant variables in the substance dependence and abuse module, the flag variables HALDAFLG and SEDDAFLG were created to indicate differences in how the hallucinogen and sedatives variables in the substance dependence and abuse module have been edited since 2008. In the case of STMDAFLG, levels 1 through 6 have been present since 2005 to document the effects of the noncore methamphetamine data on the editing of the stimulant dependence and abuse variables, and levels 11 through 16 have been present since 2008 to document the effects of ADDEREC on the editing of these variables. Documentation for these levels in STMDAFLG in 2014 was as follows, where "Now" refers to this survey year (i.e., 2014):

- 1 = Never used based on 04/Now: Not past yr user
- 2 = Not past yr user based on 04/Now: Blank
- 3 = Never used based on 04/Now: Blank
- 4 = Blank based on 04/Now: Past yr user
- 5 = Not past yr user based on 04/Now: Past yr user
- 6 = Never used based on 04/Now: Past yr user
- 11 = Never used based on 07/Now: Not past yr user
- 12 = Not past yr user based on 07/Now: Blank
- 13 = Never used based on 07/Now: Blank
- 14 = Blank based on 07/Now: Past yr user
- 15 = Not past yr user based on 07/Now: Past yr user
- 16 = Never used based on 07/Now: Past yr user
- 98 = BLANK (NO DIFFERENCE BETWEEN 04 AND NOW)

Thus, a code of 98 in STMDAFLG meant that the presence of the noncore methamphetamine and Adderall® items in the special drugs module had no effect on how the stimulant dependence and abuse variables have been edited since 2004. For example, if respondents indicated past year nonmedical use of stimulants in the core stimulants module (i.e., but they did not indicate methamphetamine use) or if they indicated that they used stimulants with a needle in the past year (from MTNLDREC or OSTNLREC in the special drugs module), they would have been asked questions about stimulant dependence or abuse regardless of how they answered the methamphetamine and Adderall® questions in the special drugs module. Similarly, if respondents were classified as lifetime nonusers of stimulants based on their answers in the core stimulants module, and the special drugs data (including the additional methamphetamine Adderall® items) continued to indicate that these respondents never used methamphetamine or other prescription-type stimulants (or Adderall®), the stimulant dependence and abuse items were coded as 91, as would have been the case in 2004 and earlier years.

For codes 1 through 6 in STMDAFLG since 2008, ADDEREC did not indicate past year nonmedical use of Adderall®. A code of 1 in STMDAFLG meant that respondents who skipped the stimulant dependence and abuse questions also would have skipped out of these items in 2004 and earlier years. The only difference was that in 2004 (and earlier), the edited stimulant dependence and abuse variables would have been coded as 91 (NEVER USED STIMULANTS). Based on data from the additional methamphetamine variables in the special drugs module, these variables currently were coded as 93 (DID NOT USE STIMULANTS IN THE PAST 12 MONTHS).

A code of 2 meant that respondents would have been coded as lifetime but not past year users of stimulants based solely on the variables that existed in 2004. The stimulant dependence and abuse variables currently were coded as 98 (blank) because these respondents reported methamphetamine use in the additional special drugs items, but it was not clear whether they used methamphetamine in the past year.

A code of 3 meant that respondents would have been coded as nonusers based solely on the variables that existed in 2004. The stimulant dependence and abuse variables were coded as 98 (blank) for the same reason as that given for situations in which STMDAFLG was coded as 2.

Beginning with a code of 4, nonmissing data existed in the current stimulant dependence and abuse variables that would not have existed in 2004.

- A code of 4 meant that the variables that existed in 2004 indicated that respondents were at least lifetime nonmedical users of stimulants, but it was not clear whether they had used in the past year. However, these respondents indicated past year methamphetamine use in the additional special drugs items.
- A code of 5 meant that respondents would have been classified as lifetime but not past year nonmedical users of stimulants based on the variables that existed in 2004, but they indicated past year methamphetamine use in the additional special drugs items.
- A code of 6 indicated the greatest potential difference between 2004 and currently. Based on the variables that existed in 2004, these respondents would have been classified as never having used stimulants nonmedically, but they indicated past year methamphetamine use in the additional special drugs items.

Levels 11 through 16 in STMDAFLG were analogous to levels 1 through 6. For levels 11 through 16, however, the noncore methamphetamine data would not have affected how the stimulant dependence and abuse variables were edited but ADDEREC did. For example, a code of 11 in STMDAFLG meant that respondents who skipped the stimulant dependence and abuse questions in 2008 onward also would have skipped out of these items in 2007, even with the noncore methamphetamine data. The difference was that in 2007, the edited stimulant dependence and abuse variables would have been coded as 91 (NEVER USED STIMULANTS) because the respondent was classified as a lifetime nonuser of stimulants based on the core stimulant data and noncore methamphetamine data. However, ADDEREC indicated that these respondents used Adderall® nonmedically more than 12 months ago but in their lifetime. Consequently, the stimulant dependence and abuse variables in this situation were assigned a code of 93 (DID NOT USE STIMULANTS IN THE PAST 12 MONTHS); this was analogous to the situation where STMDAFLG = 1.

Similarly, since 2008, codes 14 through 16 indicated that nonmissing data existed in the stimulant dependence and abuse variables that would not have existed in 2007. In particular, a code of 16 indicated the greatest potential difference between 2007 and currently. Based on the variables that existed in 2007, these respondents would have been classified as never having used stimulants nonmedically (including never having used methamphetamine based on the noncore special drugs data). However, they indicated past year nonmedical use of Adderall®.

In HALDAFLG and SEDDAFLG, levels 1 through 6 corresponded to levels 1 through 6 in STMDAFLG. In HALDAFLG and SEDDAFLG, however, data were compared relative to the results of how the relevant dependence and abuse variables would have been edited in 2007 (rather than 2004). For example, a code of 1 in HALDAFLG meant that respondents who skipped the hallucinogen dependence and abuse questions in 2008 onward also would have skipped out of these items in 2007. The difference was that in 2007, the edited hallucinogen dependence and abuse variables would have been coded as 91 (NEVER USED HALLUCINOGENS) because the core hallucinogen data indicated that the respondents never used hallucinogens. However, there was some indication of use more than 12 months ago in KETAREC, TRYPREC, or SALVREC but no indication of use in the past 12 months. Consequently, the current hallucinogen dependence and abuse variables were assigned a code of 93 (DID NOT USE HALLUCINOGENS IN THE PAST 12 MONTHS) in this situation.

7.4.4.3 Editing of Cigarette Dependence Variables

The questionnaire items that were used to determine cigarette dependence (also known as "nicotine dependence") in the 2014 NSDUH have been included in the survey since 2001. The method for determining dependence involved calculating a continuous score from the Nicotine Dependence Syndrome Scale (NDSS) (Shiffman, Hickcox, Gnys, Paty, & Kassel, 1995; Shiffman, Waters, & Hickcox, 2003). The score was calculated from the 17 questions shown in Table 7.2 that were asked of respondents who used cigarettes in the past 30 days. For each of these items, respondents who had used cigarettes in the past 30 days were asked to choose an answer from the following scale: 1 = Not at all true; 2 = Somewhat true; 3 = Moderately true; 4 = Very true; or 5 = Extremely true. For details on how the estimate for cigarette dependence was calculated based on the NDSS, see Section B.4.2 in Appendix B of the 2009 summary of national findings (Office of Applied Studies, 2010a; 2010b).

As for other variables in the substance dependence and abuse module, editing of the cigarette dependence variables also involved assignment of a legitimate skip code of 99 according to the skip logic within this section of the module. For example, question DRCGE06A (not shown in Table 7.2) asked if respondents who reported cigarette use in the past 30 days had any friends who did *not* smoke cigarettes. If DRCGE06A was answered as "no," question DRCGE06B was skipped (i.e., choosing at times not to be around friends who do not smoke because they do not like it when the respondent smokes). The edited variable CIGFNLKE corresponding to DRCGE06B was assigned a code of 99 if DRCGE06A was answered as "no."[115]

Table 7.2 Mapping of Cigarette Dependence Questions to Edited Variables

Question	Edited Variable	Question Text
DRCGE01	CIGIRTBL	After not smoking for a while, you need to smoke in order to feel less restless and irritable.
DRCGE02	CIGCRAVE	When you don't smoke for a few hours, you start to crave cigarettes.
DRCGE03	CIGCRAGP	You sometimes have strong cravings for a cigarette where it feels like you're in the grip of a force you can't control.
DRCGE04	CIGINCTL	You feel a sense of control over your smoking – that is, you can "take it or leave it" at any time.
DRCGE05	CIGAVOID	You tend to avoid places that don't allow smoking, even if you would otherwise enjoy them.
DRCGE07	CIGPLANE	Even if you're traveling a long distance, you'd rather not travel by airplane because you wouldn't be allowed to smoke.
DRCGE08	CIGRNOUT	You sometimes worry that you will run out of cigarettes.
DRCGE09	CIGREGDY	You smoke cigarettes fairly regularly throughout the day.
DRCGE10	CIGREGWK	You smoke about the same amount on weekends as on weekdays.
DRCGE11	CIGREGNM	You smoke just about the same number of cigarettes from day to day.
DRCGE12	CIGNMCHG	It's hard to say how many cigarettes you smoke per day because the number often changes.
DRCGE13	CIGSVLHR	It's normal for you to smoke several cigarettes in an hour, then not have another one until hours later.
DRCGE14	CIGINFLU	The number of cigarettes you smoke per day is often influenced by other things – how you're feeling or what you're doing, for example.
DRCGE15	CIGNOINF	Your smoking is not affected much by other things. For example, you smoke about the same amount whether you're relaxing or working, happy or sad, alone or with others.
DRCGE16	CIGINCRS	Since you started smoking, the amount you smoke has increased.
DRCGE17	CIGSATIS	Compared to when you first started smoking, you need to smoke a lot more now in order to be satisfied.
DRCGE18	CIGLOTMR	Compared to when you first started smoking, you can smoke much, much more now before you start to feel anything.

[115] Consistent with the principles in Section 2.4.2, CIGFNLKE retained missing values if DRCGE06B was skipped because respondents answered DRCGE06a as "don't know" or "refused."

No additional editing was done to the cigarette dependence variables if respondents reported cigarette use in the past 30 days or they used cigarettes in the past 30 days but skipped out of questions that did not apply. In particular, no editing was done when respondents entered the same response for all items (e.g., keying a "1" to every item). If respondents entered the same response to all cigarette dependence items, however, that would strongly suggest that they were not paying careful attention to the questions. For example, if a respondent chose "1" (i.e., not at all true) in question DRCGE03 (Table 7.2), the respondent would be expected to choose a response other than "1" to indicate some degree of truth to the statement in question DRCGE04 about having some sense of control over smoking. Nevertheless, when respondents entered the same response to all cigarette dependence items, these data were retained in order to allow analysts to decide how they wanted to handle these cases.

7.4.4.4 Imputation of Cigarette Dependence Variables

Unlike the majority of variables that are imputed for NSDUH, imputation of the cigarette dependence variables was not performed using the PMN methodology. For respondents who used cigarettes in the past 30 days and provided complete data for all 17 of the dependence questions that were used to calculate the NDSS scale value, imputation-revised cigarette dependence variables were simply assigned the values from the corresponding edited variables.

For respondents who used cigarettes in the past 30 days and gave a valid response to 16 of the 17 NDSS items, the predicted mean for the one missing item was obtained using the coefficients corresponding to the other 16 nonmissing covariates from the appropriate weighted least squares regression. For example, if CIGIRTBL was the variable whose missing value was to be imputed, CIGIRTBL would be specified as the dependent variable in the model, and the remaining 16 NDSS variables served as the covariates: CIGCRAVE, CIGCRAGP, CIGINCTL, CIGAVOID, CIGPLANE, CIGRNOUT, CIGREGDY, CIGREGWK, CIGREGNM, CIGNMCHG, CIGSVLHR, CIGINFLU, CIGNOINF, CIGINCRS, CIGSATIS, and CIGLOTMR. The imputation-revised variable was then set to the predicted mean.[116]

Respondents who used cigarettes in the past 30 days but answered 15 or fewer of the cigarette dependence questions were left out of the modeling process entirely. The missing values for these respondents remained missing in the imputation-revised variables that were derived from the edited NDSS variables.

Across all respondents (regardless of how many of the NDSS questions they answered), no response propensity adjustments were performed for the item respondent weights used in the regression-based imputation models. The analysis weights described in Section 3.3.1 were applied instead.

The NDSS mean value was calculated from imputation-revised versions of the 17 nicotine dependence questionnaire variables. The NDSS mean value was set to missing for respondents who used cigarettes in the past 30 days but answered 15 or fewer of the cigarette dependence questions.

---

[116] Because the response variable and covariates were treated as continuous in the models, it is possible for a predicted mean, and therefore an imputation-revised value, to exceed five or be less than one.

7.4.4.5 Summary Information for Cigarette Dependence Variables

In the 2014 NSDUH, 166 cases had valid values for 16 of the 17 questions. These cases were imputed. In 2014, 107 cases had fewer than 16 valid values for the NDSS scale. For these cases, no imputations were performed. Note that it was possible that the respondent was eligible to answer the questions about cigarette dependence because he or she was imputed to be a past month cigarette user. This situation occurred seven times in the 2014 NSDUH.

Table 7.3 summarizes the eligibility of respondents to answer the cigarette dependence questions and reasons why respondents were classified as eligible or not eligible to be imputed. Furthermore, this table provides details about the amount of cigarette dependence data that were missing for eligible respondents. It also provides information on whether the respondent was imputed to be a past month cigarette user. Consequently, the respondent would be eligible to have cigarette dependence data but would have missing data for all the cigarette dependence variables.

Of the 273 eligible respondents in 2014 who did not answer all 17 NDSS questions, the majority (166 cases, 60.8 percent of the eligible cases with incomplete data) were missing a response for only one of the items (Table 7.3). Any respondent with more than one of the 17 items missing (107 cases, 39.2 percent of the eligible cases with incomplete data) did not have his or her missing responses replaced with imputed values, and no cigarette dependence score was calculated for those respondents. For the respondents missing only one response, imputation was used to fill in the values for the missing variable, using information from the other 16 nonmissing variables, through weighted least squares regression models. This resulted in 17 regression-based imputation models, where the response variable for each model was the edited variable that corresponded to each item in the NDSS, and the covariates in each model were the remaining NDSS variables.

Table 7.3 Summary of Response Patterns for 2014 NDSS Variables

Number of Valid NDSS Variables	Past Month Smoker	Past Month Smoker Status Imputed	Eligible to Answer NDSS Questions	NDSS Variables Imputed	Frequency	Percentage of Total
N/A	No	No	No	N/A[1]	53,780	79.20
N/A	No	Yes	No	N/A[1]	16	0.02
Subtotal					53,796	79.23
0	Yes	No	Yes	No	14	0.02
0	Yes	Yes	Yes	No	7	0.01
1-15	Yes	No	Yes	No	86	0.13
Subtotal					107	0.16
16	Yes	No	Yes	Yes	166	0.24
17	Yes	No	Yes	N/A[2]	13,832	20.37
Total					67,901	100.00

N/A = not applicable; NDSS = Nicotine Dependence Syndrome Scale.
[1] The respondent was not eligible to answer the NDSS questions.
[2] None of the NDSS variables were missing for this scenario.

7.4.5 Special Topics Module

The special topics module asked about arrests in the respondents' lifetime and in the past 12 months, including arrests for specific offenses in the past 12 months (not counting minor traffic violations). This section also included questions about respondents being on probation or parole in the past 12 months, operating vehicles under the influence of alcohol or illegal drugs in the past 12 months, and respondents' knowledge about their states' marijuana laws.

If respondents reported that they had never been arrested in their lifetime and they did not report being on probation or parole in the past 12 months (see below), the edited variables pertaining to arrests in the past 12 months were assigned legitimate skip codes. Other standard edits described in Section 2.4.3 pertaining to situations where respondents answered "don't know" or "refused" to the lifetime arrest question were applied to the past year arrest variables that had been skipped.

Similarly, if respondents reported being arrested in their lifetime but reported being arrested 0 times in the past 12 months, the questions pertaining to arrests for specific offenses in the past 12 months were assigned legitimate skip codes. Respondents who did not know how many times they were arrested in the past 12 months or who refused to answer this question were asked whether they were arrested for specific offenses in the past 12 months. This was consistent with the logic in 1999 and since 2001 but differed from the logic in 2000, when respondents who answered "don't know" or "refused" to the question about the number of specific arrests in the past 12 months were skipped out of questions about arrests for specific offenses in the past 12 months.

Respondents also were skipped out of questions pertaining to driving under the influence of alcohol or illegal drugs if they reported in the core modules that they never used alcohol, marijuana, cocaine, heroin, hallucinogens, inhalants, or prescription psychotherapeutics for nonmedical reasons; this included situations in which respondents were classified as having never used stimulants nonmedically based on their answers in the core stimulants module, and the special drugs data (including the added special drugs items on use of methamphetamine) continued to indicate that the respondents had never used stimulants. When respondents were lifetime nonusers of alcohol and the other drugs mentioned above, all skipped variables pertaining to driving under the influence of alcohol or illegal drugs were assigned a code of 91 to indicate that the respondents were lifetime nonusers of all of these substances (Section 2.4.2). If respondents were skipped out of one or more of the substance use and driving items because their most recent use of a drug was more than 12 months ago, the edited variables were assigned legitimate skip codes.

In addition, respondents were asked about driving under the influence of illegal drugs (alone or in combination with alcohol) if they reported use of methamphetamine in the past year in the methamphetamine follow-up items from the special drugs module. No editing was done to the variables about driving under the influence of illegal drugs in this situation. However, the additional methamphetamine items from the special drugs module would have affected routing to the items in special topics about driving under the influence of illegal drugs only if these methamphetamine items were the sole indication of use of drugs other than alcohol in the past year. Stated another way, if respondents had reported past year use of marijuana, cocaine, heroin, hallucinogens, or inhalants, or if they had reported past year nonmedical use of pain relievers, tranquilizers, stimulants other than methamphetamine, or sedatives, they would have been asked the questions about driving under the influence of illegal drugs regardless of how they answered the questions from the special drugs module about methamphetamine. A flag (DRVFLAG) was created for use in subsequent data analysis to identify any respondents whose only use of drugs other than alcohol came from the methamphetamine items in the special drugs module; updated logic for this flag has been in place since 2006 to take into account the consistency checks between the core and noncore responses that were added to the special drugs module in 2006 (Section 6.2.6.

The variable pertaining to a state's maximum legal penalty for first offense possession of an ounce or less of marijuana was renamed from MXMJPENL to MXMJPNLT. The variable name was changed to reflect the addition of a response category for "No penalty" and to alert analysts that data in MXMJPNLT from 2006 onward may not be comparable with data in MXMJPENL prior to 2006.

Minimal editing was done to the variables MXMJPNLT, MXMJSURE, and MEDMJUSE, corresponding to questions SP07, SP08, and SP09, respectively; these variables pertained to knowledge about state marijuana laws and penalties. If respondents indicated in MXMJPNLT that they did not know the maximum legal penalty in their states for possession of an ounce or less of marijuana for personal use, or if they refused to answer this question, the edited variable MXMJSURE (regarding respondents' degree of certainty about their answer to question SP07) was assigned a legitimate skip code. In addition, if interviewers had entered incorrect information in the FIPE4 question regarding the state where the respondent's sampled dwelling unit was located, the variables MXMJPNLT, MXMJSURE, and MEDMJUSE were assigned bad data codes.[117] This latter edit was done because the state that respondents were asked about in these questions was governed by the state that interviewers entered in FIPE4. Hence, if interviewers entered incorrect state information in FIPE4, the answers that respondents provided in questions SP07 through SP09 were deemed to be questionable. For example, if a respondent lived in California (FIPE4 = 5) but the interviewer entered that the respondent's sampled dwelling unit was in Colorado (FIPE4 = 6), the respondent would be asked for information on marijuana laws in Colorado.

Table B.17 in Appendix B discusses additional edits that were specific to the special topics module. For example, respondents could report that they had never been arrested in their lifetime but could report that they were on probation, parole, or supervised release in the past 12 months. Because someone could not be on probation or parole without first having been arrested for a crime, these respondents were logically inferred to have been arrested in their lifetime. When this situation occurred, the skipped variables pertaining to arrests in the past 12 months retained a value of blank.

Since 2002, respondents who reported in question SP02 that they were arrested and booked for breaking the law at least once in the past 12 months and gave negative answers to every question about specific arrests (including arrests for "some other offense") have been routed to a follow-up question, SP03R, that asked respondents to verify their previous answer

[117] Creation of the edited variable STATELOC from FIPE4 is discussed in detail in Chapter 4.

from SP02. If respondents did not indicate that their previous answer from question SP02 was correct, they were routed to question SP03S, where they were allowed to update their answer regarding the number of times they were arrested or booked in the past 12 months. If respondents indicated in SP03S that they had been arrested or booked 0 times in the past 12 months, the edited variable NOBOOKYR was assigned a value of 0. Further, when SP03S indicated that respondents had been arrested and booked 0 times in the past 12 months, it was logically inferred that all items pertaining to specific arrests in the past 12 months should have been skipped. Therefore, all of the variables associated with specific arrests in the past 12 months were assigned a code of 89 (Section 2.4.2).

If respondents indicated in SP03R that their previous answer from SP02 was correct, or if they reported being arrested and booked for at least one offense in SP03S, they were asked to specify at least one offense for which they were arrested and booked in the past 12 months (questions SP03RSP or SP03SSP). If respondents specified a legitimate offense in SP03RSP or SP03SSP after giving negative answers to every question about specific arrests, the "OTHER, Specify" variable BKOTHOFF was updated to incorporate the response from SP03RSP or SP03SSP. A code of 3 (Yes LOGICALLY ASSIGNED) also was assigned to the "some other offense" variable (BKOTH) to indicate that this edit had taken place (Section 2.4.5).

Any information that respondents specified in SP03RSP or SP03SSP also was used to edit variables pertaining to offenses that respondents previously had been asked about. For example, if respondents had answered question SP03A as "no" (i.e., had not been arrested and booked for motor vehicle theft in the past 12 months) but then specified in SP03RSP or SP03SSP that motor vehicle theft was one of the offenses for which they had been arrested, the edited variable BKMVTHFT was assigned a code of 3.

If respondents reiterated in SP03R or SP03S that they had been arrested and booked for at least one offense in the past 12 months but still did not report a legitimate offense in questions SP03RSP or SP03SSP (including situations in which they answered SP03RSP or SP03SSP as "don't know" or "refused"), then a code of 5 was assigned to BKOTH. This code of 5 had the following meaning: 5 = Offense unknown. Stated another way, the response from SP02 or SP03S was retained in NOBOOKYR to indicate that the respondents were arrested in the past 12 months, but it was not possible to determine the specific crime for which they were arrested.

7.4.6 Marijuana Purchases Module

The marijuana purchases module focused on the acquisition of marijuana. Administration of questions in this module was limited to respondents who had reported previously in the core marijuana module that they used marijuana in the past 12 months (i.e., regardless of whether they reported in the blunts module that they smoked marijuana in a blunt in the past 12 months). These respondents were asked how they obtained the last marijuana they used, including buying it, trading something for it, getting it for free (or sharing someone else's), or growing it. The module also included questions about the contexts in which respondents engaged in transactions involving marijuana, including where respondents were when they bought, traded for, or got marijuana for free; from whom the respondents got the marijuana (if they did not grow it themselves); and whether they sold or gave away any of this marijuana (including those respondents who grew their own).

If respondents did not report buying the last marijuana they used, they were asked a follow-up question to identify those who had bought any marijuana in the past 12 months. Respondents who reported purchasing the last marijuana they used or who reported purchasing it at any time during the past 12 months were asked more detailed questions about their purchases of marijuana, and they were skipped out of questions pertaining to trading for marijuana, getting it for free, or growing it.

Similarly, respondents who reported that they traded something for the last marijuana they used and who had not bought marijuana at any time during the past 12 months were asked more detailed questions about trading for marijuana. If respondents did not report trading for the last marijuana they used, they were asked a follow-up question to identify those who had traded something for marijuana in the past 12 months. Respondents who had not been routed into questions about buying marijuana and who were asked more detailed questions about trading for marijuana were skipped out of questions pertaining to getting marijuana for free or growing it.

Respondents who were routed into more detailed questions about purchases of marijuana were asked whether they last bought marijuana in "joints" or in loose form, the quantity they purchased the last time they bought marijuana, and the price they paid. Similar questions were asked of respondents who were routed into questions about trading for marijuana, except that these respondents were asked to estimate the worth of the marijuana they obtained through trading.

Edits in this module principally involved assigning appropriate legitimate skip codes based on the logic for determining whether respondents should be administered the module, or the routing logic within the module, if respondents had used marijuana in the past 12 months (Section 2.4.2). If respondents reported in the marijuana module in the core that they had never used marijuana, the edited variables in the marijuana purchases module were assigned a code of 91 (or 991, etc.) to indicate that respondents had skipped out of the module because they were lifetime nonusers of marijuana. Similarly, if respondents' edited marijuana recency MJREC indicated that they last used marijuana more than 12 months ago, the edited variables in the marijuana purchases module were assigned a code of 93 (or 993, etc.) to indicate that respondents had skipped out of the module because they had used marijuana, but not in the past year. If respondents had been skipped out of the marijuana purchases module but their edited marijuana recency had a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED), the skipped marijuana purchases variables retained codes of blank because at least some of these respondents potentially used marijuana in the past 12 months and would have been eligible to be asked questions in the marijuana purchases module.

If respondents previously reported that they had used marijuana in the past 12 months, a key aspect of the editing of variables in the marijuana purchases module involved assignment of legitimate skip codes (99, 999, etc.) according to how respondents were routed through the module. As discussed previously, for example, respondents who gave some report of having bought marijuana were skipped out of questions about trading for marijuana, growing it, or getting it for free. Similarly, respondents who gave some indication of having traded for marijuana (without having indicated buying it) were skipped out of questions related to growing it or getting it for free. If respondents reported buying or trading for marijuana and bought or traded for it in joints, they were skipped out of questions pertaining to buying or trading for marijuana in loose form, and vice versa. If respondents bought or traded for marijuana in loose form, respondents also were routed into or skipped out of questions about the quantities they obtained based on whether they reported purchasing or trading for grams, ounces, or pounds of marijuana. In addition, respondents who reported that they grew the last marijuana they used (without having indicated that they bought or traded for marijuana) were skipped out of questions related to getting marijuana for free, and respondents who reported that they got their last marijuana for free were skipped out of questions related to growing it.

The remaining processing of the variables in the marijuana purchases module involved creating summary variables for the price that respondents paid for the last marijuana they bought or the estimated value of the marijuana they got through a trade. Respondents were first asked to report broad categories of prices. For some of these broader categories (e.g., if respondents reported paying $21.00 to $50.99), respondents were asked to report more detailed price categories (e.g., $21.00 to $30.99; $31.00 to $40.99; $41.00 to $50.99) in order to define more narrowly how much they paid for the marijuana (or how much they estimated the marijuana to be worth). The routing to these more detailed questions was contingent on the broader price category that respondents reported, such that responses to the more detailed price questions were mutually exclusive. Therefore, "composite" summary cost variables were created based on this routing logic.

For example, if respondents reported buying marijuana in loose form the last time, the broad price category variable was called MMLSPCTB (corresponding to question MJE20), where "LS" stood for marijuana in loose form, and "PCTB" stood for "broad price category." Similarly, the detailed price category variable for buying marijuana in loose form was called MMLSPCAT and was derived from responses in questions MJE20 through MJE25. If, for example, a respondent reported in question MJE20 that he or she paid "$21.00 to $50.99" for the last marijuana purchase (level 4 in question MJE20), MMLSPCAT was coded as 41 if the respondent reported paying $21.00 to $30.99 (level 1 in question MJE21); 42 if the respondent reported paying $31.00 to $40.99 (level 2 in question MJE21); and 43 if the respondent reported paying $41.00 to $50.99 (level 3 in question MJE21).

If respondents reported a broad price category for the marijuana they bought or traded for but they did not know (or refused to report) a more detailed price, the response from the "broad" price category variable (e.g., MMLSPCTB) was used to create a value for the corresponding detailed price category variable (e.g., MMLSPCAT). For example if respondents reported paying $21.00 to $50.99 in question MJE20 but they did not recall more detailed information, the variable MMLSPCAT was assigned a code of 40. This code indicated that it could at least be determined that the respondent paid $21.00 to $50.99, but that more detailed information was not available.

In addition, respondents who reported that they bought marijuana, traded something for marijuana, or got it for free in the past 12 months were asked where they were the last time they got marijuana in these different ways (edited variables MMBPLACE, MMTPLACE, and MMFLACE for where respondents were when they last bought, traded for, or got marijuana for free, respectively). Available response categories in these items were (1) inside a public building, such as a store, restaurant, sports arena, bar, or club; (2) inside a school building; (3) outside on school property; (4) inside a home, apartment, or dorm; (5) outside in a public area, such as a parking lot, street, or park; or (6) some other place.

"OTHER, Specify" variables have been included in the marijuana purchases module since 2005 to capture further information from respondents who reported that they were in some other place when they last bought marijuana, last traded something for marijuana, or last got marijuana for free. If the other place that respondents specified corresponded to one of the response categories mentioned above, the relevant category in the "place" variable was assigned, plus a value of 10 (e.g., in MMBPLACE for buying marijuana, corresponding to question MJE27). Suppose, for example, that respondents reported that they were in some other place when they last bought marijuana (i.e., category 6 chosen in question MJE27), but they specified something that corresponded to them being outside in a public area. That would have been equivalent to respondents having selected category 5 from MJE27. Therefore, MMBPLACE was assigned a code of 15, where 15 = Outside in a public area LOGICALLY ASSIGNED.

7.4.7 Prior Substance Use Module

The prior substance use module covered a wide variety of topics:

- retrospective use of marijuana, cigarettes, alcohol, and cocaine in the year prior to the past 12 months (also referred to below as the past 12 to 24 months);
- the age, year, and month when respondents last used substances covered in the core section of the interview (i.e., age at last use [ALU], month of last use [MLU], and year of last use [YLU]) or in the noncore special drugs module for methamphetamine if they were lifetime but not past month users of these substances;
- sources of psychotherapeutic drugs that respondents used nonmedically in the past month or past year, including how friends or relatives obtained psychotherapeutic drugs that respondents subsequently got from these other individuals; and
- the sequence of initiation of use of cigarettes, alcohol, and marijuana.

One of the important aspects of the processing of variables in this module consisted of assigning codes of 91 and 99 (Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, if respondents never used marijuana or indicated use in the past 30 days, they were skipped out of the questions asking for their age, year, and month when they last used marijuana. Respondents were skipped out of the questions pertaining to the age, year, and month when they last smoked cigarettes daily if (1) they smoked every day in the past 30 days or (2) they never had a period in their lives when they smoked cigarettes every day for at least 30 days. When a given core recency variable (e.g., MJREC for marijuana) had a refusal code and the corresponding variables in the prior substance use module had been skipped, that refusal from the core recency was propagated to the edited prior substance use variables (Section 2.4.3). Documentation of edits for specific subsections of the prior substance use module is presented in the remainder of this section.

7.4.7.1 Retrospective Use

Since 2003, this module has included a question about retrospective use of marijuana in the year prior to the past 12 months (i.e., edited variable MRJYRBFR, corresponding to question LU01). In addition, retrospective questions on use of cigarettes, alcohol, and cocaine in the year before the past 12 months (edited variables CIGYRBFR, ALCYRBFR, and COCYRBFR, respectively) have been included since 2005. These questions were analogous to MRJYRBFR.

If respondents never used the drug of interest (e.g., marijuana), the retrospective variable (e.g., MRJYRBFR) was assigned a code of 91. In addition, the cigarette recency variable CIGREC indicated if respondents last smoked cigarettes more than 3 years ago (CIGREC = 4). Logically, if respondents last smoked cigarettes more than 3 years ago, they would not have smoked a cigarette in the past 12 to 24 months. Therefore, when CIGREC indicated that respondents last smoked cigarettes more than 3 years ago and question LU37 (corresponding to CIGYRBFR) had been skipped, CIGYRBFR was assigned a code of 99 (LEGITIMATE SKIP).

The retrospective variables MRJYRBFR, ALCYRBR, CIGYRBFR, and COCYRBFR were not edited for consistency with other data for these drugs in the prior substance use module. For example, if MRJYRBFR indicated that respondents used marijuana in the past 12 to 24 months but respondents indicated that they last used marijuana at an age in edited variable MRJAGLST that would suggest that they last used marijuana more than 2 years ago, no editing was done to either MRJYRBFR or MRJAGLST. However, a codebook "NOTE" was added to alert analysts that these types of inconsistencies could exist between the related variables for these drugs.

7.4.7.2 Last Use of Drugs

The prior substance use module included questions about the last use of all drugs that were covered in the core section of the interview. If respondents were lifetime users of specific substances in the core section but had not used these substances in the past 30 days, they were asked in this module for the age, year, and month when they last used these drugs or tobacco (i.e., ALU, MLU, and YLU). If respondents ever had a period of smoking cigarettes daily but had not smoked every day in the past 30 days, they also were asked for the age, the year, and the month when they last smoked cigarettes on a daily basis.

Specifically, respondents who last used a given drug more than 30 days ago[118] were asked how old they were when they last used that drug. If respondents reported last using the drug within 1 year of their current age, they were asked to report the specific month and year when they last used, with the allowable years ranging from 2012 to 2014. If respondents reported last using the drug at their current age and their birth month was earlier than the interview month (i.e., they reached their current age in the same year that they were interviewed), the CAI program assumed that the last use of the drug occurred in the current year (i.e., 2014). These respondents were asked only for the month that they last used in the current year. The remaining respondents who last used a drug within 1 year of their current age could be routed to one of two

[118] Subsequent discussion also applies to respondents whose last period of smoking cigarettes every day occurred more than 30 days ago.

possible questions on the specific year they last used. They then were routed to a question to report on the specific month that they last used the drug in the year they had reported previously.

Because the routing logic to the different versions of the MLU and YLU questions was mutually exclusive, composite sets of MLU and YLU variables were created from the individual unedited variables. In addition, if respondents indicated a specific year that they last used a drug, the final YLU variables for 2014 were recoded to replace codes from the questionnaire with values for the years (i.e., 2012 through 2014). If respondents confirmed that they last used a drug at their current age and were interviewed subsequent to their birthday, a code of "2014" was assigned to the YLU; this was done even if respondents did not know what month they last used in the current year, or if they refused to report what month they last used in the current year. If the MLU and YLU questions had been skipped because respondents last used the drug more than 1 year younger than their current ages, legitimate skip codes were assigned to the final MLU and YLU variables.

Since 2005, consistency checks have been included in the module that were triggered when the values for the MLU and YLU were inconsistent with the ALU. Specifically, for respondents who recently stopped use of a given drug, the CAI program calculated a second ALU based on the MLU and YLU data by comparing these data with the respondent's date of birth. This comparison was not done if the respondent reported last use of the drug in the same month that he or she was born; a unique ALU could not be determined from the MLU and YLU in these situations because it was not known whether the drug use occurred before or after the respondent's birthday. Similarly, a consistency check was not triggered if the respondent had missing data in either of the month or year questions, such as if the respondent knew the year when he or she last used a drug but did not know the MLU.

In the remaining situations in which respondents provided complete data for the MLU and YLU, a consistency check was triggered if the MLU and YLU suggested that respondents stopped use of the drug at an earlier or a later age than what they had previously reported in their ALU question. For example, a consistency check was triggered if a 16-year-old respondent reported last using a drug at age 16 but then reported last using the drug in a month and year that would have meant the respondent was 15 years old when he or she last used the drug. No editing needed to be done if respondents indicated twice in a row that the ALU that was calculated from the MLU and YLU was correct. The CAI program updated the value for the ALU (e.g., AGELSTCG for cigarettes) to agree with the values for the MLU and YLU.

If respondents indicated at some point in the consistency check sequence that the value they had reported for their ALU (e.g., question LU03 for cigarettes) was correct, they had an opportunity to revise the values for their MLU and YLU. If a consistency check was triggered between the ALU and data in the MLU and YLU, the MLU and YLU were updated with any year and month data that the respondent entered in the consistency checks (e.g., LUCG07 and LUCG07a for any cigarette use). These data were used in subsequent editing steps. Otherwise, the MLU and YLU data were picked up from the original source variables (e.g., LU03a through LU03d for any cigarette use) for use in subsequent editing.

Table B.18 in Appendix B presents the edits that were implemented when consistency checks were triggered between the ALU, MLU, and YLU. The default when a respondent did not resolve an inconsistency between the ALU and the MLU and YLU was to favor the ALU in subsequent editing decisions.

Respondents also were skipped out of the MLU and YLU questions if they indicated that they last used a drug (or last smoked cigarettes every day) at an age that was more than 1 year younger than their current age. In these situations, the edited MLU and YLU variables were assigned legitimate skip codes.

Prior substance use variables were not edited with respect to *imputed* core drug use variables. Suppose, for example, that respondents did not know or refused to report when they first used marijuana, but they gave ages at last use for marijuana in question LU02 that were consistent with their current ages. Although the potential existed for the imputed marijuana AFU (imputed variable: IRMJAGE) to be imputed to a value greater than the ALU, the marijuana ALU (MRJAGLST) was not edited for consistency with IRMJAGE.

Similarly, if a core drug recency variable (e.g., MJREC for marijuana) had been assigned an "indefinite" value of 8 or 9 (indicating use at some point in the past 12 months or lifetime, respectively; see Section 6.2.2.3), the corresponding imputed recency (e.g., IRMJRC for marijuana) could be statistically imputed to indicate past month use (IRMJRC = 1). Although the prior substance use module was designed for respondents who were not past month users, any data in this module were retained for respondents who were statistically imputed to be past month users. In this situation, analysts would have the option of deciding whether to use or disregard data from respondents who were imputed to be past month users.

Table B.19 describes the specific edits that were implemented for the ALU, MLU, and YLU variables in the prior substance use module (i.e., other than the edits described in Table B.18 when consistency checks were triggered between the ALU, MLU, and YLU variables). For example, a consistency check was triggered if respondents entered an ALU that was earlier than the age when they reported first using drugs or cigarettes, or when they first smoked cigarettes daily. Because these ALU questions occurred in a noncore module, respondents were not allowed to change their answers to the corresponding AFU questions from the core modules. Thus, the only way that respondents could resolve the inconsistency between the ALU and AFU was to change their answer to the ALU question. If respondents indicated that their inconsistent ALU was correct or they entered a new ALU value that was still inconsistent with the AFU, the edited ALU variable (e.g., MRJAGLST for marijuana) was assigned a bad data code. Thus, the relevant AFU from the core modules was used as the standard against which the corresponding noncore ALU variable was compared. Similarly, the MFU and YFU questions for a given drug from the core section of the interview were used as standards for editing the MLU and YLU variables for that drug.

As was the case in prior years, methamphetamine users who reported lifetime but not past month use of methamphetamine in the core stimulants module were routed into the questions for last use of methamphetamine. Since 2007, methamphetamine users identified in the special drugs module (Section 6.2.6) also have been routed into these questions. The question wording for the age, year, and month when respondents last used methamphetamine differed depending on whether use was reported in the core stimulants module or in the noncore special drugs module. If respondents reported methamphetamine use in the core stimulants module, the wording of questions for the ALU, MLU, and YLU asked about use of "Methamphetamine, Desoxyn, or Methedrine." For users identified through the special drugs module, the questions about the ALU, MLU, and YLU simply referred to use of "Methamphetamine."

Because methamphetamine users who reported use in the core stimulants module were asked about prescription forms of the drug (e.g., Desoxyn®), the edits described in Table B.19 continued to hold for the ALU, MLU, and YLU of prescription stimulants when methamphetamine users from the core stimulants module reported a later ALU, MLU, and YLU than they reported for last use of *prescription* stimulants. However, these edits did not apply when respondents were identified as methamphetamine users through the special drugs module. In this latter situation, the question wording "Methamphetamine" was assumed to refer to the "street" form of the drug and not to prescription forms. For this reason, documentation of codes that indicated lifetime nonuse of stimulants was modified from "NEVER USED STIMULANTS" to "NEVER USED PRESCRIPTION STIMULANTS" beginning in 2007.

7.4.7.3 Sources of Psychotherapeutic Drugs

Since 2005, the prior substance use module has included questions on how nonmedical users of prescription pain relievers, prescription tranquilizers, prescription stimulants, methamphetamine, and prescription sedatives obtained the medications they misused in the past 30 days or past 12 months. For stimulants, these questions differentiated between stimulants that are typically available by prescription and methamphetamine, which is typically manufactured outside the legitimate pharmaceutical industry by illicit laboratories and distributed through illegal trafficking.[119]

For all of these drugs except methamphetamine, respondents were given a list of 10 potential sources of prescription medications. These sources included prescriptions from one or more doctors, fake prescriptions, thefts from medical facilities, the Internet, drug dealers, and friends or relatives (obtained with or without the knowledge of friends or relatives). Respondents also had the option of indicating that they obtained these medications "in some other way." Respondents who indicated that they obtained these medications in some other way were asked to specify what that other way was.

Questions about how methamphetamine users obtained this drug included a reduced list of six potential sources of the drug. Unlike the psychotherapeutic drugs that often may be available by prescription, the methamphetamine questions did not include options for respondents obtaining methamphetamine by prescription (including fake prescriptions written for methamphetamine) or by stealing methamphetamine from medical facilities or pharmacies. Response options for methamphetamine included obtaining the drug from a friend or relative for free, buying it from a friend or relative, taking it from a friend or relative without asking, buying it from a drug dealer or other stranger, buying it on the Internet, or getting it "in some other way." Again, respondents who reported that they got methamphetamine in some other way were asked to specify how they obtained it.

---

[119] Although methamphetamine also is available in prescription form (e.g., Desoxyn®), legitimate prescribing of methamphetamine in the United States is relatively rare.

Since 2006, respondents who reported that they obtained pain relievers, tranquilizers, stimulants, sedatives, or methamphetamine from a friend or relative for free have been asked follow-up questions (e.g., LU27A for pain relievers) on how the friend or relative originally obtained the drug. These items were added in 2006 due to the large number of respondents in 2005 who reported obtaining these drugs from a friend or relative for free. Similar to their predecessor questions, for all of these drugs except methamphetamine, respondents were given the same list of 10 sources for where a friend or relative could have obtained the drugs. Respondents also had the option of indicating that the friend or relative obtained these medications "in some other way"; in that situation, respondents then were asked to specify the other way that the friend or relative obtained the medication.

Respondents who reported that they last used a given psychotherapeutic drug in the past 30 days[120] were asked to report all of the ways that they obtained that drug in the past 30 days (e.g., question LU27 for pain relievers). Thus, the questions pertaining to how past month nonmedical users obtained these drugs in that period were "enter all that apply" questions in which respondents could indicate more than one source of these drugs. Each response option, such as "I got the pain reliever from a friend or relative for free" in question LU27 for pain relievers, was captured as a separate variable (ANLFRFRE for this option). The individual variables were coded as 1 if the response was chosen and were coded as 6 if the response was not chosen (Section 2.4.4). Codes of 94 and 97 (for "don't know" and "refused" respectively) in these variables indicated that respondents did not know or refused to report the source of the psychotherapeutic drugs that they used nonmedically in the past 30 days (Section 2.2.1).

Similar edits applied to the variables pertaining to how friends or relatives obtained medication that respondents subsequently used nonmedically. Specifically, if respondents reported that they last used a psychotherapeutic drug nonmedically in the past 30 days and they reported that they got the drug from a friend or relative for free, respondents were asked to report all of the ways that the friend or relative obtained that drug in the past 30 days (e.g., LU27A for pain relievers). In this situation, the corresponding variables were coded as 1, 6, 94, or 97, as described previously.

Respondents who used these drugs in the past 30 days and indicated more than one source of these drugs in that period were asked to report how they obtained the drugs that they last used (e.g., question LU28 for pain relievers). Similarly, respondents whose most recent use of a given psychotherapeutic drug was more than 30 days ago but within the past 12 months[121] were asked to report how they obtained the drug the last time that they used it. Unlike the "enter all that apply" variables pertaining to how past month nonmedical users obtained these drugs, respondents were allowed to enter only one response for how they obtained the

[120] For pain relievers, this included respondents who reported using OxyContin® nonmedically in the past 30 days. For methamphetamine, this included respondents who reported last using methamphetamine in the follow-up questions in the special drugs module and those who reported last using methamphetamine with a needle in the special drugs module (see Section 6.2.6), as well as those respondents who reported last using methamphetamine in the past 30 days in the core stimulants module.

[121] Again, for pain relievers, this included nonmedical use of OxyContin® more than 30 days ago but within the past 12 months. For methamphetamine, this included indications of use more than 30 days ago but within the past 12 months from either the core stimulants module or from questions in the noncore special drugs module (see Section 3.1).

psychotherapeutics for their last nonmedical use. Thus, for the edited variables ANLGTLAS, TRNGTLAS, STMGTLAS, and SEDGTLAS (pertaining to how respondents got pain relievers, tranquilizers, prescription stimulants, or sedatives, respectively, the last time they used them), codes in these variables corresponded to those in the corresponding questions LU28, LU30, LU32, and LU36 for these drugs.

If respondents reported that they last obtained a psychotherapeutic drug from a friend or relative for free in the past 12 months (e.g., in ANLGTLAS for pain relievers), similar edits applied to the variables pertaining to how friends or relatives obtained the medication that respondents last used nonmedically. Respondents could indicate only one way that the friend or relative obtained the psychotherapeutic drug that respondents last used nonmedically.

For methamphetamine, question LU34 pertaining to how respondents obtained methamphetamine the last time they used it (and question LU34A pertaining to how friends or relatives obtained the methamphetamine that respondents last used) contained a reduced number of response options. Specifically, response options were not offered to respondents for obtaining methamphetamine via prescription (i.e., including fake prescriptions) or stealing it from medical facilities or pharmacies. Response categories in questions LU34 and LU34A were recoded in the corresponding edited variables MTHGTLAS (i.e., how the respondent obtained the methamphetamine the last time he or she used it) and MTHFFLAS (i.e., how the friend or relative obtained the methamphetamine that the respondent last used) to match the corresponding levels in the variables pertaining to the other psychotherapeutics. For example, level 5 in ANLGTLAS was "I got the pain reliever from a friend or relative for free." Similarly, level 5 in ANLFFLAS (i.e., how the friend or relative obtained the prescription pain reliever that the respondent last used nonmedically) was "He or she got the pain reliever from another friend or relative for free." In question LU34 for methamphetamine, however, "I got the Methamphetamine from a friend or relative for free" was the first response category. Similarly, the first response category in question LU34A was "He or she got the Methamphetamine from another friend or relative for free."

Therefore, if respondents reported that the last time they used methamphetamine, they got it from a friend or relative for free (LU34 = 1), that response was recoded to 5 in the edited variable MTHGTLAS. Similarly, if a respondent reported that he or she got the methamphetamine that he or she last used from a friend or relative for free, and the friend or relative had gotten that methamphetamine from another friend or relative for free (LU34A = 1), that response was recoded to 5 in the edited variable MTHFFLAS.

Consistent with overall editing procedures for the prior substance use module, an important aspect of editing the variables pertaining to sources of psychotherapeutic drugs involved assigning appropriate legitimate skip codes (Section 2.4.2). Conditions under which specific legitimate skip codes were assigned are discussed below for pain relievers, tranquilizers, and sedatives. Special issues for prescription stimulants and methamphetamine are discussed separately.

- If respondents reported in the relevant core section of the interview that they never were nonmedical users of prescription pain relievers, tranquilizers, or sedatives, the variables pertaining to how respondents obtained these drugs for nonmedical use and how friends or relatives obtained these drugs were assigned a code of 91 (e.g., NEVER USED PAIN RELIEVERS).

- If respondents were logically inferred in the relevant core section of the interview to have never used prescription pain relievers, tranquilizers, or sedatives nonmedically, the variables pertaining to how respondents obtained these drugs for nonmedical use and how friends or relatives obtained these drugs were assigned a code of 81 (e.g., NEVER USED PAIN RELIEVERS Logically assigned).
- If the variables for a given psychotherapeutic drug (e.g., pain relievers) had been skipped because respondents refused to answer all lifetime questions in the corresponding core module regarding whether they had ever used that type of drug nonmedically, the refusal was propagated to the skipped source of psychotherapeutics variables, including the variables on how friends or relatives obtained them.
- If respondents reported that their last nonmedical use was more than 30 days ago but within the past 12 months, the variables pertaining to how respondents obtained that drug for nonmedical use in the past 30 days and how friends or relatives obtained it were assigned a code of 93 (e.g., DID NOT USE PAIN RELIEVERS IN THE PAST 30 DAYS).
- If respondents reported that their last nonmedical use was more than 12 months ago, the variables pertaining to how respondents obtained that drug for nonmedical use in the past 30 days and how friends or relatives obtained it were assigned a code of 93 (same meaning as above). In addition, the variables pertaining to how respondents obtained that drug the last time that respondents misused it in the past 12 months (e.g., ANLGTLAS and the associated "OTHER, Specify" variable ANLGTOSP for pain relievers) and how friends or relatives obtained the drug were assigned a code of 93 (e.g., DID NOT USE PAIN RELIEVERS IN THE PAST 12 MONTHS).
- If respondents reported that they got the drug from only one source in the past month (e.g., only one answer chosen in question LU27 for pain relievers), the corresponding variable for how respondents got the drug the last time they used it (e.g., ANLGTLAS for pain relievers) was assigned a legitimate skip code. The corresponding "OTHER, Specify" variable (e.g., ANLGTOSP for pain relievers) also was assigned a legitimate skip code. In these situations, it was not necessary to ask respondents how they got the drug the last time because they logically would have gotten it from that single source.
- If respondents used a drug nonmedically in the past 30 days but they did not report that one of the ways they obtained it was from a friend or relative for free, the corresponding edited variables pertaining to how friends or relatives obtained the drug were assigned legitimate skip codes.
- If respondents were asked how they obtained a given psychotherapeutic drug the last time they used it nonmedically and they did not report that they obtained it from a friend or relative for free, the corresponding edited variable pertaining to how the friend or relative obtained the drug was assigned a legitimate skip code.

As noted previously, questions in the prior substance use module distinguished between how respondents obtained *prescription* stimulants and how they obtained methamphetamine. The logic for assigning codes of 91 or 93 that was described previously in this section for pain relievers, tranquilizers, and sedatives also applied to the variables for the source of prescription stimulants if respondents reported in the core stimulants module that they never used methamphetamine or any prescription-type stimulants nonmedically, or if their most recent reported use of any stimulants from the core stimulants module caused them to be skipped out of the questions pertaining to how they obtained prescription stimulants in the past 30 days or the past 12 months. Similarly, if respondents reported that they obtained prescription stimulants from only one source in the past 30 days, the questions regarding how they obtained prescription stimulants the last time they used them nonmedically (i.e., STMGTLAS and STMGTOSP) were assigned legitimate skip codes.

In addition, respondents were not asked how they obtained prescription stimulants for nonmedical use if the only stimulant they reported ever using in the core stimulants module was methamphetamine. In this situation, the variables pertaining to the source of *prescription* stimulants were assigned legitimate skip codes. This edit also was implemented when respondents reported lifetime use of only two stimulants: methamphetamine and "some other stimulant," but the only "other" stimulant they specified using was methamphetamine. In this situation, data in the source of prescription stimulants variables were overwritten with a code of 89.

Although the core stimulants module did not explicitly ask when respondents last used stimulants other than methamphetamine, no editing was done to the prescription stimulants variables if the core recency-of-use variables for any stimulants (STIMREC) and methamphetamine (METHREC) indicated use in the same period, such as if STIMREC and METHREC both indicated use in the past 30 days. For example, if respondents were lifetime nonmedical users of methamphetamine and other stimulants and they indicated that they last used methamphetamine in the past 30 days, they also were expected to have reported that they used any stimulants in the past 30 days. Because questions LU31 and LU32 explicitly asked respondents how they obtained prescription stimulants, it was assumed when STIMREC and METHREC both indicated use in the same period that respondents' answers to questions LU31 and LU32 pertained to how they obtained prescription stimulants, and not how they obtained methamphetamine.

Because of the relationship between the core recency variables STIMREC and METHREC, however, a consistency check was triggered in the core stimulants module if respondents reported more recent use of methamphetamine (from the core question ST19) than they reported for any stimulants (from the core question ST09). In some of these situations, it was less clear whether respondents' answers in LU31 or LU32 referred specifically to prescription stimulants and not to methamphetamine. Therefore, a flag variable (STMGTFLG) was created when a consistency check was triggered between the recency of use for any stimulant and methamphetamine in the core stimulants module. The default value in STMGTFLG was 98 (BLANK [NO ANSWER]). Situations in which STMGTFLG had values other than 98 are described below.

STMGTFLG was set to a value of 1 if valid values existed in LU31 or LU32 for how respondents obtained prescription stimulants for nonmedical use and either of the following occurred:

- respondents answered the consistency check question STCC18 (which would indicate that their previous answer in the general stimulant recency question ST09 was incorrect),[122] regardless of whether they resolved the inconsistency between the most recent use of any stimulant and methamphetamine; or
- respondents did not resolve the inconsistency, but the edited stimulant recency STIMREC was logically edited to more recent use based on the methamphetamine recency METHREC.

For example, suppose respondents reported in question ST09 that they last used any stimulant "more than 12 months ago," but they reported in ST19 that they last used methamphetamine "more than 30 days ago but within the past 12 months." If these respondents changed their stimulant recency to "more than 30 days ago but within the past 12 months" in STCC18, they would be routed to question LU32, regarding how they obtained *prescription* stimulants the last time they used them nonmedically. If these respondents reported obtaining prescription stimulants in a way that individuals also might obtain methamphetamine, it could be questionable to assume that these answers in LU32 referred specifically to prescription stimulants and not to methamphetamine. STMGTFLG also could be set to a value of 1 if respondents reported in question STCC17 that "neither answer was correct" for their most recent use of any stimulant or methamphetamine and they indicated use of stimulants in the past month or past year in STCC18.

STMGTFLG was set to a value of 2 if valid values existed in LU31 or LU32 when the following occurred:

- respondents answered STCC19 (but not STCC18, which would indicate that their previous answer for when they last used methamphetamine was incorrect, but that the stimulant answer was correct), and
- they revised their methamphetamine recency (METHREC) to be consistent with the recency for any stimulants (STIMREC).

For these cases where STMGTFLG = 2, it could be more reasonable to assume that answers in LU31 or LU32 pertained to prescription stimulants and not methamphetamine. However, this value in STMGTFLG would still alert analysts to the occurrence of an inconsistency in the core stimulants data between when respondents reported last using any stimulants and methamphetamine.

The procedures for assigning values to STMGTFLG did not include situations in which LU31 or LU32 had missing values because the respondents had not resolved the inconsistency between the most recent use of any stimulant and methamphetamine, and STIMREC had been edited to infer use in the past month or past year, based on data in METHREC. Suppose, for example, that respondents initially reported last using methamphetamine in the past 30 days but they reported last using any stimulant more than 30 days ago but within the past 12 months, and

[122] Question STCC18 is asked if respondents indicated in question STCC17 that their methamphetamine recency from question ST19 was correct (i.e., and by extension, that their general stimulant recency was incorrect) or that neither answer to their general stimulant recency and methamphetamine recency was correct. In question STCC18, respondents are asked again to report when they last used any stimulant nonmedically.

they did not resolve this inconsistency when prompted to do so. For these respondents, STIMREC was assigned a value of 11 (Used in the past 30 days LOGICALLY ASSIGNED; see Sections 6.2.2.3 and 6.2.2.6). Because these respondents reported last using stimulants more than 30 days ago but within the past 12 months, they would be asked LU32 but they would be skipped out of LU31. In this example, the default assumption was applied that the answers in LU32 pertained to how respondents obtained prescription stimulants because respondents appeared to be making a distinction between "stimulants" and methamphetamine.

Also in this example, no editing was done to the variables from LU31 regarding how respondents obtained prescription stimulants for nonmedical use in the past 30 days. Specifically, it was not inferred that the source of prescription stimulants for respondents' last nonmedical use of prescription stimulants from LU32 applied to how respondents obtained prescription stimulants in the past 30 days because respondents may have used only methamphetamine but not prescription stimulants in the past 30 days. Likewise, values of 93 were not assigned to the 30-day prescription stimulant variables because some respondents may have misused prescription stimulants in the past month.

The skip logic for the variables pertaining to how respondents obtained methamphetamine took into account respondents' answers to the core methamphetamine questions in the stimulants module and the follow-up questions on methamphetamine from the special drugs module (Section 6.2.6). Thus, a code of 91 (Section 2.4.2) was assigned to the methamphetamine variables if respondents reported one of the following:

- they indicated in both the core stimulants module and on follow-up in special drugs that they never used methamphetamine;
- they did not know or refused to report in the core stimulants module whether they ever used methamphetamine, but they indicated in special drugs that they never used it; or
- they explicitly indicated in the core stimulants module that they never used methamphetamine, but they did not know or refused to report on follow-up in the special drugs module whether they had ever used it.

Similarly, a code of 93 was assigned to the source of methamphetamine variables in one of two ways: (1) respondents reported in the core stimulants module that their last use of methamphetamine was outside of the period(s) of interest for asking the methamphetamine questions LU33 or LU34, or (2) respondents did not report methamphetamine use in the core stimulants module, but they reported use in special drugs, with their last use being outside of the period(s) of interest for LU33 or LU34. For example, if respondents did not report methamphetamine use in the core stimulants module but they reported in the special drugs question SD17b that they last used it more than 30 days ago but within the past 12 months, the variables corresponding to question LU33 were assigned a code of 93, and data from LU34 (and the "OTHER, Specify" variable LU34SP, if applicable) were assigned to the edited variable MTHGTLAS (and to MTHGTOSP, if applicable). A code of 93 also was applied to the variables pertaining to how friends or relatives obtained methamphetamine if they did not use methamphetamine in the period of interest. For example, if respondents did not use methamphetamine in the past 12 months, then the edited variables MTHFFLAS and MTHFFLSP (pertaining to how friends or relatives obtained the methamphetamine that respondents last used) were assigned a code of 93.

Miscellaneous edits based on skip logic also applied to the source of psychotherapeutics variables. These edits applied to all of the psychotherapeutic drugs, with situations being cited for pain relievers. For example, if respondents used prescription pain relievers nonmedically in the past 30 days and they did not indicate that they obtained prescription pain relievers "in some other way" in that period in question LU27, then the corresponding "OTHER, Specify" variable (e.g., ANLOTHSP, corresponding to question LU27SP) was assigned a code of 99. If respondents refused to report how they got pain relievers that they used nonmedically in the past 30 days, ANLOTHSP also was assigned a code of 97 (REFUSED). Similarly, if respondents used pain relievers nonmedically in the past 12 months, were asked question LU28, and did not indicate that they got pain relievers "in some other way" the last time they used them nonmedically, then the edited "OTHER, Specify" variable ANLGTOSP (corresponding to question LU28SP) was assigned a code of 99. If respondents refused to report how they got pain relievers the last time they used them nonmedically, ANLGTOSP also was assigned a refusal code.

Levels 1 through 9 in the "OTHER, Specify" variables were used for responses that corresponded to existing response options. For example, a code of 5 was assigned to ANLOTHSP (corresponding to question LU27SP) if respondents reported in LU27SP that they got pain relievers in the past 30 days from a friend or relative for free. These same coding categories applied to the "OTHER, Specify" variables for methamphetamine. Thus, a code of 5 also was assigned to MTHOTHSP (corresponding to question LU33SP) if respondents reported getting methamphetamine from a friend or relative for free in the past 30 days, even though this was the first response option in question LU33.

Table B.20 describes additional edits that were relevant to the source of psychotherapeutics variables. For these edits, the source of pain relievers variables are used as examples, although these edits also applied to the other psychotherapeutic drugs. Where relevant, the edits that are described in Table B.20 for how respondents obtained psychotherapeutic drugs also were applied to the variables for ways that friends or relatives obtained psychotherapeutic drugs.

7.4.7.4 Sequence of Initiation

If respondents first used alcohol and cigarettes, cigarettes and marijuana, alcohol and marijuana, or all three substances at the same age, they were asked to report which of these they used first. For example, if respondents indicated that they first used alcohol and marijuana at the same age, they were asked which of these they had used first.

- Questions LU22, LU23, and LU24 (corresponding to the edited variables USEALCG, USEMJCG, and USEALMJ, respectively) were asked when respondents reported first use of only two of these substances at the same age (i.e., USEALCG = use of alcohol and cigarettes at the same age; USEMJCG = use of marijuana and cigarettes at the same age; and USEALMJ = use of alcohol and marijuana at the same age).
- Questions LU25 and LU26 (corresponding to the edited variables USEACM and USENEXT) were asked when respondents reported first use of all three of these

substances at the same age; USEACM indicated which of these three substances the respondents used first, and USENEXT indicated which of the remaining two substances the respondents used next.

As was the case for the processing of other variables in the prior substance use module, an important aspect of editing these sequence-of-use variables involved assigning various legitimate skip codes, as appropriate. Conditions under which specific legitimate skip codes were assigned are discussed below.

- If respondents never used alcohol or cigarettes (regardless of whether they ever used marijuana), USEALCG was assigned a code of 91. Documentation of a code of 91 for this variable was 91 = NEVER USED ALCOHOL/CIGARETTES.
- If respondents never used marijuana or cigarettes (regardless of whether they ever used alcohol), USEMJCG was assigned a code of 91. Documentation of a code of 91 for this variable was 91 = NEVER USED MARIJUANA/CIGARETTES.
- If respondents never used alcohol or marijuana (regardless of whether they ever used cigarettes), USEALMJ was assigned a code of 91. Documentation of a code of 91 for this variable was 91 = NEVER USED ALCOHOL/MARIJUANA.
- If respondents never used alcohol, cigarettes, or marijuana, USEACM and USENEXT were assigned a code of 91. Documentation of a code of 91 for these variables was 91 = NEVER USED ALCOHOL/CIGARETTES/MARIJUANA.
- If the values in the edited AFU variables for alcohol (ALCTRY), cigarettes (CIGTRY), and marijuana (MJAGE) all were valid and equal, USEALCG, USEMJCG, and USEALMJ were assigned a code of 99 (LEGITIMATE SKIP).
- If at least one value for ALCTRY, CIGTRY, or MJAGE was valid but the values for all three were not equal, USEACM and USENEXT were assigned a code of 99.
- If at least one value for ALCTRY or CIGTRY was valid but the values were not equal, USEALCG was assigned a code of 99. Similarly, if at least one value for MJAGE or CIGTRY was valid but the values were not equal, USEMJCG was assigned a code of 99. If at least one value for ALCTRY and MJAGE was valid but the values were not equal, USEALMJ was assigned a code of 99.

Miscellaneous skip issues also applied to the data, such as if USEALCG (corresponding to LU22) was blank for some reason other than those mentioned above but some value other than 85, 91, 98, or 99 existed in questions LU23, LU24, or LU25. In these situations, a legitimate skip code was assigned to the relevant skipped variable (e.g., USEALCG). Similarly, if USEACM and USENEXT (corresponding to LU25 and LU26) had been skipped but data existed in LU22, LU23, or LU24, then USEACM and USENEXT were assigned legitimate skip codes. This logic covered residual situations in which variables might be skipped but data existed in one of the alternate variables.

Table B.21 describes additional edits pertaining to these sequence-of-use variables. For example, if ALCTRY, CIGTRY, and MJAGE all had a code of 997 (REFUSED) because respondents refused to answer the AFU questions for these drugs, or because respondents refused to answer the lifetime use question (for alcohol or marijuana), all blank values in USEALCG through USENEXT were replaced with the two-digit refusal code of 97 (Section 2.4.3). If for some reason respondents were routed into any of the questions corresponding to these variables when all of these ages at first use had a refusal code, nonblank values in USEALCG through USENEXT were overwritten with bad data codes.

7.4.8 Substance Treatment Module

The substance treatment module asked about receipt of treatment services for the use of alcohol or other drugs, not counting cigarettes. Questions about the receipt of treatment services included questions about receipt of treatment in the respondents' lifetime and in the past 12 months, specific locations where respondents received treatment in the past 12 months, emergency room visits in the past 12 months related to their use of specific drugs, whether they were still in treatment, the length of time since they were last in treatment (if they were not currently in treatment), specific questions about their last (or current) treatment episode, whether they were enrolled in treatment on October 1, 2013, and whether the only treatment they received in the past 12 months was detoxification.

Since 2007, respondents who reported methamphetamine use in the special drugs module have been eligible to be asked questions in the substance treatment module. However, this logic change affected only those respondents who reported methamphetamine use in the special drugs module and did not report use of alcohol or other drugs in the core modules; users of methamphetamine and other drugs would have been eligible to be asked the substance treatment questions based on their reported use of other drugs. A small number of respondents in 2014 (fewer than five) reported that they received substance treatment in their lifetime based solely on their reported methamphetamine use in the special drugs module. Most of these respondents did not report that they received treatment in the past 12 months.

In addition, respondents who reported methamphetamine use in the special drugs module were eligible to be asked question TX05 about visits to a hospital emergency room in the past 12 months to receive treatment for their use of cocaine, heroin, marijuana, PCP, LSD, or methamphetamine, provided that (1) they reported having received treatment in the past 12 months in question TX02 and (2) they reported in question TX03 that they received treatment in the past 12 months for their use only of drugs or their use of both alcohol and drugs. Again, all respondents who reported methamphetamine use in the special drugs module and reported receiving treatment in the past 12 months for their use of drugs (with or without receipt of treatment for alcohol use) either reported that they did not visit a hospital emergency room to receive treatment for their use of cocaine, heroin, marijuana, PCP, LSD, or methamphetamine, or they reported use of one of these other drugs in addition to their use of methamphetamine.

Questions also have been included since 2004 to capture information about respondents' life history of substance treatment. Respondents who had ever received treatment for their use of alcohol or other drugs but did not receive treatment in the past 12 months were asked questions regarding their receipt of treatment for their use of alcohol, other drugs, or both, depending on the substances they had reported using.[123] Where relevant, these respondents subsequently were asked to report the ages when they first received treatment for alcohol, drugs, or both. Respondents who reported that they had received treatment in the past 12 months were asked similar questions. Questions about treatment life history for respondents who reported that they had received treatment in the past 12 months also took into account their answers to question TX03, regarding whether they received treatment in the past 12 months for their use of alcohol, drugs, or both.

Since 2006, questions TX52 and TX53 have been included in the substance treatment module to ask whether respondents had attended a self-help group in the past 12 months (question TX52; edited variable TX12MSHG), and if so, whether this was for alcohol use, drug use, or both (TX53; edited variable TX12SGAD). These items were administered to respondents who previously indicated that they used alcohol or other drugs but they did not report receiving *treatment* through a self-help group in the past 12 months; use of the term "treatment" is inconsistent with the language used in the self-help/recovery community. Therefore, these items were added to assess whether asking respondents about treatment they received could result in underestimates of attendance at self-help group meetings.

Questions about the last or current treatment episode were asked principally of respondents who reported that they received treatment in the past 12 months (question TX02 answered as "yes"); the logic also routed respondents to the last or current treatment questions if they did not know or refused to report in question TX02 whether they had received treatment in the past 12 months. If respondents received treatment in the past 12 months (or answered question TX02 as "don't know" or "refused") and reported in question TX07 that they were currently in treatment,[124] subsequent questions asked about the main location where they were receiving treatment, specific drugs for which they were receiving treatment, the primary drug for which they were receiving treatment (if treatment for more than one drug was reported), the length of time that they had been in treatment thus far, and anticipated payment sources for their current treatment. If respondents were asked question TX07 and did not report currently being in treatment, these subsequent questions pertained to their last treatment episode, such as the duration of their last treatment and the payment sources for their last treatment. Respondents who did not report that they were currently in treatment also were asked about the outcome of their last treatment.

The substance treatment module also included questions about respondents' perceived need for treatment in the past 12 months if they never received treatment or did not report that they received treatment in the past 12 months. Questions about respondents' perceived need for treatment included questions about specific drugs for which respondents thought they needed treatment and whether they made specific efforts to receive treatment in the past 12 months. In addition, respondents who received treatment in the past 12 months but did not report that they were currently in treatment were asked whether they felt the need for *additional* treatment in the past 12 months. Those respondents who reported that they felt the need for additional treatment

[123] Since 2007, reports of substance use have been based on reports in the core modules or reports of methamphetamine use from the special drugs module.

[124] Question TX07 asks, "Are you currently receiving treatment or counseling for your [TXFILL1]?" where [TXFILL1] could be replaced with "alcohol use," "drug use," or "alcohol or drug use."

were asked about the specific drugs for which they needed additional treatment and whether they made specific efforts to receive additional treatment.

As noted previously, the substance treatment module was relevant only for respondents who reported some lifetime use of alcohol or other drugs, not counting cigarettes. Therefore, all of the edited treatment variables were assigned a code of 91 (i.e., NEVER USED ALCOHOL OR DRUGS) if respondents were skipped out of the entire substance treatment module because they never used alcohol, illicit drugs (including methamphetamine), or prescription-type psychotherapeutics for nonmedical reasons (i.e., pain relievers, tranquilizers, stimulants, or sedatives).

In situations where respondents' only lifetime use of drugs involved use of OTC medications that were reported in one or more of the psychotherapeutics modules (i.e., and they did not report methamphetamine use in the special drugs module), a code of 81 was assigned to all of the edited substance treatment variables (i.e., NEVER USED ALCOHOL OR DRUGS Logically assigned). This was done to signify that these respondents were logically inferred to be lifetime nonusers of alcohol or drugs. This code of 81 also set these respondents apart from those whose original answers indicated that they had never used any of these drugs. These edits for assigning codes of 91 or 81 also applied to the self-help variables TX12MSHG and TX12SGAD.

7.4.8.1 Receipt of Substance Treatment Services

An important aspect of the processing of the substance treatment variables involved assignment of relevant legitimate skip codes when it could be determined unambiguously from respondents' answers that subsequent questions did not apply. In particular, respondents who were lifetime users of alcohol or at least one other drug were asked if they had ever received treatment for their alcohol or other drug use, not counting cigarettes. If respondents reported that they never received treatment (i.e., TXEVER = 2), the CAI program skipped them out of all remaining questions pertaining to the receipt of treatment services. Thus, if respondents clearly indicated that they never received treatment, the skipped treatment service variables were assigned legitimate skip codes (Section 2.4.2). As described in Section 2.4.3, when the treatment service questions were skipped because respondents refused to indicate whether they ever received treatment, the edited variables were assigned a refusal code; if treatment service questions were skipped because respondents did not know whether they ever received treatment, the edited variables retained a value of blank.

Similarly, respondents were not asked subsequent questions about receipt of treatment services in the past 12 months if they did not report having ever received treatment in that period (i.e., TXYREVER = 2). Thus, if respondents reported that they did not receive treatment in the past 12 months and there were no other responses in the substance treatment module to suggest that they had (see below), legitimate skip codes were assigned to the variables pertaining to receipt of treatment in specific locations in the past 12 months. The procedures for editing 12-month treatment variables that had been skipped when respondents refused to indicate whether they had received treatment in the past 12 months or did not know whether they had received treatment in this period were the same as those described above.

If respondents reported that they received treatment in the past 12 months, it was possible for them to be asked subsequent questions about treatment in an emergency room in the past 12 months for their use of marijuana, cocaine, heroin, LSD, PCP, or methamphetamine. Respondents were not asked these questions if they previously reported that their treatment in the past 12 months was only for their use of alcohol. Thus, "legitimate skip" codes were assigned to the edited variables pertaining to emergency room use (TXYRVSER and TXYRNMER), provided there were no other answers in the substance treatment module that indicated treatment for use of these drugs, which would suggest that respondents should have been asked these questions. Similarly, legitimate skip codes were assigned to the edited variable pertaining to the number of emergency room episodes for treatment of these six drugs (TXYRNMER) if respondents reported that they never received treatment in an emergency room related to their use of these drugs.

In addition, respondents who reported receiving treatment in the past year were not asked certain questions about receipt of treatment related to their use of specific drugs if they were lifetime nonusers of these drugs. For example, respondents who never used heroin were not asked whether they last received (or were currently receiving) treatment for their use of heroin. Similarly, respondents who reported receiving treatment in the past 12 months but who never used marijuana, cocaine, heroin, LSD, PCP, or methamphetamine were not asked the questions about use of hospital emergency room services for the use of these drugs. Rather than assign the usual type of legitimate skip code (i.e., 99 or 89), however, a special code of 6 was assigned in these situations, provided that the respondent had not indicated receipt of treatment for any of these drugs elsewhere in the substance treatment module. This code had the following meaning: 6 = Never used the relevant drug.

This coding was done because respondents could be routed into or skipped out of a number of different combinations of questions depending on their reported drug use history. For example, a respondent who reported that he or she had received treatment in the past 12 months and was a lifetime user of alcohol, marijuana, cocaine, hallucinogens, prescription pain relievers, and prescription stimulants would selectively be asked the questions about treatment for these drugs during his or her last treatment or current episode and would not be asked the questions pertaining to treatment for heroin, inhalants, prescription tranquilizers, and prescription sedatives.

When respondents were skipped out of a question related to treatment for a given drug because they refused to indicate whether they had ever used that drug, the refusal was propagated onto the edited variable pertaining to treatment for that drug. For example, if a respondent reported receiving treatment in his or her lifetime but refused to indicate whether he or she had ever used heroin, the question about treatment for heroin during the last treatment episode was skipped. The edited variable pertaining to treatment for heroin (TXLTYHER) was therefore assigned a refusal code.

As noted above, respondents who did not report that they received treatment in the past 12 months were not asked questions about their last treatment episode. Therefore, if the final edited variable pertaining to receipt of treatment in the past 12 months indicated that respondents had not received treatment during this period (i.e., TXYREVER = 2), the variables pertaining to the last treatment episode were assigned legitimate skip codes.

Most of the editing of the substance treatment questions TX45 through TX51A that have been present since 2004 also involved assigning legitimate skip codes where relevant. Consistent with the logic described above, if respondents had never used alcohol or other drugs, these variables were assigned codes of 91 or 991 (or 81 or 981, if their only use of drugs involved OTC medications). In addition, if respondents reported in question TX01 that they had never received treatment, these variables were assigned legitimate skip codes (e.g., 99 or 999). Questions TX45 through TX48A also pertained to respondents who had received treatment but not in the past 12 months. Therefore, if respondents reported in question TX02 that they received treatment in the past 12 months, the edited variables corresponding to questions TX45 through TX48A were assigned legitimate skip codes. In addition, when respondents reported receiving treatment in the past 12 months, subsets of the variables corresponding to questions TX49 through TX51A were assigned legitimate skip codes based on answers in question TX03 regarding receipt of treatment in the past 12 months for alcohol, drugs, or both. Variables corresponding to TX49 through TX51A also were assigned legitimate skip codes based on indications in the core modules or in the special drugs module for methamphetamine that respondents never used alcohol or never used any illicit drugs. Similar assignment of legitimate skip codes occurred for the variables corresponding to TX49 through TX51A if respondents received treatment in their lifetime but not in the past 12 months, or depending on respondents' answers in the core drug modules or in special drugs for methamphetamine. For example, if respondents answered question TX02 as "no" regarding receipt of treatment in the past 12 months, the variables corresponding to TX49 through TX51A were assigned legitimate skip codes; by definition, these respondents had to have answered the lifetime treatment question TX01 as "yes."

Table B.22 in Appendix B presents additional edits that were specific to the variables for the receipt of treatment services for variables that existed prior to 2004. For example, the answers to the questions on receipt of treatment in the past 12 months and the last time that respondents received treatment could be inconsistent. Specifically, respondents could report that they received treatment in the past 12 months (TX02 = 1) but then subsequently report that the last time they received treatment was more than 12 months ago (TX24 = 3). For these respondents, the recency of treatment was inferred to be at some point within the past 12 months (TXLASREC = 8). Respondents also could provide an answer other than "yes" when asked in question TX02 whether they had received treatment in the past 12 months and then indicate that they last received treatment in the past 30 days or more than 30 days ago but within the past 12 months (TX24 = 1 or 2). In these situations, the respondents were logically inferred to have received treatment in the past 12 months. Similarly, respondents could answer "don't know" or "refused" when asked whether they had received treatment in the past 12 months and then report that they last received treatment more than 12 months ago. In this situation, a negative response was logically inferred for the variable pertaining to receipt of treatment services in the past 12 months (TXYREVER = 4).

In addition, composite variables combining data from more than one individual item were created for the following:

- the main place where respondents received (or were receiving) treatment during their last (or current) treatment episode (TXLTYMN);

- the outcome of the last treatment episode, for respondents who were not currently in treatment (TXLTYOUT); and
- the length of time that respondents had been in treatment or currently had been in treatment thus far (TXLTYDUR).

For the first two variables listed above, respondents could select a response category from a list, including selection of an "other" category (e.g., treatment in some other place). Only those respondents who chose the other category were routed into a second item where they were asked to specify the other location or the other outcome of their treatment. Consequently, the final variables for the main place where respondents received (or were receiving) treatment during their last (or current) treatment episode and the outcome for that last episode included data both from the existing response categories that respondents were allowed to choose and valid "other" responses that they specified. If respondents chose the other category but specified something that was coded with a missing value (i.e., "bad data," "don't know," "refused," or blank), a final code of "other" was retained for these two variables.

The variable pertaining to the length of time that respondents had been in treatment (TXLTYDUR) was derived from a question that asked respondents to indicate whether they wanted to give their answer in terms of days, months, or years, and from questions that asked for the number of days, months, or years that they were in treatment. TXLTYDUR was expressed as the number of days that respondents were in treatment. If respondents answered in terms of a number of months, their reported number of months was multiplied by 30. If respondents answered in terms of a number of years that they had been in treatment, their reported number of years was multiplied by 365.

If respondents answered in terms of a number of months in treatment, the treatment duration data also were compared for consistency with the respondent's age. Specifically, the number of months in treatment was divided by 12 to yield an estimated number of years in treatment. If the reported number of years in treatment exceeded the respondent's current age, then TXLTYDUR was assigned a bad data code. If the difference between the respondent's current age and the number of years in treatment was 10 or fewer years, this data pattern was flagged. Such respondents would have been reporting that they had *not* been in treatment for 10 or fewer years. However, TXLTYDUR was not set to bad data for this latter situation.

Table B.23 presents edits that were specific to the substance treatment variables that were added to the survey in 2004. For example, respondents could report that they first received treatment for their use of alcohol at ages that were earlier than when they first reported using alcohol. No editing was done to these data. However, flags were created to indicate whether the ages for first treatment of alcohol or other drugs were consistent with reported ages at first use from the core modules or in the special drugs module for methamphetamine (since 2007), and if not, the flags indicated the degree of inconsistency between these data. The meaning of the values in these flag variables is discussed in more detail in Table B.23.

In addition, data from substance treatment variables that existed prior to 2004 were used to edit these added variables. However, data from these added substance treatment variables were not used to edit the substance treatment variables that existed prior to 2004. Consequently, variables that also existed prior to 2004 (as well as in 2004 through 2014) were created in a manner that was comparable with how these variables were created in prior years.

An important aspect of editing the variables TX12MSHG (corresponding to question TX52) and TX12SGAD (corresponding to TX53) involved assigning relevant codes based on the skip logic for these items. For example, if respondents were routed to question TX52 and answered it as "no," then TX12SGAD was assigned a legitimate skip code. Similarly, if TX52 was refused, that refusal was propagated to TX12SGAD. In addition, if the lifetime treatment variable TXEVER (corresponding to question TX01) had been set to bad data, then any nonblank values also were replaced with codes for bad data in TX12MSHG and TX12SGAD (see Table B.22 in Appendix B).

If question TX52 had been skipped because respondents had reported earlier in the substance treatment module that they received treatment in a self-help group in the past 12 months (from question TX04H), then TX12MSHG was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern). In addition, TX53 was skipped when question TX04H was answered as "yes." Therefore, if values of 1, 2, or 3 existed in TXYRSGAD (corresponding to question TX04H1 and indicating treatment in a self-help group in the past 12 months for alcohol only, drugs only, or both alcohol and drugs, respectively), TX12SGAD (corresponding to TX53) was assigned corresponding codes of 11, 12, or 13. "Bumping" these values by 10 in the assignment of codes from TXYRSGAD to TX12SGAD indicated to analysts that a logical inference was made in TX12SGAD; in contrast, codes of 1, 2, or 3 in TX12SGAD came from respondents answering TX53 directly. Other values greater than 3 that existed in TXYRSGAD were moved over to TX12SGAD without further editing. In particular, if TXYRSGAD (from question TX04H1) had been legitimately skipped because respondents reported in question TX03 that they were treated in the past 12 months for alcohol only or drugs only, then a legitimate skip code could be assigned to TX12SGAD as well.

Variables pertaining to the lifetime receipt of treatment (TXEVER), treatment in the past 12 months (TXYREVER), treatment in a self-help group in the past 12 months (TXYRSHG), and treatment for alcohol, drugs, or both in a self-help group in the past 12 months (TXYRSGAD) were not edited to make them consistent with data in questions TX52 and TX53. Consequently, the integrity of trends in the receipt of treatment (particularly in the past 12 months) can be preserved. This decision also was in agreement with the decision that was made in 2004 not to edit substance treatment variables that existed in the module prior to 2004 for consistency with data from questions TX45 through TX51A that were added to the substance treatment module in 2004.

However, a flag variable (TXSHGFLG) was created to alert analysts to situations in which data were inconsistent between TX12MSHG and TXYRSHG. The default value of TXSHGFLG was 98 (blank). In addition, TXSHGFLG was assigned values of 91 if respondents had never used alcohol or other drugs, and it was assigned values of 81 if the logical inference had been made that respondents had never used alcohol or other drugs. Assignment of values other than 81, 91, or 98 to TXSHGFLG is described below.

- TXSHGFLG was set to a value of 1 if respondents answered TX52 as "yes" (i.e., attended a self-help group for their own use of alcohol or other drugs in the past year) but the respondents answered question TX04H as "no" (i.e., did not receive treatment in a self-help group in the past 12 months). In this situation, the answers in TX04H and TX52 may be seen as contradictory, except that respondents may not have thought of self-help group attendance as "treatment" when they were answering question TX04H.

- TXSHGFLG was set to a value of 2 if respondents answered TX52 as "yes" but they reported either that they did not receive any treatment in the past 12 months or in their lifetime. A separate level was created for this pattern because these respondents may not have thought of self-help groups as "treatment" when they were answering question TX01 (lifetime receipt of treatment) or TX02 (receipt of treatment in the past 12 months). Consequently, respondents who attended self-help groups in the past 12 months may have answered TX01 or TX02 as "no."
- TXSHGFLG was set to a value of 3 if the edited variable TXYRSHG (corresponding to question TX04H) indicated that respondents were logically inferred to have received treatment in a self-help group in the past year, but TX12MSHG also did not indicate that these respondents attended a self-help group in the past 12 months. In this situation, TX12MSHG was inconsistent with TXYRSHG, but only because a logical inference had been made that TXYRSHG should have been answered as "yes."
- TXSHGFLG was set to a value of 4 if TX52 was answered as "yes" but the edited variable TXYRSHG had a value for "don't know," "refused," or "blank." In particular, TXYRSHG could have been blank if question TX04H had been skipped and it had subsequently been inferred that respondents got treatment in the past 12 months. In these types of situations, data in TX12MSHG and TXYRSHG were not necessarily contradictory.
- TXSHGFLG was set to a value of 5 if question TX04H was answered as "no" and TX52 was answered as "don't know" or "refused." Although a logical inference was not made in TX12MSHG that respondents did not attend self-help groups in the past 12 months, analysts would have the option of inferring that respondents absolutely did not attend self-help groups in the past 12 months for their own use of alcohol or other drugs based on TXYRSHG. Alternatively, analysts could decide to treat self-help group attendance data as missing based on TX12MSHG.

7.4.8.2 Perceived Need for Substance Treatment

The basic content of the section of the substance treatment module in 2014 pertaining to respondents' perceived need for substance abuse treatment did not change relative to 2013. Since 2007, however, respondents who reported methamphetamine use in the noncore special drugs module and who did not report that they received treatment for their alcohol or other drug use in the past 12 months have been eligible to be asked whether they felt that they needed treatment for their alcohol or other drug use at any time in the past 12 months. As noted previously, this logic change was relevant only for those respondents who reported methamphetamine use in the special drugs module and did not report use of alcohol or other drugs in the core modules. A small number of respondents in 2014 (fewer than five) reported lifetime use of methamphetamine solely from the special drugs module. These cases did not report that they received substance treatment in the past 12 months and also did not report that they felt that they needed treatment in the past 12 months.

Since 2002, respondents have been asked to report the reasons why they did not receive substance treatment services despite feeling the need for treatment (question TX22A). Similarly, respondents who reported that they received treatment but needed additional treatment have been asked why they did not receive additional treatment or counseling (question TX23A).

Based on a review of what respondents had specified in quarter 1 of 2002 as leading "other" reasons for not getting substance treatment or additional treatment, additional questions (TX22B and TX23B) have been included in the interview since 2003 to capture commonly endorsed other reasons for not receiving treatment. Therefore, if respondents answered question TX08 as "yes" (i.e., perceived the need for substance treatment), they eventually were routed to question TX22A, where they could report up to 10 reasons for not receiving substance treatment, including "some other reason or reasons." If respondents chose the "some other reason or reasons" option in question TX22A, they were routed next to question TX22B, where they could choose additional reasons for not getting substance treatment; again, respondents were given the option in question TX22B to report "some other reason or reasons." If respondents chose this "other" response category in TX22B, they were asked to specify the *most important* other reason why they did not get treatment. Thus, respondents were asked to specify only the most important other reason for not getting substance treatment. Similar logic was in place for questions TX23A and TX23B, when respondents reported in TX09 that they perceived a need for additional substance treatment.

Questions TX22A, TX22B, TX23A, and TX23B were "enter all that apply" questions in which respondents could choose more than one reason from each list. Each response option (e.g., "You had no health care coverage, and you couldn't afford the cost") was captured as a separate variable. The edited variables corresponding to the individual response options in TX22A, TX22B, TX23A, and TX23B were coded as 1 (Response entered) or 6 (Response not entered), if at least one item was chosen from the TX22A, TX22B, TX23A, and TX23B lists (Section 2.4.4).

In addition, the 10th response option in questions TX22A and TX23A ("some other reason or reasons") was principally considered to be a "toggle" to questions TX22B and TX23B, respectively. Therefore, separate "some other reason or reasons" variables were not created to correspond to the last response category in TX22A and TX22B. Similarly, separate variables were not created to correspond to the last category in TX23A and TX23B.

The edits described below were implemented in 2003 to take into account the addition of questions TX22B and TX23B. The edits described use TX22A and TX22B as examples but also applied to TX23A and TX23B.

- If respondents chose the 10th response option in TX22A, any response that was entered from the TX22B series was coded as 1, and anything that respondents did not choose from the TX22B list was coded as 6.
- If respondents chose a response from TX22A but did not choose the 10th response category in TX22A, the variables corresponding to the response categories in TX22B (i.e., NDTXNOND through NDTXSOR) all were given a code of 6 (Response not

entered), rather than being assigned "legitimate skip" codes. That is, TX22A and TX22B were considered together to be one big series of reasons.

- If respondents chose category 10 in question TX22A, chose at least one reason from TX22B, but did not choose category 6 in TX22B (some other reason or reasons), the edited variable NDTXSOR (some other reason or reasons for not receiving substance treatment) was assigned a code of 6. That is, it was inferred in this situation that the list of specific reasons in TX22B was adequate for capturing why respondents did not get substance treatment. For example, if a respondent chose category 10 in TX22A and then chose only category 2 in TX22B ("You thought you could handle the problem without treatment"), it would be reasonable to infer that this response in TX22B was the only other reason why the respondent did not get treatment.
- If respondents chose response category 10 in question TX22A, it was possible for them to answer TX22B as "don't know" or "refused" (i.e., did not know or refused to report what the other reasons were). When this occurred, the "some other reason" variable NDTXSOR was set to 1 (Response entered) in order to retain information that the respondent chose "some other reason or reasons" somewhere in the series. Remaining variables corresponding to the TX22B series retained codes of 94 ("don't know") or 97 ("refused").
- If respondents answered question TX22A as "don't know" or "refused," question TX22B was skipped. Therefore, the relevant code of 94 or 97 was propagated onto the variables corresponding to the TX22B list.
- If NDTXSOR had a value of 6 (see above), the "OTHER, Specify" variable NDTXRIMP was assigned a legitimate skip code. If NDTXSOR had a refusal code, that refusal was propagated onto NDTXRIMP.
- If NDTXSOR had a code of 1 when the respondent answered "don't know" or "refused" to the TX22B series, the "OTHER, Specify" variable NDTXRIMP retained a code of 98 (blank).

Consistent with general editing procedures, if respondents reported a reason that corresponded to a reason in the lists for TX22A/TX22B or TX23A/TX23B, that reason was logically inferred to have been chosen in the relevant edited variable (Section 2.4.5). Suppose, for example, that a respondent had not received treatment but felt the need for it, and the respondent specified that one of the reasons for not receiving treatment was that he or she was not ready to stop using alcohol or drugs. If the respondent had not chosen this response in TX22A, the edited variable NDTXREDY (corresponding to response category 5 in TX22A) was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). Similarly, if the respondents specified that they did not get treatment because they thought they could handle the problem without treatment but had not chosen that reason in TX22B, the edited variable NDTXHNDL (no substance treatment because the respondent thought he or she could handle the problem without treatment) was assigned a code of 3.

Conversely, if respondents did not report "some other reason" why they did not receive treatment in the past 12 months (edited variable NDTXSOR = 6, corresponding either to response category 10 in question TX22A not being chosen or response category 6 in question TX22B not being chosen), legitimate skip codes were assigned to the edited "OTHER, Specify" variable NDTXRIMP (corresponding to question TX22SP). Similar edits were done for the "OTHER, Specify" variable pertaining to reasons for not receiving additional treatment if respondents reported that they felt the need for additional treatment but did not indicate "some other reason" for not receiving additional treatment.

As was the case with the variables pertaining to receipt of treatment services, an important aspect of the processing of the variables pertaining to perceived need for treatment involved assigning relevant legitimate skip codes (Section 2.4.2). In particular, the variables on perceived need for treatment were compared with data on receipt of treatment services in the past 12 months. For example, if respondents had received treatment services in the past 12 months, the questions about perceived need for treatment in that period did not apply. Thus, legitimate skip codes were assigned to the variables pertaining to the perceived need for any alcohol or other drug treatment when respondents had received treatment in the past 12 months. Similarly, if respondents received treatment in the past 12 months and they reported that they were still in treatment (TXRCVNOW = 1), the questions about perceived need for additional services did not apply, and legitimate skip codes were assigned to the corresponding edited variables.

Respondents who had not indicated that they received treatment in the past 12 months and who were lifetime users of alcohol or some other drug also were skipped out of questions regarding their perceived need for additional treatment. Again, the edited variables corresponding to perceived need for additional services were assigned legitimate skip codes. Those respondents who had not indicated that they received treatment in the past 12 months were asked the general question about whether they perceived themselves as needing treatment for their use of alcohol or other drugs (edited variable NDTXNEDR). If they did not see themselves as needing treatment, they were skipped out of questions pertaining to perceived need for treatment for specific drugs in the past 12 months. Again, legitimate skip codes were assigned to the edited variables that had been skipped.

Similarly, respondents were globally skipped out of questions TX11 through TX22 (regarding their perceived need for any treatment for alcohol or specific other drugs) if they reported in question TX02 that they received treatment in the past 12 months. Therefore, the edited variables corresponding to questions TX11 through TX22 (NDTXALCR through NDTXEFTR) were assigned legitimate skip codes.

Legitimate skip codes also were assigned in situations in which respondents were lifetime nonusers of a particular drug. For example, if respondents indicated that they needed treatment for their use of alcohol or drugs, they were asked about their perceived need for treatment only for those specific drugs that they had ever used; legitimate skip codes were assigned to the skipped drug-specific variables that respondents had never used. Thus, for example, if a respondent had never used heroin but reported needing treatment in the past 12 months for alcohol or drugs (TX08 = 1), a legitimate skip code was assigned to the edited variable pertaining to the perceived need for treatment for heroin (NDTXHERR).

Procedures consistent with those described in Section 2.4.3 also were implemented when questions about the perceived need for treatment were potentially applicable, but respondents refused to report whether they had ever used a particular drug. For example, if a respondent had not received treatment in the past 12 months, reported needing treatment in the past 12 months for alcohol or other drugs, but refused to report whether he or she had ever used heroin, the item about perceived need for treatment for heroin was skipped. Because the respondent refused to report about lifetime use or nonuse of heroin, the edited variable NDTXHERR was assigned a refusal code.

Table B.24 presents additional edits that were specific to the variables pertaining to the perceived need for treatment services. As noted above, for example, respondents were skipped out of questions TX11 through TX22 if they reported that they received treatment in the past 12 months. If respondents had not originally reported receiving treatment in the past 12 months but were logically inferred to have done so (see Table B.22), these respondents would have been routed to questions TX11 through TX22. Rather than replace respondents' answers with codes for "bad data," however, special codes were assigned to indicate that respondents were routed into questions about their perceived need for treatment for use of specific drugs when they were logically inferred to have received treatment in the past 12 months. This procedure would allow analysts to decide whether to use or disregard these data in their analyses.

7.4.9 Health Care Module

The health care module in the 2014 NSDUH covered a wide variety of health-related topics:

- pregnancy status for females aged 12 to 44;
- utilization of medical services in the past 12 months;
- lifetime and past year histories of specific health conditions;
- height and weight; and
- screening for substance use or talks with a health care professional about substance use in the context of health care visits in the past 12 months.

7.4.9.1 Pregnancy Status

Female respondents aged 12 to 44 were asked whether they were currently pregnant and, if so, the number of months that they had been pregnant, as well as lifetime and past year histories of specific health conditions.

An important aspect of processing the pregnancy variables for females aged 12 to 44 involved assignment of legitimate skip codes, where relevant (Section 2.4.2). For example, males of any age and women over the age of 44 were assigned legitimate skip codes to the pregnancy variables. Similarly, if females aged 12 to 44 reported that they were not currently pregnant (PREGNANT = 2), legitimate skip codes were assigned to the variable pertaining to the number of months that they were pregnant (PREGMOS).

In the pregnancy variables, if women reported currently being pregnant, the allowable range for the number of months that they were pregnant ranged from 1 to 9 months. Thus, women who reported that they were currently pregnant were not allowed to report that they had been pregnant for "0" months.

7.4.9.2 Utilization of Medical Services

Prior to 2013, the NSDUH questionnaire included questions about respondents' utilization of hospital emergency room services and overnight inpatient hospitalizations. Questions about respondents' outpatient medical visits in the past 12 months have been included in the questionnaire since the 2013 NSDUH, along with questions about use of other medical services that were described previously in this paragraph.

As for the pregnancy variables that were described previously, an important aspect of processing the health care utilization variables involved assignment of legitimate skip codes where relevant (Section 2.4.2). Specifically, respondents who did not report that they were hospitalized overnight in the past 12 months (edited variable INHOSPYR) were not asked for the number of times they were hospitalized in that period (edited variable NMNGTHSP). If respondents reported that they were not hospitalized overnight in the past 12 months (INHOSPYR = 2), the variable NMNGTHSP was assigned a legitimate skip code. If respondents refused to report whether they were hospitalized overnight in the past 12 months (INHOSPYR = 97), that refusal was propagated onto NMNGTHSP (Section 2.4.3).

In addition, question HLTH16 (edited variable NMVSOPT), regarding the number of times that respondents visited health care professionals in outpatient settings, was a continuous variable. If respondents did not know or refused to report the number of times they visited health care professionals in outpatient settings in the past 12 months, they were asked in a follow-up question (HLTH16DK) to give their "best guess" of the number of outpatient visits that they had in the past 12 months. To encourage respondents to report a number of visits, HLTH16DK was categorical and included ranges for respondents to report their number of visits (e.g., 2–3 times, 4–5 times). HLTH16DK also had response categories of 0 ("I have not visited a health care professional in the past 12 months") and 1 (i.e., 1 time). If NMVSOPT indicated a specific number of outpatient visits in the past 12 months, then the edited variable NMVSOEST corresponding to HLTH16DK was assigned a legitimate skip code.

For the questions about the number of nights that respondents were inpatients in a hospital or the number of outpatient visits in the past 12 months, allowable values that respondents could report ranged from 1 to 366. No editing was done to the variables NMNGTHSP or NMVSOPT when respondents reported that they had spent 365 nights in a hospital in the past 12 months, or that they visited a health care professional in an outpatient setting 365 or 366 times, respectively. In addition, no editing was done if the sum of the number of nights as an inpatient and the number of outpatient visits in the past 12 months was greater than 365. However, if respondents reported spending 366 nights as an inpatient in a hospital in the past 12 months, NMNGTHSP was set to 365. Because NMVSOPT refers to the number of outpatient *visits* rather than the number of days, no editing was done to NMVSOPT if respondents reported 366 outpatient visits in the past 12 months.

7.4.9.3 Lifetime and Past Year Health Conditions

Questions have been included in the health care module since 2005 about the occurrence of the following health conditions in the lifetime and past 12 month periods: anxiety disorder, asthma, bronchitis, cirrhosis of the liver, depression, diabetes, heart disease, hepatitis, high blood pressure, HIV/AIDS (i.e., human immunodeficiency virus or acquired immunodeficiency syndrome), lung cancer, pancreatitis, sexually transmitted disease (STD, such as chlamydia, gonorrhea, herpes, or syphilis), sinusitis, sleep apnea, stroke, tinnitus, tuberculosis, or ulcers. The content of these items and associated edits have not changed.

Respondents were asked whether a doctor or other medical professional had ever told them that they had any of the specific health conditions mentioned above. Respondents could report that they had been told that they had as many of these conditions as applied (i.e., this was an "enter all that apply" question). Respondents also could report that they never had any of these conditions. However, if they reported having one of the conditions listed above and also that they never had any of these conditions, the CAI program triggered an error message that required the respondents to resolve the inconsistency before they could proceed further.

If respondents reported that they had been told that they ever had some of these specific conditions, they were asked whether a doctor or other medical professional told them they had these specific conditions in the past 12 months. The CAI logic restricted respondents' choices in the past 12 months to those conditions that they reported for the lifetime period. For example, if a respondent reported ever being told by a doctor or other health professional that he or she had asthma and bronchitis, but the respondent did not indicate being told that he or she had any of the other health conditions, the respondent's choices for the past 12 month period were limited to reporting whether a doctor or health professional told the respondent that he or she had asthma, bronchitis, or none of these conditions; if the respondent attempted to choose another response for a condition in the past 12 months (e.g., diabetes), the CAI program triggered an error message that this was not one of the respondent's choices. Similarly, respondents were not allowed to report that they had one or more health conditions in the past 12 months and that they had "none of the above" (i.e., none of these conditions in the past 12 months).

Because these were "enter all that apply" variables, separate variables were created for each health condition for the lifetime and past year periods (e.g., LIFANXD and LIFASMA for lifetime occurrence of anxiety disorders or asthma, respectively; YRANXD and YRASMA for the occurrence of these respective conditions in the past year). The individual edited variables for these lifetime and past year health conditions were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (Section 2.4.4).

The CAI logic discussed above eliminated the occurrence of the inconsistent data patterns noted above for these health condition variables, namely reports of having specific health conditions and having none of them, or respondents not reporting these conditions for the lifetime period but reporting them for the past 12 months. Consequently, the editing procedures for these health condition variables involved assignment of legitimate skip codes based on the skip/routing logic. These edits are discussed below.

- If respondents reported that they never had any of these conditions in the lifetime period (i.e., edited variable LIFNONE coded as 95), all of the variables pertaining to lifetime medical conditions (LIFANXD through LIFULCER) were assigned legitimate skip codes. In addition, all of the past year variables (YRANXD through YRULCER and also YRNONE) were assigned legitimate skip codes.

- If respondents reported the lifetime occurrence of at least one of these conditions, LIFNONE was assigned a legitimate skip code.
- If respondents reported the lifetime occurrence of at least one of these conditions but that a doctor or other health professional did not tell them that they had any of these conditions in the past 12 months (i.e., edited variable YRNONE coded as 95), all of the variables pertaining to past year medical conditions (YRANXD through YRULCER) were assigned legitimate skip codes.
- If respondents reported the lifetime occurrence of at least one of these conditions and they did not report the lifetime occurrence of specific other conditions, the corresponding past year variables for the conditions they did not report were assigned legitimate skip codes. For example, if respondents reported that a doctor or other health professional had ever told the respondents that they had diabetes (LIFDIAB = 1) but they did not report ever being told that they had high blood pressure (LIFHBP = 6), the past year high blood pressure variable YRHBP was assigned a legitimate skip code.
- If respondents reported the lifetime occurrence of a particular condition but they did not report that a doctor or other health professional told them that they had this condition in the past year, the edited variable for the condition in the past year was coded as 6 (Response not entered). Suppose, for example, that a respondent reported the lifetime occurrence of bronchitis and high blood pressure (LIFBRONC = 1 and LIFHBP = 1, respectively), and the respondent reported being told in the past year that he or she had high blood pressure but the respondent did not report being told in the past year that he or she had bronchitis, then YRBRONC was coded as 6 and LIFHBP was coded as 1 (Response entered).
- If respondents reported that a doctor or other health professional told them in the past year that they had one or more specific health conditions, YRNONE was assigned a legitimate skip code.

Because these health condition questions were "enter all that apply" items, if respondents indicated that they did not know or refused to report whether they had any of these conditions in the lifetime or past year periods, codes of 94 (for "don't know") or 97 (for "refused") were propagated to all of the unedited variables corresponding to these health conditions in the relevant time period, including the unedited variables indicating that respondents had none of these conditions. Therefore, if respondents reported at least one lifetime health condition but they answered the past year question as "don't know" or "refused," the corresponding codes of 94 or 97 that were assigned to the past year health conditions that the respondents did not report having in their lifetime were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned). In this situation, YRNONE retained the code of 94 or 97 because respondents may not have been told in the past year that they had any of the conditions that they reported for the lifetime period. These edits preserved those responses of "don't know" or "refused" in the past year variables that correspond to conditions that respondents reported that they had in their lifetime.

For example, suppose a respondent reported being told that he or she had diabetes in his or her lifetime but did not choose the lifetime item for lung cancer (i.e., LIFDIAB = 1 but LIFLUNCA = 6). If the respondent answered the past year question as "don't know," the edited past year variable for diabetes (YRDIAB) retained a code of 94 but the past year variable for lung cancer (YRLUNCA) was assigned a code of 89. In addition, YRNONE retained a code of 94.

In addition, if respondents answered the question for the lifetime list of health conditions as "don't know" or "refused," the corresponding code of 94 or 97 was propagated to LIFNONE as well. The item also was skipped pertaining to the occurrence of these conditions in the past year. In this situation, the code of 94 or 97 was retained in LIFNONE because respondents may never have been told by a health professional that they had any of these conditions. The relevant code of 94 or 97 also was propagated to the individual past year variables. For example, if respondents did not know whether a health professional had ever told them that they had any of these conditions, a code of 94 was assigned to the edited past year variables YRANXD through YRULCER and to YRNONE. That is, if these respondents did not know whether they had ever been told that they had any of these conditions, it could reasonably be inferred that the respondents did not know whether they had any of them in the past year. This edit served to reduce the number of codes of "blank" in the corresponding past year variables.

No editing was done in situations in which respondents reported that they had been told at some point in their lifetime that they had certain long-term chronic medical conditions (e.g., cirrhosis of the liver, HIV/AIDS) but did not report these conditions for the past year period. The rationale for not editing the data in this situation was that the past year question asked respondents to indicate which of these conditions a doctor or medical professional told them that they had in the past 12 months. Consequently, respondents may have had these chronic conditions in the past 12 months, but a doctor or other health professional literally may not have told them in the past 12 months that they had these conditions. Nevertheless, analysts would have the option of deciding how to handle these types of special situations.

7.4.9.4 Height and Weight

The 2014 NSDUH included questions for respondents to report their height and weight. Respondents were given four options for reporting their height: (1) in feet and inches, (2) in meters and centimeters, (3) in inches only, or (4) in centimeters only. The edited variable HTANSWER was a two-level variable that indicated whether respondents preferred to report their height in (1) English units (feet and inches or inches only) or (2) metric units (meters and centimeters or centimeters only).

Respondents could choose to report their weight in pounds or in kilograms. If female respondents aged 12 to 44 reported that they were currently pregnant (Section 7.4.9.1), they were asked to report their weight before they became pregnant. As for HTANSWER, the edited variable WTANSWER was a two-level variable that indicated whether respondents preferred to report their height in (1) English units (pounds) or (2) metric units (kilograms).

A single variable for height in inches (HTINCHES) was created from respondents' reports of their height. If respondents chose to report their height in feet and inches, their reports were converted to inches (e.g., 5 feet, 10 inches = 70 inches). If respondents chose to report their height in meters and centimeters, their answers first were converted to centimeters (e.g., 1 meter, 50 centimeters = 150 centimeters). Reported heights in centimeters (including heights in meters and centimeters that were converted to centimeters) were converted to heights in inches using the conversion factor of 1 centimeter = 0.393700787 inches.

In 2014, the allowable ranges for reporting heights in inches (questions HLTH08A or HLTH08B) were revised according to whether respondents preferred to report their height in feet and inches or only in inches. Similarly, the allowable ranges for reporting heights in centimeters (questions HLTH10A or HLTH10B) were revised according to whether respondents preferred to report their height in meters and centimeters or only in centimeters. The allowable ranges for reporting weights in pounds (questions HLTH12 and HLTH14) were changed from 50 to 550 pounds in 2013 to 40 to 999 pounds in 2014. Similarly, the allowable ranges for reporting weights in kilograms (questions HLTH13 and HLTH15) were changed from 22 to 275 kilograms in 2013 to 18.00 to 999.00 kilograms in 2014. As a consequence, codes for missing data for respondents' final weights were four digits in length.

HTINCHES could have missing values based on the types of responses that are described below.

- If respondents had missing data for their preferred way of reporting their height (edited variable HTANSWER), remaining questions for height were skipped. The corresponding code for missing data was assigned to HTINCHES (Section 2.2.1). For example, if HTANSWER = 94 (don't know), then HTINCHES was assigned a code of 994 because heights of 90 or more inches were allowed in the data.
- If respondents preferred to report their height in feet and inches or in meters and centimeters, it was possible for them to report the major unit for their height (e.g., feet) but to have a missing value for the minor unit (e.g., inches). The corresponding code for missing data was assigned to HTINCHES. For example, if a respondent reported being 5 feet tall but did not know the number of additional inches beyond 5 feet for his or her height, then HTINCHES was assigned a code of 994.

A single variable for weight in pounds (WTPOUNDS) was created from respondents' reports of their weight (or the weight of pregnant females before they became pregnant). Reported weights in kilograms were converted to weights in pounds using the conversion factor of 1 kilogram = 2.20462262 pounds.

As for HTINCHES, relevant codes for missing data were assigned to WTPOUNDS according to the type of nonresponse. For example, if WTANSWER had a code of 94 (don't know), then WTPOUNDS was assigned a code of 9994. As noted previously, codes for missing data in WTPOUNDS were four digits in length for 2014 to account for the changes to the allowable ranges for weight in the 2014 survey.

7.4.9.5 Screening for Substance Use or Talks about Substance Use with a Health Care Professional

In addition to new questions being included for outpatient health care visits in the past 12 months, questions have been included in this module since the 2013 NSDUH regarding whether respondents who had health care visits in the past 12 months were asked by a doctor or other health care professional about their tobacco, alcohol, or other substance use in the context of these medical visits (i.e., screening for substance use), regardless of whether they reported use of tobacco, alcohol, or illicit drugs. Respondents also were asked new questions about talks that they had with a doctor or other health care professional about their use of tobacco, alcohol, or illicit drugs if they reported health care visits in the past 12 months and they reported use of tobacco products, use of alcohol, or use of specific illicit drugs (i.e., marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, or methamphetamine) (for HLTH20). Table 7.4 lists these new questions, the names for the corresponding edited variables, and descriptions for these questions.

Respondents were defined as having had health care visits in the past 12 months if any of the following occurred:

- the respondent reported in question HLTH03 (edited variable NMERTMT) that he or she had been treated in a hospital emergency room one or more times in the past 12 months;
- the respondent answered question HLTH04 (edited variable INHOSPYR) as "yes," indicating that the respondent stayed overnight or longer as an inpatient in a hospital in the past 12 months; or
- the respondent reported in question HLTH16 (edited variable NMVSOPT) or in question HLTH16DK (edited variable NMVSOEST) that he or she had one or more visits to a health care professional in an outpatient setting in the past 12 months.

Thus, an important aspect of processing the variables pertaining to screening for substance use and talks about substance use with health care professionals in health care settings involved assignment of legitimate skip codes where relevant (Section 2.4.2). Further details are provided elsewhere in this section for assignment of legitimate skip codes, depending on the questions.

Table 7.4 Mapping of Edited Variables for Health Care Screening or Talks about Substance Use with a Health Care Professional

Question	Edited Variable	Description
HLTH17A	HPUSETOB	Doctor asked in person or on a form in the past 12 months if the respondent smokes cigarettes or uses any other tobacco products.
HLTH17B	HPUSEALC	Doctor asked in person or on a form in the past 12 months if the respondent drinks alcohol.
HLTH17C	HPUSEDRG	Doctor asked in person or on a form in the past 12 months if the respondent uses illegal drugs.
HLTH18	HPQTTOB	Doctor advised the respondent in the past 12 months to quit smoking cigarettes or quit using other tobacco products.
HLTH19, response option 1	HPALCAMT	Doctor asked the respondent in the past 12 months how much the respondent drank alcohol.
HLTH19, response option 2	HPALCFRQ	Doctor asked the respondent in the past 12 months how often the respondent drank alcohol.
HLTH19, response option 3	HPALCPRB	Doctor asked the respondent in the past 12 months whether the respondent had any problems because of his or her drinking.

Table 7.4 Mapping of Edited Variables for Health Care Screening or Talks about Substance Use with a Health Care Professional (continued)

Question	Edited Variable	Description
HLTH19, response option 4	HPALCCUT	Doctor advised the respondent in the past 12 months to cut down on his or her drinking.
HLTH19, response option 5	HPALCTX	Doctor offered to give the respondent more information about alcohol use and treatment for alcohol use problems in the past 12 months.
HLTH19, response option 95	HPALCNOT	Doctor did not discuss the respondent's alcohol use with the respondent in the past 12 months.
HLTH20	HPDRGTLK	Doctor discussed use of specific illicit drugs with the respondent in the past 12 months.

NOTE: HLTH19 is an "enter all that apply" question. Therefore, separate variables were created for each response option. Illicit drugs for HLTH20 referred to marijuana, cocaine, crack, heroin, hallucinogens, inhalants, or methamphetamine.

7.4.9.5.1 Screening for Substance Use

Eligibility for being asked questions HLTH17A, HLTH17B, and HLTH17C was based only on reports of health care visits in the past 12 months, regardless of whether respondents used tobacco, alcohol, or other substances. HPUSETOB, HPUSEALC, and HPUSEDRG were assigned a code of 99 (Legitimate skip) if respondents reported (1) no hospital emergency room visits in the past 12 months (i.e., NMERTMT = 0), (2) no overnight hospitalizations in the past 12 months (i.e., INHOSPYR = 2), and (3) no outpatient visits in the past 12 months (i.e., NMVSOPT = 0 or NMVSOEST = 0). HPUSETOB, HPUSEALC, and HPUSEDRG retained missing values if these questions had been skipped because respondents did not report having any health care visits in the past 12 months but (1) NMERTMT or INHOSPYR had missing values (i.e., "don't know" or "refused") or (2) both NMVSOPT and NMVSOEST had missing values.

7.4.9.5.2 Talks about Tobacco Use

Since 2014, respondents who had health care visits in the past 12 months and reported *past year* use of cigarettes, snuff, chewing tobacco, or cigars, or *past month* use of pipe tobacco (i.e., there is no question in NSDUH to identify respondents who used pipe tobacco more than 30 days ago but within the past 12 months) were eligible to be asked question HLTH18 (HPQTTOB), regarding whether a doctor or other health care professional advised these respondents to quit smoking cigarettes or to stop using other tobacco products.[125] Data in HPQTTOB for 2014 were edited according to this skip logic as described below.

- If question HLTH18 had been skipped and respondents *never* used cigarettes, snuff, chewing tobacco, cigars, or pipe tobacco, then HPQTTOB was assigned a code of 91, where 91 = NEVER USED TOBACCO PRODUCTS. This code was assigned

[125] In 2013, respondents who had health care visits in the past 12 months and reported *lifetime* use of tobacco products were eligible to be asked HTLH18. For documentation of the procedures for editing HPQTTOB based on the logic in the 2013 NSDUH, see the editing and imputation report in the 2013 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2015b).

regardless of whether respondents reported any health care visits in the past 12 months.

- If question HLTH18 had been skipped because respondents did not report health care visits in the past 12 months and the most recent use of any tobacco products was more than 12 months ago (which could include reports that respondents never used some types of tobacco products),[126] then HPQTTOB was assigned a code of 93, where 93 = DID NOT USE TOBACCO IN THE PAST 12 MONTHS.
- If question HLTH18 had been skipped because respondents had no health care visits in the past 12 months but they used tobacco products in the past 12 months,[127] then HPQTTOB was assigned a code of 99, where 99 = LEGITIMATE SKIP. Thus, in the editing procedures, precedence was given to assigning codes of 91 or 93 to HPQTTOB ahead of assigning a code of 99.

HPQTTOB retained missing values if HLTH18 had been skipped because of the following:

- respondents reported health care visits in the past 12 months but it was unknown whether respondents were users of tobacco products in the past year; or
- respondents reported past year use of one or more tobacco products but it was unknown whether respondents had any health care visits in the past 12 months.

7.4.9.5.3 Talks about Alcohol Use

For question HLTH19, respondents who had health care visits in the past 12 months and reported past year use of alcohol were eligible to be asked the question. Because HLTH19 was an "enter all that apply" question, separate variables were created that corresponded to the response options in HLTH19. When HLTH19 was skipped, variables corresponding to HLTH19 were edited as described below.

- If question HLTH19 had been skipped and respondents *never* used alcohol, then the variables corresponding to HLTH19 were assigned a code of 91, where 91 = NEVER USED ALCOHOL. This code was assigned regardless of whether respondents reported any health care visits in the past 12 months.
- If question HLTH19 had been skipped because respondents last used alcohol more than 12 months prior to the interview date, then the edited variables corresponding to HLTH19 were assigned a code of 93, where 93 = DID NOT USE ALCOHOL IN THE PAST 12 MONTHS. Again, this code was assigned regardless of whether respondents reported any health care visits in the past 12 months.
- If question HLTH19 had been skipped because respondents had no health care visits in the past 12 months but they used alcohol in the past 12 months, then the edited variables corresponding to HLTH19 were assigned a code of 99, where 99 =

[126] For brevity, this includes reports that respondents did not use pipe tobacco in the past 30 days and there were no reports of past year use of other tobacco products.

[127] For brevity, this includes reports that respondents used pipe tobacco in the past 30 days.

LEGITIMATE SKIP. Thus, in the editing procedures for HLTH19, precedence was given to assigning codes of 91 or 93 ahead of assigning a code of 99.

The edited variables for HLTH19 retained missing values if HLTH19 had been skipped because of the following:

- respondents reported health care visits in the past 12 months but it was unknown whether respondents were past year alcohol users; or
- respondents reported past year alcohol use but it was unknown whether respondents had any health care visits in the past 12 months.

In addition, the last category in question HLTH19 allowed respondents to report that a doctor or other health professional did not discuss the respondents' alcohol use with them during any health care visits that they had in the past 12 months (HLTH19 = 95). The CAI logic did not allow respondents to choose this last response if they chose any of the preceding five response options for discussions about alcohol use, or vice versa (Table 7.4). If this last response option was chosen in HLTH19, then the edited variable HPALCNOT corresponding to this response option retained a code of 95. The edited variables HPALCAMT through HPALCTX were assigned a code of 99 (LEGITIMATE SKIP). If respondents chose one or more of the first five response options in HLTH19, then HPALCNOT was assigned a code of 99.

The edited variables for question HLTH19 were examples of variables with the new codes for "enter all that apply" questions that were described in Section 2.4.4 in Chapter 2. If respondents chose any of the responses for categories 1 through 5 from HLTH19, then these responses that were chosen were assigned a code of 1 (Yes). Other responses from the list that were not chosen were assigned a code of 2 (No [not entered]).

7.4.9.5.4 Talks about Illicit Drug Use

Respondents were eligible to be asked question HLTH20 (HPDRGTLK) if they had health care visits in the past 12 months and they reported past year use of any of the following: marijuana, cocaine (including crack), heroin, hallucinogens, inhalants, or methamphetamine. For cocaine and heroin, lifetime users who did not report past year use in the corresponding core modules were eligible to be asked HLTH20 if they reported past year use of these substances in the noncore special drugs module (e.g., use of cocaine with a needle in the past 12 months) (Section 7.4.1.2). Eligibility for being asked HLTH20 for past year users of methamphetamine was based on reports of past year use from the core stimulants module or the noncore special drugs module. When HLTH20 was skipped, HPDRKTLK was edited as described below.

- If question HLTH20 had been skipped and respondents *never* used any of the illicit drugs that were described in the preceding paragraph, then HPDRGTLK was assigned a code of 91, where 91 = NEVER USED MJ/COC/HER/HALL/INH/METH. This code was assigned regardless of whether respondents reported any health care visits in the past 12 months.
- If question HLTH20 had been skipped because the most recent use of any of these illicit drugs was more than 12 months ago (which could include reports that

respondents never used some of these types of illicit drugs), then HPDRGTLK was assigned a code of 93, where 93 = DID NOT USE MJ/COC/HER/HAL/INH/METH IN THE PAST 12 MONTHS.

- If question HLTH20 had been skipped because respondents had no health care visits in the past 12 months but they used any of these illicit drugs in the past 12 months, then HPDRGTLK was assigned a code of 99, where 99 = LEGITIMATE SKIP. Thus, in the editing procedures for HPDRGTLK, precedence was given to assigning codes of 91 or 93 ahead of assigning a code of 99.

HPDRGTLK retained missing values if HLTH20 had been skipped because of the following:

- respondents reported health care visits in the past 12 months but it was unknown whether respondents were users of any of these illicit drugs in the past year; or
- respondents reported past year use of one or more illicit drugs but it was unknown whether respondents had any health care visits in the past 12 months.

In addition, the CAI instrument defined most recent use of methamphetamine in the special drugs module following question SD17B (see Section 6.2.6 and Table 6.4). However, this logic in the CAI instrument occurred before the consistency checks SD17A1 and SD17ALT; these consistency checks were triggered when respondents previously reported that they never used methamphetamine in the core stimulants module but they reported methamphetamine use in the special drugs module. However, the editing of the core-plus-noncore (CPN) methamphetamine recency took into account data from SD17A1 and SD17ALT, as described in Section 6.2.6. Consequently, a small number of respondents in 2014 (fewer than 15) were routed to question HLTH20 because their *only* reported past year use of illicit drugs came from question SD17B but they subsequently reported in SD17A1 or SD17ALT that they never used methamphetamine (i.e., and their edited data for most recent methamphetamine use indicated that they never used it).

The edits described below were implemented for HPDRGTLK when this pattern occurred. For these edits, the respondents were defined as having never used methamphetamine based on their final edited CPN recency for methamphetamine.

- If respondents (1) had never used any of the illicit drugs that were relevant to HLTH20 (including nonuse of methamphetamine) and (2) answered question HLTH20 as "no," then HPDRGTLK was assigned a code of 81 (i.e., logical inference that these respondents never used the relevant illicit drugs). The assumption was that the answer of "no" in HLTH20 could mean that these respondents were not asked about their use of these drugs because they never used marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine.
- If respondents had never used methamphetamine, but they last used some other illicit drug more than 12 months ago, and they answered HLTH20 as "no," then HPDRGTLK was assigned a code of 83 (i.e., logical inference that the respondent was not a past year user of the relevant illicit drugs). The assumption was that the answer of "no" in HLTH20 could mean that these respondents were not asked about

their use of these drugs because they were lifetime (but not past year) users of marijuana, cocaine, heroin, hallucinogens, or inhalants.

- If respondents had never used any of the illicit drugs that were relevant to HLTH20 (including nonuse of methamphetamine) but answered question HLTH20 as "yes," then HPDRGTLK was set to a value of 11 (i.e., "yes" with a "bump by 10"). This edit preserved the information that these respondents answered the question as "yes" but also was intended to indicate to analysts that there was some question about whether these respondents should have been asked the question because there was no indication that they ever used any of the relevant illicit drugs.
- If respondents had never used methamphetamine, but they last used some other illicit drug more than 12 months ago, and they answered HLTH20 as "yes," then HPDRGTLK was set to a value of 21 (i.e., "yes" with a "bump by 20"). This edit preserved the information that these respondents answered the question as "yes" but also was intended to indicate to analysts that there was some question about whether the respondents should have been asked the question because there was no indication that they used the relevant illicit drugs in the past year.

7.4.10 Adult Mental Health Service Utilization Module

The module on adult mental health service utilization asked adult respondents about (1) their receipt of specific sources of inpatient or outpatient mental health services in the past 12 months, (2) the length of time that respondents spent in specific inpatient mental health settings or the number of outpatient visits that respondents made to specific types of outpatient mental health providers, (3) payment sources for mental health services, (4) use of prescribed medication for a mental health condition, (5) unmet demand for services (i.e., the respondent felt the need for mental health services but did not receive them), (6) use of alternative sources of treatment (e.g., acupuncture), and (7) how respondents were prompted to get treatment. If the lifetime substance treatment question TX01 indicated that respondents had received treatment for their use of alcohol or other drugs (or if they answered the question as "don't know" or "refused"), respondents were instructed not to include substance abuse treatment in their answers about mental health treatment.

Sources of inpatient mental health treatment or counseling that were asked about in the module included (1) a private or public psychiatric hospital, (2) a psychiatric unit within a general hospital, (3) a medical unit within a general hospital, (4) another type of hospital, (5) a residential treatment center, or (6) "some other type of facility." Sources of outpatient mental health treatment or counseling that were asked about in the module included (1) an outpatient mental health clinic or center, (2) the office of a private therapist not associated with a clinic, (3) a doctor's office that was not part of a clinic, (4) an outpatient medical clinic, (5) a partial day hospital or day treatment program, or (6) "some other place." Sources of alternative treatment that were asked about in the module included treatment from (1) an acupuncturist or acupressurist; (2) a chiropractor; (3) an herbalist; (4) an in-person support group or self-help group; (5) an Internet support group or chat room; (6) a spiritual or religious advisor, such as (but not limited to) a pastor, priest, or rabbi; (7) a telephone hotline; (8) a massage therapist; or (9) "other" (i.e., some other source).

An important aspect of processing the variables in this section involved assignment of legitimate skip codes, where relevant (Section 2.4.2). That included (1) assignment of legitimate skip codes to variables in the entire module for respondents who were aged 12 to 17 and (2) assignment of legitimate skip codes to adult respondents' data based on routing logic within the adult mental health service utilization module. For example, if respondents reported that they did not stay overnight or longer in a hospital or other facility to receive mental health counseling in the past 12 months (AUINPYR = 2), all subsequent variables pertaining to inpatient mental health services were assigned legitimate skip codes.

In addition, if respondents did not report receiving treatment in a particular facility or setting in the past 12 months, the questions pertaining to the number of times they received treatment in that setting were skipped. For example, if respondents reported receiving outpatient mental health services in the past 12 months (AUOPTYR = 1) but did not indicate that they received outpatient services in a day treatment program, the edited variable pertaining to receipt of day treatment services (AUOPDTMT) was assigned a legitimate skip code. If respondents in 2014 reported receiving outpatient services in one or more locations from the lists they were provided but they did not report receiving services in "some other place," the edited "OTHER, Specify" variable AUOPYRSP for outpatient services in some other place was assigned a legitimate skip codes.

Similarly, if respondents reported only one source of payment for inpatient or outpatient mental health services, there was no need to ask them who paid for (or would pay for) most of the inpatient or outpatient services that they received. For example, if respondents reported that they received outpatient mental health services in the past 12 months but reported only that private insurance paid for their outpatient mental health services, the edited variable pertaining to the principal payment source (AUPOPMOS) was assigned a legitimate skip code.

In questions pertaining to the specific places where respondents received inpatient or outpatient mental health services in the past 12 months, they were allowed to enter more than one place from the list where they received services. Similarly, respondents could select more than one response from lists of payment sources for their inpatient or outpatient services. Information for each of these mental health service locations or payment sources was subsequently captured as a discrete variable. For example, information about receipt of inpatient mental health services in a psychiatric hospital, the psychiatric unit of a general hospital, the medical unit of a general hospital, another type of hospital, a residential treatment center, or some other type of facility was captured in the variables AUINPSYH, AUINPGEN, AUINMEDU, AUINAHSP, AUINRESD, and AUINSFAC, respectively. These individual edited variables corresponding to these "enter all that apply" questions in the adult mental health service utilization module were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (Section 2.4.4).

Codes of 94 and 97 (for "don't know" and "refused," respectively) were assigned to an entire list of variables if respondents did not know or refused to report what specific places they received mental health services or what specific sources paid (or would pay) for their mental health treatment (Section 2.2.2). If an entire list was blank but respondents had previously reported receiving inpatient services (e.g., if respondents broke off the interview), then the lists of variables pertaining to locations for inpatient services or payment for inpatient services retained a code of 98 (i.e., "blank"); similar logic was applied if respondents reported receiving outpatient mental health services but the location or payment variables were entirely blank.

Adult respondents also were asked if there was any time in the past 12 months when they felt the need for mental health treatment but did not get services. Respondents who answered this question as "yes" then were asked to indicate the reason (or reasons) why they did not get treatment. Based on a review of what respondents had specified in quarter 1 of 2002 as leading "other" reasons for not getting mental health treatment, an additional question (ADMT27A) has been included in the module since 2003 to capture commonly endorsed other reasons for not getting treatment. Therefore, if respondents answered question ADMT26 as "yes" (i.e., perceived the need for mental health treatment), they were routed first to question ADMT27, where they could report up to nine reasons for not receiving mental health treatment, including "some other reason or reasons." If respondents chose the "some other reason or reasons" option in question ADMT27, they were routed next to question ADMT27A, where they could choose additional reasons for not getting mental health treatment; again, respondents were given the option in question ADMT27A to report "some other reason or reasons." If respondents chose this "other" response category in ADMT27A, they were asked to specify the *most important* other reason why they did not get treatment. Respondents since 2003 also have been asked to specify only the most important other reason for not getting mental health treatment.

Questions ADMT27 and ADMT27A were "enter all that apply" questions, in which respondents could choose more than one reason from each list. Each response option (e.g., "You couldn't afford the cost") was captured as a separate variable. The edited variables corresponding to the individual response options in ADMT27 and ADMT27A were coded as 1 (Response entered) or 6 (Response not entered), if at least one item was chosen from the lists in ADMT27 or ADMT27A (Section 2.4.4).

In addition, the ninth response option in question ADMT27 ("some other reason or reasons") was considered principally to be a "toggle" to question ADMT27A. Therefore, separate "some other reason or reasons" variables were not created to correspond to the last response category in ADMT27 and the last category in ADMT27A.

The edits described below have been implemented since 2003 to take into account the data from question ADMT27A.

- If respondents chose the ninth response option in ADMT27, any response that was entered from the ADMT27A series was coded as 1, and anything that respondents did not choose from the ADMT27A list was coded as 6.
- If respondents chose a response from ADMT27 but did not choose the ninth response category in ADMT27, the variables corresponding to the response categories in ADMT27A (i.e., AUUNOND through AUUNSOR) all were given a code of 6 (Response not entered), rather than being assigned "legitimate skip" codes. That is, ADMT27 and ADMT27A were considered together to be one big series of reasons.
- If respondents chose category 9 in question ADMT27, chose at least one reason from ADMT27A, but did not choose category 7 in ADMT27A (some other reason or reasons), the edited variable AUUNSOR (some other reason or reasons for not

receiving mental health treatment) was assigned a code of 6 (Response not entered). That is, it was inferred in this situation that the list of specific reasons in ADMT27A was adequate for capturing why respondents did not get mental health treatment. For example, if a respondent chose category 9 in ADMT27 and then chose only category 2 in ADMT27A ("You thought you could handle the problem without treatment"), it would be reasonable to infer that this response in ADMT27A was the only other reason why the respondent did not get treatment.

- If respondents chose response category 9 in question ADMT27, it was possible for them to answer ADMT27A as "don't know" or "refused" (i.e., did not know or refused to report what the other reasons were). When this occurred, the "some other reason" variable AUUNSOR was set to 1 (Response entered) in order to retain information that the respondent chose "some other reason or reasons" somewhere in the series. Remaining variables corresponding to the ADMT27A series retained codes of 94 ("don't know") or 97 ("refused").
- If respondents answered question ADMT27 as "don't know" or "refused," question ADMT27A was skipped. Therefore, the relevant code of 94 or 97 was propagated onto the variables corresponding to the ADMT27A list.
- If AUUNSOR had a value of 6 (see above), the "OTHER, Specify" variable AUUNRIMP was assigned a legitimate skip code. If AUUNSOR had a refusal code, that refusal was propagated onto AUUNRIMP.
- If AUUNSOR had a code of 1 when the respondent answered "don't know" or "refused" to the ADMT27A series, the "OTHER, Specify" variable AUUNRIMP retained a code of 98 (blank).

Consistent with the editing procedures in prior years, if AUUNMTYR indicated that there was not a time in the past 12 months when respondents felt the need for mental health treatment but did not receive services (AUUNMTYR = 2), the edited variables corresponding to questions ADMT27 and ADMT27A were assigned legitimate skip codes. Similarly, if AUUNMTYR was refused, that refusal was propagated onto the skipped variables from questions ADMT27 and ADMT27A.

Also, since 2004, respondents have been shown a list of alternative sources of mental health treatment in question ADMT29A (edited variable AUALTYR) and were asked whether they had received treatment, counseling, or support from other sources such as these in the past 12 months. Respondents who answered ADMT29A as "yes" were asked question ADMT29B, pertaining to specific sources of alternative mental health treatment in the past 12 months. ADMT29B was an "enter all that apply" type of question. Therefore, the individual variables AUALACUP (acupuncturist or acupressurist), AUALCHIR (chiropractor), AUALHERB (herbalist), AUALSGRP (in-person support group or self-help group), AUALINET (Internet support group or chat room), AUALRELG (spiritual or religious advisor, such as a pastor, priest, rabbi), AUALHLIN (telephone hotline), AUALMASG (massage therapist), and AUALOTH (other source) pertaining to the individual sources of alternative treatment in ADMT29B were assigned codes of 1 or 6, as described previously in this section and in Section 2.4.4, when AUALTYR = 1 (i.e., "yes"). When AUALTYR = 2 (i.e., "no"), AUALCHIR through AUALOTH and AUALOTSP (the "OTHER, Specify" variable for other sources of alternative treatment) were assigned legitimate skip codes. When AUALTYR was refused, that refusal was propagated to AUALACUP through AUALOTSP. When AUALOTH was coded as 6 (Response not entered), a legitimate skip code was assigned to AUALOTSP.

Coding of AUALOTSP, regarding other alternative practitioners, was based on information from the National Institutes of Health's (NIH's) National Center for Complementary and Integrative Health (NCCIH) (https://nccih.nih.gov/health/integrative-health).[128] NCCIH groups complementary health approaches into the following categories:

- *Natural Products*: Herbalists fall under this type of therapy, along with use of dietary supplements and probiotics.
- *Mind and Body Practices*: Forms of mind and body practices include the following
  - acupuncture,
  - massage therapy,
  - meditation,
  - movement therapies, such as Pilates and Rolfing Structural Integration,
  - relaxation techniques, such as breathing exercises and guided imagery,
  - spinal manipulation practiced by health care professionals such as chiropractors and osteopaths,
  - tai chi or qi gong practices of movement from traditional Chinese medicine,
  - yoga, and
  - other mind and body practices, such as healing touch and hypnotherapy.
- *Other Complementary Health Approaches*: These include traditional healers, Ayurvedic medicine from India, traditional Chinese medicine, homeopathy, and naturopathy.

In addition, the "OTHER, Specify" variable AUOPYRSP (other source of outpatient mental health treatment or counseling) had previously included a code 11 for support groups, self-help groups, or group counseling. However, AUALSGRP pertained only to support groups or self-help groups that were not commonly part of treatment or counseling from a mental health professional. In contrast, group counseling likely would be administered from a mental health professional. For this reason, respondents who reported group counseling were assigned to a category in AUOPYRSP that has been present since 2003 (43 = Group counseling, self-help not specified). Only those respondents who reported receiving treatment from support groups or self-help groups retained a code of 11 for AUOPYRSP. Therefore, documentation for code 11 in AUOPYRSP has read as "Support group/self-help group" since 2004. Respondents who reported in AUOPYRSP that they had received treatment or counseling from a support group or self-help

[128] As defined by NCCIH, if a non-mainstream practice is used together with conventional medicine, it is considered "complementary." If a non-mainstream practice is used in place of conventional medicine, it is considered "alternative," although true alternative medicine is rare (i.e., without the use of any conventional medical treatment). "Integrative" health care involves bringing conventional and complementary approaches together in a coordinated way.

group were logically inferred in AUALSGRP to have received treatment from this source, if AUALSGRP had not already been coded as 1; this issue is described further in Table B.25 in Appendix B.

In subsequent analyses of the adult mental health service utilization data, respondents were not classified as having received outpatient mental health treatment if the only "outpatient" location that they reported was a support group or self-help group. In contrast, group counseling was considered a valid other form of outpatient treatment. Therefore, to facilitate analysis of trends in adults' receipt of outpatient mental health treatment in the past 12 months, the variable AUOPYRSP also was revised in 2003 to reclassify respondents into category 43 if they reported group counseling and to retain a code of 11 in AUOPYRSP only for those respondents who reported receiving services from a support group or a self-help group.

Table B.25 discusses additional edits that were relevant to the adult mental health service utilization variables. For example, respondents could report receipt of outpatient mental health services in "some other place" and then specify a location (e.g., a private therapist's office) that they had not already chosen as a place where they received services. In these situations, respondents were logically inferred to have received services at that location (Section 2.4.5). For example, if respondents had not already indicated that they received outpatient mental health treatment in the office of a private therapist, the edited variable AUOPTHER was assigned a code of 3 (Response entered LOGICALLY ASSIGNED).

7.4.11 Social Environment Module

As noted in Section 7.1, the social environment module was administered only to adults. This section included questions about respondents' changes of residence in the past 5 years, involvement in criminal or potentially criminal activities, attitudes about adults trying marijuana once or twice, and religious involvement. A shorter version of this module has been administered since 2005 with the deletion of items related to neighborhood cohesiveness. Consequently, the name of this module has been changed from "social and neighborhood environment" prior to 2005 to "social environment" since 2005.

As was the case in prior years, minimal processing of data was done to variables in this section. The primary data processing involved assignment of legitimate skip codes for respondents who were aged 12 to 17 (Section 2.4.2). Adults were asked all questions within the social environment module. As noted in Section 7.2, responses were replaced with codes for bad data if adult respondents keyed "1" (or multiple-digit responses of "1" if they occurred) to all of the questions that were asked in the module. Although a code of "1" would be plausible for all variables in this module, responses were set to bad data if the pattern of keying "1" wherever possible continued from prior modules.

7.4.12 Parenting Experiences Module

The parenting experiences module was intended to be administered only in dwelling units (DUs) where (1) two people had been selected for an interview, (2) a 12- to 17-year-old had been selected for an interview (regardless of whether the youth completed the interview), and (3) the respondent being interviewed was the parent or legal guardian of the 12- to 17-year-old who also was selected for an interview. Editing of the parenting experiences data first involved editing the field interviewer (FI) checkpoint variables (FIPE1, FIPE2, and FIPE3) completed by the interviewers toward the beginning of the interview.[129] The variables in the parenting experiences module then were edited based on the final values assigned to the edited FIPE variables.

The content of this module did not change in 2014. The ACASI "lockout" feature that was described in Section 7.3 did not affect how the FI checkpoint data were edited but did affect the editing of the parenting experiences data. Issues associated with this ACASI "lockout" are described in Section 7.4.12.2.

7.4.12.1 Editing of the Field Interviewer Checkpoint Variables

Interviewers were instructed to enter into these checkpoints the relevant information described above for determining whether respondents were eligible to be administered the parenting experiences questions. These checkpoint variables were edited for consistency with the pair-selection and pair-respondent sample variables (PAIRSEL and PAIRRESP, respectively). These checkpoints were interviewer-administered and not self-administered. Editing of these checkpoints was related to the edits for the parenting experiences questions (which were self-administered), however, because the final values in the edited checkpoints were critical for determining whether respondents were in fact eligible to be asked the parenting experiences questions.

Editing of the FIPE1 Checkpoint (and Related Edits). First, the FIPE1 variable was edited for consistency with the pair-selection variable PAIRSEL. Specifically, this checkpoint pertained to whether two people were selected for an interview at that DU. There were no situations in 2014 when two people were interviewed at a given DU without two people having first been selected. Therefore, editing FIPE1 involved reviewing only information on the number of people selected for an interview at that DU based on PAIRSEL.

If the pair-selection data indicated that two people were selected from that DU, then FIPE1 should have been answered as "yes." Therefore, if the pair-selection data indicated that two people were selected and FIPE1 was not answered as "yes," a code of 3 (i.e., Yes LOGICALLY ASSIGNED) was assigned to the edited FIPE1 variable (SKPX2PER). Similarly, if the pair-selection data indicated that only one person was selected from that DU, then FIPE1 should have been answered as "no." Therefore, if the pair-selection data indicated that only one person was selected and FIPE1 was not answered as "no," the editing procedures logically inferred that "no" should have been the answer. If the edited version of FIPE1 indicated that two people were not selected for an interview, then the edited versions of FIPE2 (SKPX1217) and FIPE3 (SKPXPRNT) were assigned legitimate skip codes. If data existed in FIPE2 or FIPE3 when the edited SKPX2PER was inferred to be answered as "no," SKPX1217 and SKPXPRNT were assigned a code of 89 (i.e., LEGITIMATE SKIP Logically assigned) to signify that these two checkpoints should have been skipped (Section 2.4.2).

[129] "Checkpoint" refers to an item completed by the interviewer about the location of the sampled dwelling unit (SDU) or characteristics of the sample within the SDU; these checkpoints are not seen by the respondent and are used to determine what the respondent is asked in subsequent sections of the interview.

Editing of the FIPE2 Checkpoint (and Related Edits). Next, FIPE2 was edited for consistency with PAIRSEL, PAIRRESP, and the age of the respondent. Specifically, this checkpoint pertained to whether a 12- to 17-year-old was selected for an interview at that DU, *regardless of whether the selected youth actually responded*. Edits of the FIPE2 checkpoint data involved review of both the pair-selection data (PAIRSEL) and the pair-respondent data (PAIRRESP) in case either indicated that a 12- to 17-year-old was selected or interviewed.

The age of the respondent was taken into account because interviewers were skipped past this checkpoint if respondents were aged 12 to 17. Therefore, the edited version of FIPE2 (SKPX1217) was assigned legitimate skip codes (i.e., 99 if FIPE2 was blank and 89 if FIPE2 was not blank) when the respondent was a youth.

The remaining edits for FIPE2 were implemented when the respondent was an adult. If both PAIRSEL and PAIRRESP indicated that a 12- to 17-year-old was neither selected nor interviewed, it reasonably could be inferred that FIPE2 should have been answered as "no." If FIPE2 was not already answered as "no," the edits assigned a code to SKPX1217 to indicate that a response of "no" was logically inferred. This included situations in which the pair-selection data indicated that a 12- to 17-year-old was not selected, and a completed interview was obtained from only one respondent, who was not aged 12 to 17, regardless of whether PAIRSEL and PAIRRESP were totally consistent. For example, if the pair-selection data indicated that an 18- to 25-year-old and a 26- to 34-year-old were selected, but a single interview was obtained from a 35- to 49-year-old, the pair-selection and pair-respondent data were not totally consistent, but neither would suggest that a 12- to 17-year-old should have been selected. When the edited SKPX1217 indicated that a 12- to 17-year-old was not selected, including situations described above in which the edits inferred that no 12- to 17-year-old was selected, then legitimate skip codes were assigned to the edited variable SKPXPRNT corresponding to FIPE3 (code of 99 if FIPE3 was blank; or 89 if it was not blank).

If either PAIRSEL or PAIRRESP indicated that a 12- to 17-year-old was selected or interviewed, it could be inferred that FIPE2 should have been answered as "yes." Therefore, if FIPE2 was not already answered as "yes," a special code was assigned to SKPX1217 to indicate that a response of "yes" was logically inferred. This included the following situations: (1) PAIRSEL indicated that a 12- to 17-year-old was selected, and PAIRRESP indicated that an interview was obtained from a 12- to 17-year-old, regardless of whether PAIRSEL and PAIRRESP matched exactly (e.g., a 12- to 17-year-old and a 26- to 34-year-old were selected, but interviews were obtained from a 12- to 17-year-old and a 35- to 49-year-old); and (2) PAIRSEL indicated that a 12- to 17-year-old was selected, but a single interview from an adult was obtained at the DU, regardless of whether the adult category from PAIRSEL matched the category in PAIRRESP (e.g., a 12- to 17-year-old and 26- to 34-year-old were selected, but a single interview was obtained from a 35- to 49-year-old). In the latter situation, the respondent result (from PAIRRESP) was not totally consistent with what would be expected based on the pair selection, but PAIRRESP would not provide any information to directly contradict the indication from PAIRSEL that a 12- to 17-year-old was selected.

If PAIRSEL and PAIRRESP disagreed when two people were interviewed, with one indicating the selection or interview of a 12- to 17-year-old but the other variable did not, then special codes were assigned to SKPX1217. When this type of inconsistency occurred, a code of

11 was assigned to SKPX1217 when FIPE2 was originally answered as "yes," and a code of 12 was assigned when FIPE2 was originally answered as "no."

Suppose, for example, that PAIRSEL indicated that a 12- to 17-year-old and a 35- to 49-year-old were selected for the interview, but PAIRRESP indicated that an 18- to 25-year-old and a 35- to 49-year-old were actually interviewed, with the interviewer keying FIPE2 = 1 in the adult's interview (i.e., "yes," a 12- to 17-year-old was selected for an interview at this DU). In this situation, the "yes" in FIPE2 was consistent with who was *selected* (according to the information provided by the screening respondent), but it was not consistent with the ages provided by the respondents themselves. Therefore, the edited variable SKPX1217 would be set to a value of 11 in this example.

This latter edit preserved the information that the interviewer originally entered but also denoted that an inconsistency existed between PAIRSEL and PAIRRESP. This edit also was designed to preserve any possible parenting experiences data when both FIPE2 and FIPE3 (see below) were answered as "yes" but there was an inconsistency between PAIRSEL and PAIRRESP. When an inconsistency occurred between PAIRSEL and PAIRRESP, an analyst would have discretion about whether to use parenting experiences data in an analysis.

Editing of the FIPE3 Checkpoint. This checkpoint pertained to whether the respondent was the parent or legal guardian of the 12- to 17-year-old who also was selected to be interviewed at that DU. Respondents who were aged 12 to 17 were skipped out of FIPE3 and did not have an opportunity to be routed into the parenting experiences module. Therefore, when FIPE3 had been skipped because the respondent was 12 to 17, the edited FIPE3 variable SKPXPRNT was assigned a legitimate skip code.

No further editing of FIPE3 was done when PAIRSEL indicated that a 12- to 17-year-old was selected and PAIRRESP had some result *other than* that of two adults having been interviewed at that DU. The rationale for this approach was that FIPE3 was based on who the actual respondent was, provided that a 12- to 17-year-old was selected. For example, if PAIRSEL indicated that a 12- to 17-year-old and a 26- to 34-year-old were selected, but a 35- to 49-year-old and a 12- to 17-year-old were interviewed, and FIPE3 was answered as "yes" (i.e., this adult respondent is the parent of the youth who was selected), that 35- to 49-year-old respondent may indeed have been a parent or legal guardian of the youth who was selected. This principle also would have held if the selected youth did not respond. Therefore, any data that were present in the parenting experiences module would be preserved.

In contrast, the following situations could occur when FIPE3 was inconsistent with either PAIRSEL or PAIRRESP: (1) PAIRSEL indicates that a youth/adult pair was selected, but two adult interviews were obtained at that DU, or (2) PAIRRESP indicated that a youth/adult pair was interviewed, but PAIRSEL indicated that an adult/adult pair was selected. When either of these inconsistencies occurred, a code of 11 was assigned to SKPXPRNT when FIPE3 was originally answered as "yes," and a code of 12 was assigned when FIPE3 was originally answered as "no."

Suppose, for example, that PAIRSEL indicated that an 18- to 25-year-old and a 35- to 49-year-old were selected for the interview, but PAIRRESP indicated that a 12- to 17-year-old and a

35- to 49-year-old were actually interviewed, and the interviewer keyed FIPE2 = 1 and FIPE3 = 1 in the adult's interview. Stated another way, the interviewer indicated that "yes," a 12- to 17-year-old was selected for an interview at this DU, and "yes," this 35- to 49-year-old respondent was the parent of the 12- to 17-year-old youth who was selected. In this situation, FIPE3 was consistent with PAIRRESP but not PAIRSEL. Furthermore, based on who was interviewed at that DU, the 35- to 49-year-old may indeed be the parent of the 12- to 17-year-old who also was interviewed at that DU. In this situation, the edited SKPXPRNT would be set to a value of 11 to denote that this type of inconsistency has occurred. Again, this edit would preserve any possible parenting experiences data—especially in situations in which an adult/child *respondent* pair was obtained.

7.4.12.2 Editing of the Variables in the Parenting Experiences Module

The variables in the actual parenting experiences module were edited according to the final values assigned to SKPX2PER, SKPX1217, and SKPXPRNT based on the edits described in Section 7.4.12.1. In particular, if these three variables indicated that the respondent was not eligible to be administered the parenting experiences questions, then the edits assigned the appropriate legitimate skip codes to the parenting experiences variables. This included replacing blank values with legitimate skip codes when a code of 12 had been assigned SKPXPRNT and the parenting experiences module has been skipped. The rationale for this latter edit was that even if FIPE3 was answered as "no" when PAIRSEL and PAIRRESP were inconsistent, the adult respondent still may not have been the parent or legal guardian of the youth who also was selected for an interview at that DU.

Conversely, if a respondent had been skipped out of the parenting experiences module and the edited FIPE variables SKPX2PER, SKPX1217, or SKPXPRNT indicated that the respondent was *potentially* eligible to be administered the parenting experiences questions (i.e., the respondent skipped the module based on the original answers in the FIPE questions, but other data suggested that the respondent may have been eligible to be asked these questions), then the edited parenting experiences variables retained a value of "blank." For example, if FIPE2 had been keyed as "no" and it was inferred for SKPX1217 that a 12- to 17-year-old was selected (i.e., SKPX1217 = 3), then FIPE3 and the parenting experiences questions also would have been skipped. In this situation, the respondent's eligibility or ineligibility to be administered the parenting experiences questions could not be determined because the field interviewer (FI) was not routed to the final checkpoint. Therefore, it could not be determined whether the respondent should have been asked the parenting experiences questions or should have been skipped.

As noted previously, the interview has included a "lockout" feature since 2003 that does not allow interviewers or respondents to go back into the ACASI sections and change their answers once that section of the interview had been completed. However, interviewers could go back to the beginning of the interview after respondents had been administered the ACASI sections and change FIPE1 through FIPE3 in a manner that made the final value in FIPE3 inconsistent with the presence of data in the parenting experiences module. Specifically, FIPE3 ("Is this respondent the parent or legal guardian of the 12 - 17 year old child who was selected for an interview?") could be answered as something other than "yes," with at least some data existing in the parenting experiences module. According to the CAI logic, however, the module was to be administered only when the interviewer indicated that the respondent was the parent or guardian of the selected 12- to 17-year-old (i.e., FIPE3 = 1).

The following is an example of a scenario where parenting experiences data could exist when FIPE3 was not answered as "yes." If the FI initially answered FIPE1 through FIPE3 as "yes" (i.e., two people were selected at that DU, a 12- to 17-year-old was selected at that DU, and the respondent is the parent/guardian of that 12- to 17-year-old), the respondent would be routed through the ACASI parenting experiences questions. In the household roster section of the "back-end" demographics section, however, a "hard error" would be triggered if the second person selected to be interviewed was not identified in the roster. Before the interview could proceed, the interviewer would need to change the information in the household roster to make it consistent with the information in FIPE1 through FIPE3, or else the interviewer would need to go back and change the information in FIPE1 through FIPE3 to make it consistent with the roster. In particular, an interviewer could resolve this inconsistency by going back and changing one of the answers in FIPE1 through FIPE3 from "yes" to "no." Because the interviewer and respondent are locked out of ACASI, however, the parenting experiences data would be saved as it had been entered originally. This type of scenario was reproducible in testing of the CAI instrument and yielded a result where FIPE3 was not answered as "yes" but data existed in one or more questions in the parenting experiences module.

If interviewers changed the value in FIPE3 to "no," this would not present a problem in editing the parenting experiences data because the corresponding edited variable SKPXPRNT indicated that the respondent was not the parent or legal guardian of the youth who was selected for an interview. Consequently, parenting experiences data were edited to infer that these respondents should have legitimately skipped the parenting experiences module. Any data that existed in the parenting experiences module were overwritten with a code of 89 (or 9989, etc.).

However, changes made by the interviewer to FIPE1 through FIPE3 that resulted in SKPXPRNT having a final value of 98 (i.e., blank) were more problematic. Because the edits for SKPX2PER, SKPX1217, and SKPXPRNT were consistent with the pair-selection and pair-respondent data (PAIRSEL and PAIRRESP, respectively), the decision was made to retain the value of 98 in SKPXPRNT. Therefore, any nonblank values that existed in the parenting experiences module were replaced with codes for bad data.

One set of variables involved skip logic within the parenting experiences module. Specifically, respondents were skipped out of question PE04 (length of most serious discussion about the dangers of tobacco/alcohol/other drug use) when question PE03 had a value of 1 (i.e., talked with child 0 times in the past year about the dangers of tobacco/alcohol/other drug use), or if PE03 was answered as "don't know" or "refused." Standard procedures for assigning legitimate skip codes (Section 2.4.2) or propagating refusal codes (Section 2.4.3) were implemented in the edited version of question PE04 (PXSERDIS) depending on the response in PE03 (edited variable PXKIDYR).

Parents were asked to report the birth date of the youth who was selected for an interview at that DU (question PE01). However, the birth year that respondents could enter for the youth in question PE01 was restricted to ages that would be more consistent with selection of a 12- to 17-year-old (but also allowed for birth dates that would include 18 year olds, in case a 17-year-old respondent just recently had a birthday). Thus, respondents were prevented from entering birth dates that would be extremely inconsistent with selection of a 12- to 17-year-old (such as entry of the current interview year for the birth year).

A refinement has been implemented for the parenting experiences edit logic since 2003 to take into account the situation in which all remaining parenting experiences questions had been skipped because the respondent did not provide a date of birth for the selected youth in question PE01 or did not provide an age for the youth in question PE01B. Prior to 2003, the skipped parenting experiences variables were assigned legitimate skip codes. Since 2003, the parenting experiences variables have retained codes for "blank" when this pattern occurred. The effect of this refinement in 2003 was to make the frequencies of legitimate skip values in PXCHCIG (corresponding to question PE02) and subsequent parenting experiences variables agree with the total count of codes of 2 or 12 (i.e., "no"), or 89 or 99 (i.e., legitimate skip) in SKPXPRNT.

The CAI program also calculated an age for the youth who was selected for an interview based on the youth's date of birth (as reported by the parent) and the interview date at the start of the parenting experiences module. Respondents were asked to confirm this age (question PE01A). If parents did not confirm the age that the CAI program calculated for the youth, they were asked to provide a corrected age for the youth who was selected for an interview (question PE01B). Similarly, if respondents did not know or refused to report the date of birth of the selected youth, they were asked to report an age in question PE01B without having to indicate the youth's date of birth.

This information was captured in the created variable PXCHLDAG. Specifically, PXCHLDAG contained the age based on the reported date of birth for the youth and the interview date (if respondents confirmed that this age was correct), or else PXCHLDAG contained the age supplied by the respondent from question PE01B. If respondents supplied a corrected age for the youth in question PE01B that was between 12 and 18 and it mismatched the age of the youth that was calculated from the birth date and interview date information, the edited variables containing the birth date information for the youth (PXBMONTH, PXBDAY, and PXBYR) were assigned bad data values. If respondents answered question PE01B as "don't know" or "refused" when they were asked to provide a corrected age for the selected youth, that response of "don't know" or "refused" was assigned to PXCHLDAG. In addition, if respondents answered question PE01A as "don't know" or "refused" when they were asked to confirm the age of the youth who was selected for the interview, the age for the youth that the CAI program had previously calculated was retained in PXCHLDAG (see above). When values in parenting experiences variables had been set to bad data because SKPXPRNT was blank, PXCHLDAG also was assigned a code for bad data.

A recoded variable (PXCMPAGE, for "compare age") also was created that compared the selected youth's age (from PXCHLDAG) with the respondent's age for the second interview that was conducted at that DU. If two interviews were obtained at that DU and a 12- to 17-year-old was selected for an interview, then PXCMPAGE was calculated as the absolute value of the difference between PXCHLDAG and the actual age of the second respondent, within defined categories (i.e., 0 year difference in ages; 1 year difference in ages; 2 year difference in ages; 3 to 4 year difference in ages; and 5 or more year difference). If the adult respondent answered "don't know" or "refused" to the question about the youth's date of birth, or if the youth's date of birth information was set to bad data because of invalid dates, these codes were reflected in PXCMPAGE.

For the large majority of cases where an interview was obtained from a 12- to 17-year-old, PXCMPAGE indicated no difference between the age based on the date of birth reported by the parent and the youth's age that was recorded in the second interview at that DU. Nevertheless, information about more extreme differences in ages as recorded by PXCMPAGE (e.g., a difference of 2 or more years between the two ages) could be used by analysts in deciding whether to use the parenting experiences data in an analysis. When the second interview was from an 18 year old, PXCMPAGE was assigned a value of 18. When the second interview was from an adult older than age 18 (i.e., and the parent was supposed to be reporting about a 12- to 17-year-old), the edit program assigned a code of 50 to PXCMPAGE. Again, these codes were designed to give analysts discretion in using or disregarding parenting experiences data when the second interview at a DU came from an adult.

If a 12- to 17-year-old was supposed to be selected at a given DU but only the adult was interviewed, PXCMPAGE was assigned a code of 93. This code was assigned because there were no data to corroborate the youth's date of birth reported by the parent.

If the edited FIPE variables from above indicated that the respondent was not eligible to be administered the parenting experiences questions, then PXCMPAGE was assigned a code of 99 (i.e., legitimate skip). That included situations in which the edited FIPE3 was assigned a code of 12 because of an inconsistency between PAIRSEL and PAIRRESP, and the parenting experiences module had been skipped (see above). Otherwise, if the parenting experiences module was all blank or if PXCMPAGE was undefined for some other reason, then PXCMPAGE was assigned a code of 98. This code of 98 in PXCMPAGE meant "other missing." This code of 98 also was applied in PXCMPAGE when SKPXPRNT was blank and parenting experiences data had been replaced with codes for bad data.

7.4.13 Youth Experiences Module

As noted in Section 7.1, the youth experiences module was administered only to respondents aged 12 to 17. This section included questions about changes of residence in the past 5 years; school enrollment and related issues (e.g., opinions about the importance of assigned schoolwork) in the past 12 months, including homeschooling; other social and family characteristics (e.g., substance use behaviors of other students or friends, personal attitudes about substance use, parental attitudes about substance use); people with whom the youth could confide about a serious problem; exposure to alcohol- and other drug-related prevention messages in school or outside school; and personal behaviors (e.g., involvement in criminal or potentially criminal activities, involvement in extracurricular activities) that might be associated positively or negatively with the use of alcohol or other drugs. The youth experiences module also included questions about youths' religious involvement in the past 12 months and opinions about religious issues.

Minimal processing of data was done to variables in this section. The primary data processing involved assignment of legitimate skip codes based on the CAI routing logic. That included (1) assignment of legitimate skip codes to variables in the entire module for respondents who were aged 18 or older and (2) assignment of legitimate skip codes to youths' data based on routing logic within the youth experiences module.

Some special issues were encountered in editing the variables corresponding to question YE22, which pertained to people whom youths could turn to if they had a serious problem. Specifically, youths were asked to enter all the different types of people to whom they could turn to (e.g., a parent, a friend). This question also included a response category for youths who felt that there was no one they could talk to about a serious problem.

The questions indicating the youths' relationships to people whom they could turn to if the youths had a serious problem were "enter all that apply" questions. The individual edited variables for relationships were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (Section 2.4.4). If the entire list of responses was blank (e.g., if a youth broke off the interview before getting to these questions), the edited variables retained a code of "blank."

Youths could indicate that there was no one they could talk to about a serious problem but then indicate that they could talk to one or more of the people or types of people in the list from question YE22. In this situation, the variable pertaining to the first item in the list ("There is nobody I can talk to about a serious problem") was assigned a code of 11 (if that response was chosen along with another response from the list). Similarly, a code of 11 was assigned to the edited relationship variables (e.g., my mom, my dad) when they were chosen along with the response that there was nobody that the youth could talk to.

7.4.14 Mental Health Module

From 2004 to 2007, the mental health module for adults consisted of two primary components. First, a 12-month Kessler-6 (K6) distress scale was administered, and then questions about lifetime and 12-month major depressive episode (MDE) were asked. Since 2008, however, the K6 questions have collected data on distress in the past 30 days and in the past 12 months (see Section B.4.5 in the 2008 national findings report for details) (Office of Applied Studies, 2009). Respondents were routed into the 12-month version of the K6 if they reported having a period in the past 12 months when they felt more depressed, anxious, or emotionally stressed than they felt in the past 30 days.

In 2008, adult respondents also were administered one of two impairment scales—an abbreviated World Health Organization Disability Assessment Schedule (WHODAS) (see Section B.4.6 in the 2008 national findings report for details) (Office of Applied Studies, 2009) or the Sheehan Disability Scale (SDS)—and suicidal ideation questions. These impairment and suicide questions were placed after the K6 questions, but before the MDE questions. A random split-sample design was implemented in 2008 for adults where respondents in sample A were administered the WHODAS scale and respondents in sample B were administered the SDS. All adult respondents were administered the suicidal ideation questions after the impairment items, but before the MDE items. Starting with 2009, the SDS items were no longer included in the mental health module, and the WHODAS has been used to assess impairment.

Apart from propagating refusals from the lead questions for the WHODAS, minimal processing of data was done to the mental health variables. The primary data processing involved assignment of legitimate skip codes for respondents who were aged 12 to 17. Legitimate skip codes also were assigned to data from adults based on other skip logic within the module. If adults broke off the interview before they reached the WHODAS questions, then all of the WHODAS variables retained a code of 98 (i.e., "blank").

7.4.15 Adult and Adolescent Depression Modules

The adult and adolescent depression modules have been present in the interview since 2004. Questions in these modules were based on those used in the National Comorbidity Survey Replication for adults (NCS-R) and the National Comorbidity Survey Replication Adolescent Supplement (NCS-A) (http://www.hcp.med.harvard.edu/ncs/). These depression modules were included in NSDUH to produce lifetime and 12-month prevalence estimates of major depressive episode (MDE), severity of 12-month MDE, age at first MDE, lifetime number of episodes, current and 12-month treatment, and the respondent's perception of treatment effectiveness.

There were some differences in wording between related items in these modules, such as use of simpler wordings for the adolescent depression questions. For example, question AD17 in the adult depression module asked respondents how "severe" their "emotional distress" was during their worst periods lasting 2 weeks or longer when they had problems with their mood. The corresponding question YD17 from the adolescent depression module asked adolescents how "strong" their "bad feelings" were during these periods. Despite these differences in wording, similar naming conventions were used for the variables in these two modules. For example, the edited variable corresponding to question AD17 from the adult depression module was ADWRDST (where WR = worst period, and DST = distress), and the edited variable corresponding to question YD17 from the adolescent depression module was YOWRDST, even though YD17 did not ask about distress. Thus, the only difference in the names for analogous edited variables in these modules was in the use of the two-letter prefix that defined which module a given variable came from: "AD" for variables from the adult depression module and "YO" for adolescent depression variables.

There also were differences in how the CAI program created indicators of MDE for adolescents and adults based on differences in the NCS-A and NCS-R. Specifically, the criteria for defining respondents as having the symptom of loss of interest or pleasure in most things was less restrictive for adolescents than for adults. In particular, the DSM-IV criteria for MDE (American Psychiatric Association, 1994) place more emphasis for adolescents on the cognitive aspects of depression, such as boredom or apathy, rather than on somatic or physical complaints, such as sleep loss, that may be manifest in adults with MDE. For example, somatic or physical complaints, such as sleep loss, may be due to factors during adolescence other than depression. Consequently, somatic or physical complaints that may be associated with MDE among adults function less well as indicators of MDE among adolescents than do cognitive indicators. For this reason, the CAI logic gave adolescents additional opportunities to be classified as having the symptom of loss of interest based on their answers to questions that were not taken into account in classifying adults as having this symptom.

Despite these differences between the adult and adolescent depression modules, the basic logic for asking questions was similar between the two modules. Therefore, the remainder of this section discusses edits for both of these modules together. Except where differences are discussed in terms of how variables were edited for these modules, the same basic edits discussed below applied to variables in both modules.

An important aspect of the processing of variables in these modules consisted of assigning legitimate codes based on the routing logic within these modules. In particular, adults were assigned legitimate skip codes to the edited variables in the adolescent depression module, and adolescents were assigned legitimate skip codes to the edited variables in the adult depression module.

As an additional example, respondents were asked a series of questions to identify changes in appetite or weight. They first were asked whether they had a much smaller appetite than usual during the most recent time when their problems were the worst (questions AD26A and YD26A, corresponding to edited variables ADWRELES and YOWRELES, respectively). If respondents answered the relevant question as "yes," they were skipped out of subsequent questions about increases in appetite and weight gain. If respondents reported having less appetite, they were asked whether they lost weight without trying to, and if so, whether their weight loss was due to them being sick or on a diet; respondents who indicated that their weight loss was due to sickness or a diet were not asked to report how many pounds they lost in weight.

Conversely, if respondents did not report in AD26A/YD26A that they had less appetite than usual, they were asked whether they had a much larger appetite than usual (questions AD26B and YD26B, corresponding to edited variables ADWREMOR and YOWREMOR, respectively). Respondents who reported that they had a much larger appetite were skipped out of remaining questions related to weight loss. These respondents subsequently were asked whether they gained weight without trying to. If respondents reported gaining weight without trying to, the CAI program asked follow-up questions to rule out weight gains due to growth (for respondents aged 21 or younger) or pregnancy (for females); respondents who indicated that they gained weight because they were growing or because they were pregnant were not asked to report how many pounds they gained. Thus, editing of the adult and adolescent depression variables related to changes in appetite or weight involved assignment of legitimate skip codes to these variables based on the routing logic for the corresponding questions.

As noted in Section 2.4.3, legitimate skip codes generally were not assigned if a lead question was answered as "don't know" or "refused." However, important exceptions to this principle were made in editing of the adult and adolescent depression variables because of consideration of other aspects of the routing logic in these modules. In particular, the lead screening questions ASC21 through ASC23 (corresponding to edited variables ADDPREV, ADDSCEV, and ADLOSEV) and YDS21 through YDS23 (corresponding to edited variables YODPREV, YODSCEV, and YOLOSEV) at the beginning of these modules had a special skip logic. If a particular lead question was not answered affirmatively, this logic routed respondents into follow-up questions that could screen respondents into further questions about depression.

In the adult depression module, for example, if question ASC21[130] was answered as "don't know" or "refused," the subsequent question AD01[131] was skipped. In this situation, however, respondents were routed to follow-up question ASC22.[132] Thus, if respondents answered question ASC21 "don't know" or "refused" but they answered question ASC22 as "yes," they were still eligible to be administered the remainder of the adult depression module, depending on how they answered subsequent questions.

Therefore, legitimate skip codes were assigned (where relevant) to the variables corresponding to questions AD01 through AD09 in the adult depression module (edited variables ADDPDISC, ADDPLSIN, ADDSLSIN, and ADLSI2WK) if at least one item from questions ASC21, ASC22, or ASC23 was answered as "yes" or "no." For example, if question ASC21 was answered as "don't know," and question ASC22 or ASC23 was answered as "yes" or "no," the editing procedures assigned a legitimate skip code to ADDPDISC, corresponding to question AD01. Similarly, if question ASC22 was answered as "don't know" but ASC23 was answered as "yes" or "no," the editing procedures assigned a legitimate skip code to ADDSLSIN, corresponding to question AD02. In turn, if ADDPDISC, ADDPLSIN, and ADDSLSIN all were answered as "no" or had legitimate skip codes after the above edits, then ADLSI2WK (corresponding to AD09) was assigned a legitimate skip code. The values in ADDPDISC, ADDPLSIN, ADDSLSIN, and ADLSI2WK determined whether subsequent variables were assigned legitimate skip codes. The logic provided in this example for adult depression also was applied to the corresponding variables in the adolescent depression module.

As discussed in Section 2.4.3, if a lead question that governs a skip pattern was refused, the editing procedures typically "propagated" that refusal from the lead question to the variables that had been skipped. For most of the adult and adolescent depression variables, however, this refusal propagation was not performed. The CAI program contained routines for scoring the symptom indicators for MDE. The CAI program coded the symptom score variable DSMMDEA2 as 2 if the sum of the numbers of codes of 1 (i.e., has symptom), "don't know," or "refused" in the individual symptom indicators was less than 5. Therefore, not propagating refusals helped to avoid situations in which a different overall score might be obtained if analysts were to calculate DSMMDEA2 based on edited variables.

The exception to this rule of not propagating refusals in the adult and adolescent depression modules concerns the final questions regarding receipt of counseling from a medical doctor or other professional about the respondents' symptoms of depression (e.g., questions AD86 through AD86F in the adult depression module). For example, if question AD86 (edited variable ADSEEDOC) was refused, the editing procedures still propagated that refusal code to the skipped variables that were dependent on AD86.

[130] Question ASC21 asked, "Have you ever in your life had a period of time lasting several days or longer when most of the day you felt sad, empty, or depressed?"

[131] If respondents answered question ASC21 as "yes," they were asked question AD01. Question AD01 asked, "During times when you felt sad, empty, or depressed most of the day, did you ever feel discouraged about how things were going in your life?"

[132] Question ASC22 asked, "Have you ever had a period of time lasting several days or longer when most of the day you were very discouraged about how things were going in your life?"

In addition, the CAI program created MDE symptom variables and overall MDE symptom scores for adults and adolescents. Table 7.5 lists the final, edited variables that were created from these symptom variables and overall symptom scores. For each variable, explanation of the meaning of that variable also is provided. The only editing that was done to these variables in Table 7.5 was to assign legitimate skip codes based on the respondent's age (i.e., 12 to 17 or 18 or older). Thus, the values that were created by the CAI program were preserved in the variables listed in Table 7.5.

Table 7.5 Depression Symptom and Score Variables

Adult Depression Variable	Adolescent Depression Variable	Explanation
AD_MDEA1	YO_MDEA1	Respondent (R) felt sad, empty, depressed, or discouraged most of the day.
AD_MDEA2	YO_MDEA2	R lost interest or pleasure in most things.
AD_MDEA3	YO_MDEA3	R had changes in appetite or weight (not due to growth, pregnancy, illness, or dieting).
AD_MDEA4	YO_MDEA4	R had sleep problems.
AD_MDEA5	YO_MDEA5	Others noticed that the R was restless or lethargic.
AD_MDEA6	YO_MDEA6	R felt tired or low on energy nearly every day.
AD_MDEA7	YO_MDEA7	R felt worthless nearly every day.
AD_MDEA8	YO_MDEA8	R was unable to concentrate or make decisions.
AD_MDEA9	YO_MDEA9	R was suicidal (had thoughts of suicide, made plans, or made an attempt).
ADSMMDEA	YODSMMDE	Score of symptom indicators 1 through 9 from above.

Relatively little additional editing was done to the adult and adolescent depression variables, aside from assigning legitimate skip codes. Additional editing issues that were relevant to these modules are described in the remainder of this section.

If respondents reported a period of time when their symptoms or problems were the worst, they were asked to report how old they were when this time started (edited variables ADWRAGE for adults and YOWRAGE for adolescents). In addition, if respondents scored as positive for MDE (edited variables ADSMMDEA and YODSMMDE) and they reported that these problems caused some interference with their work, social life, or relationships, they were asked to report the age at which these problems first occurred (edited variables ADPBAGE and YOPBAGE). If respondents reported an age of onset in any of these variables that was greater than their current age, these variables were set to bad data. For adults, if ADWRAGE had been set to bad data and the respondent's original answer was age 22 or greater, the respondent was skipped out of the question that asked about weight gain because the respondent was growing (edited variable ADWRGROW). Therefore, if ADWRAGE had been set to bad data, ADWRGROW retained a code of blank.

As noted previously, if respondents reported gaining or losing weight and these gains or losses could not be attributed to factors other than depression (e.g., growth, pregnancy, dieting), respondents were asked to report the number of pounds they gained or lost. In particular, respondents were allowed to report that they gained or lost 0 pounds. No editing was done to the variables ADWRGNLB, ADWRLSLB, YOWRGNLB, or YOWRLSLB when this response of 0 pounds occurred because respondents did not have an opportunity to report gains or losses of less than 1 pound. Furthermore, only gains or losses of 10 or more pounds resulted in respondents being scored as having the symptom of changes in appetite or weight.

In addition, a feature of the logic for asking respondents about gains or losses in weight was that if respondents were asked the question about gaining weight without trying and they answered it as "don't know" or "refused," they had an additional opportunity to be asked questions about losing weight without trying. That is, the program was looking for the first affirmative set of answers that would allow a determination to be made of whether respondents gained or lost enough pounds to qualify for being depressed. Consequently, no editing was done if respondents originally gave an answer in the questions corresponding to ADWREMOR or YOWREMOR (i.e., having a much larger appetite than usual), they answered the weight gain question (corresponding to ADWRGAIN or YOWRGAIN) as "don't know" or "refused," and then they were routed into the questions about weight loss.

If respondents reported that they talked to a medical doctor or other professional in the past 12 months about the problems they were experiencing related to depression, they were asked to report which professionals they saw or talked to. In the question pertaining to the specific professionals that they saw or talked to, respondents were allowed to enter more than one type of professional from the list they were presented. As in other modules, the "enter all that apply" variables in the adult and adolescent depression modules were coded as 1 if the response was entered and as 6 if the response was not entered (Section 2.4.4).

Codes of 94 and 97 (for "don't know" and "refused," respectively) were assigned to an entire list of variables if respondents did not know or refused to report the specific professionals that they saw or talked to about their problems. If the entire list was blank but respondents had previously reported that they saw or talked to a professional about their problems, then the specific variables corresponding to categories of helping professionals retained a code of 98 (i.e., "blank").

Respondents could report that they saw or talked to "another type of helping professional" and then specify a helping professional that they had already been asked about, such as a psychiatrist. Thus, for example, if the edited variable ADPSYMD, pertaining to services from a psychiatrist, was not coded as 1 and respondents specified that they saw or talked to a psychiatrist, ADPSYMD was assigned a code of 3, where 3 = Response entered LOGICALLY ASSIGNED.

7.4.16 Youth Mental Health Service Utilization Module

The module on youth mental health service utilization asked respondents aged 12 to 17 about their receipt of specific sources of inpatient, foster care, outpatient, or school-based mental health services in the past 12 months; the number of nights that respondents spent in specific inpatient or foster care mental health settings; the number of times they visited specific types of outpatient or school-based mental health providers; and the reasons for receiving inpatient, foster care, outpatient, or school-based services for mental health problems the last time they received such services. Specific sources of mental health services that respondents were asked about included (1) any type of hospital, (2) a residential treatment center, (3) foster care or a therapeutic foster home, (4) a partial day hospital or day treatment program, (5) a mental health clinic or center, (6) a private therapist, (7) an in-home therapist, (8) a pediatrician or other family doctor, (9) special education services, and (10) in-school counseling, such as from school counselors or school psychologists.

In 2009, a set of questions from prior years was deleted from the youth mental health utilization module; the deleted questions asked about receipt of special education services and receipt of school counseling in the past 12 months, and whether youths were ever in jail or foster care. In their place, a set of questions has been included since 2009 that measure the same constructs as the questions from prior years but with different wordings. For youths' experiences with the juvenile justice system, the questions since 2009 also focus on the past 12 months rather than the lifetime period and collect additional information about these encounters with the juvenile justice system. The revised questions that have been included since 2009 asked about the following for the past 12 months: (1) receipt of counseling or treatment for mental health problems from social workers, school psychologists, or school counselors; (2) the reason for the last talk with a social worker, school psychologist, or school counselor (if applicable); (3) attendance at a school for students with emotional or behavioral problems or participation in a special program at a regular school for students with emotional or behavioral problems; (4) time spent overnight or longer in a juvenile detention center; (5) the number of nights spent in a juvenile detention center (if applicable); and (6) receipt of treatment or counseling for emotional or behavioral problems in a juvenile detention center (if applicable).

An important aspect of processing the variables in this section involved assignment of legitimate skip codes, where relevant. That included (1) assignment of legitimate skip codes to variables in the entire module for respondents who were aged 18 or older and (2) assignment of legitimate skip codes to youths' data based on routing logic within the youth mental health service utilization module. For example, if respondents reported that they did not stay overnight or longer in a hospital to receive mental health counseling in the past 12 months (YUHOSPYR = 2), all subsequent variables pertaining to mental health services in a hospital were assigned legitimate skip codes. That included the number of nights that respondents stayed in a hospital and the reasons that they were hospitalized the last time.

Although respondents in the youth experiences module who reported that they were not enrolled in school in the past 12 months were asked whether they were homeschooled during this period, the youth experiences variable pertaining to homeschooling (YEHMSLYR, corresponding to question YE09A) was not used to edit youth mental health service utilization variables pertaining to receipt of school-based mental health services. Only the youth experiences variable pertaining to school enrollment in the past 12 months (YEATNDYR, corresponding to question YE09) was used to edit these school-based service variables.

If respondents reported that they stayed overnight or longer in foster care or in a therapeutic foster care home in the past 12 months for emotional or behavioral problems, they were not asked whether they had ever been in foster care. Therefore, the edited variable pertaining to foster care in the lifetime (YUFCAREV) was assigned a code of 5 (Yes LOGICALLY ASSIGNED [from skip pattern]). This code of 5 indicated that it could be logically inferred that respondents had ever been in foster care because they reported being in foster care in the past 12 months.

Similarly, if the variable pertaining to foster care in the past 12 months (YUFCARYR) initially had a missing value (e.g., if respondents did not know or refused to report whether they stayed in foster care in the past 12 months) but respondents reported that they had never been in foster care (YUFCAREV = 2), it could be inferred that these respondents had not been in foster care in the past 12 months. In these situations, the edited variable YUFCARYR was assigned a final code of 4 (No LOGICALLY ASSIGNED). The remaining variables related to foster care in the past 12 months were assigned legitimate skip codes.

For each type or location of mental health treatment or counseling that respondents were asked about, they could report that they received services the last time at that particular location for any of the following reasons: (1) they thought about or tried to kill themselves, (2) they felt depressed, (3) they felt very afraid or tense, (4) they were breaking rules or "acting out," (5) they had eating problems, (6) they had trouble controlling their anger, (7) they had gotten into physical fights, (8) they had problems at home or in their families, (9) they had problems with their friends, (10) they had problems with people other than friends or family, (11) they had problems at school, or (12) some other reason. The reasons pertaining to trouble controlling anger through problems at school have been included since 2005 based on identification of commonly reported "other" reasons that respondents specified prior to 2005 for why they received treatment.

For each mental health service location where youths received services, information on these reasons for receiving services subsequently was captured as a discrete variable. For example, if respondents reported receiving mental health counseling from a pediatrician or family doctor, information about why they received counseling the last time was captured in the variables YUFDSUIC (suicidal), YUFDDEPR (depressed), YUFDFEAR (afraid and tense), YUFDBKRU (breaking rules), YUFDEATP (eating problems), YUFDANGR (anger), YUFDFITE (physical fights), YUFDFMLY (problems in family), YUFDFRND (problems with friends), YUFDOTPP (problems with people other than family or friends), YUFDSCHL (problems at school), and YUFDSOR (some other reason). The individual edited variables for these "enter all that apply" variables were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (Section 2.4.4). No further editing was done if respondents endorsed every single reason on a list as pertaining to why they received mental health services at a given location in the past 12 months.

Codes of 94 and 97 (for "don't know" and "refused," respectively) were assigned to an entire list of variables if respondents did not know or refused to report why they received counseling at a specific location in the past 12 months; this applied as well to the items for the additional reasons why respondents received counseling that were added in 2005. If an entire list of reasons was blank but respondents had previously reported receiving services at a given location (e.g., if respondents broke off the interview), then the list of reasons for receiving services at that location retained a code of 98 (i.e., "blank").

For purposes of illustration, if youths reported in question YSU22 that they had received treatment or counseling from a pediatrician or other family doctor in the past 12 months (i.e., for emotional or behavioral problems that were not caused by alcohol or drugs), they were routed first to question YSU23 regarding the number of times they received treatment from a family doctor, and then they were routed to question YSU24, regarding the reason(s) for their visit; question YSU24 has been present in the module ever since the module was added to the survey.[133] As in prior years, YSU24 was an "enter all that apply" item, and individual variables were created corresponding to the specific reasons why respondents got treatment.

Respondents could report up to six reasons in YSU24 for why they received treatment from a family doctor, including "some other reason." If respondents chose the "some other reason" option in question YSU24, they were routed next to question YSU24A, where they could choose the additional reasons for receiving treatment that were noted above, such as difficulty controlling anger; again, respondents were given the option in question YSU24A to report "some other reason." If respondents chose this "other" response category in YSU24A, they were asked to specify the *most important* other reason why they did got treatment in this location; this request for the most important other reason was a noteworthy change relative to prior years (see below).

Based on the logic noted above, the sixth response option in question YSU24 (some other reason) was principally considered to be a "toggle" to question YSU24A. Therefore, a separate "some other reason" variable was not created to correspond to the last response category in YSU24.

The edits described below have been implemented since 2005 to take into account the items on additional reasons why youths received treatment (e.g., YSU24A).

- If respondents chose the sixth response option in YSU24, any response that was entered from the YSU24A series was coded as 1, and anything that respondents did not choose from the YSU24A list was coded as 6.
- If respondents chose a response from YSU24 but did not choose the sixth response category in YSU24, the variables corresponding to the response categories in YSU24A (i.e., YUFDANGR through YUFDSOR) all were given a code of 6 (Response not entered), rather than being assigned "legitimate skip" codes. That is, YSU24 and YSU24A were considered together to be one big series of reasons.
- If respondents chose the sixth category in question YSU24, chose at least one reason from YSU24A, but did not choose category 7 in YSU24A (some other reason), the edited variable YUFDSOR (some other reason for receiving treatment) was assigned a code of 6. That is, it was inferred in this situation that the list of specific reasons in YSU24A was adequate for capturing why respondents saw a family doctor about emotional or behavioral problems in the past 12 months. For example, if a respondent chose category 6 in YSU24 and then chose only category 3 in YSU24A ("You had problems at home or in your family"), it would be reasonable to infer that this response in YSU24A was the only other reason why the respondent visited a family doctor about emotional or behavioral problems.

[133] Questions in this module for treatment that youths received in other settings were structured in the same manner as in this example for treatment from a family doctor. The changes to the questions for additional reasons why youths received treatment from a family doctor also applied to the other sources of treatment in this module.

- If respondents chose the sixth response category in question YSU24, it was possible for them to answer YSU24A as "don't know" or "refused" (i.e., did not know or refused to report what the other reasons were). When this occurred, the "some other reason" variable YUFDSOR was set to 1 (Response entered) in order to retain information that the respondent chose "some other reasons" somewhere in the series. Remaining variables corresponding to the YSU24A series retained codes of 94 ("don't know") or 97 ("refused").
- If respondents answered question YSU24 as "don't know" or "refused," question YSU24A was skipped. Therefore, the relevant code of 94 or 97 was propagated onto the variables corresponding to the YSU24A list.
- If YUFDSOR had a value of 6 (see above), the "OTHER, Specify" variable YUFDIMPR (i.e., the most important other reason why the respondent received treatment from a family doctor) was assigned a legitimate skip code. If YUFDSOR had a refusal code, that refusal was propagated onto YUFDIMPR.
- If YUFDSOR had a code of 1 when the respondent answered "don't know" or "refused" to the YSU24a series, the "OTHER, Specify" variable YUFDIMPR (which has been three digits in length since 2005) retained a code of 998 (blank).

Consistent with general editing procedures, if respondents reported a reason that corresponded to a reason in the lists for YSU24/YSU24A, that reason was logically inferred to have been chosen in the relevant edited variable. Suppose, for example, that the most important other reason that a respondent reported for receiving treatment from a family doctor indicated that the respondent was breaking rules or "acting out." If the respondent had not chosen this response in YSU24, the edited variable YUFDBKRU was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). Similarly, if the respondent specified that a problem in the respondent's family was the most important other reason for receiving treatment from a family doctor and the respondent had not chosen this response in YSU24A, the edited variable YUFDFMLY was assigned a code of 3. Conversely, if respondents did not report "some other reason" why they received treatment in the past 12 months from a family doctor (edited variable YUFDSOR = 6, corresponding to response category 6 in question YSU24 not being chosen or response category 6 in question YSU24A not being chosen), legitimate skip codes were assigned to the edited "OTHER, Specify" variable YUFDIMPR (corresponding to question YSU24SP).

As noted previously, the "OTHER, Specify" items in this module underwent an important change in 2005. Prior to 2005, youths were asked to specify "the other reason" why they received services in a particular location or from a given type of provider. Some respondents in these prior years gave a considerable amount of information in the space that was allotted to them to specify their other reason(s) for receiving services. Often, multiple reasons were reported. Therefore, prior to 2005, up to five separate "OTHER, Specify" codes were assigned for a given treatment location or provider based on respondents' explanations regarding why they received services.

Because youths since 2005 have been asked to specify the "most important" other reason why they received services in a particular location, only one reason was captured in the "OTHER, Specify" variables. If respondents specified more than one reason as the "most important" other reason why they received treatment in a given location, only the first reason that respondents specified was coded as a general rule. The exception to this rule was that any reports of respondents thinking about or trying to kill themselves were given precedence in coding, regardless of whether this reason was specified first. In addition, the change in the "OTHER, Specify" variables for this module was used as an opportunity to revise the "OTHER, Specify" codes to regroup related reasons together, such as reasons that indicated that respondents had a diagnosed condition (e.g., attention-deficit/hyperactivity disorder, or ADHD). For these reasons, the "OTHER, Specify" data from 2005 onward are not comparable with data prior to 2005.

In a relatively rare number of situations, youths denied receiving mental health services as part of an "OTHER, Specify" response. In these situations, the "OTHER, Specify" response was assigned a bad data code. Data were retained that indicated that the youths received mental health services in a given location in the past 12 months.

Respondents could report that the number of nights they stayed overnight in a hospital or residential treatment program in the past 12 months (or the sum of the two, if respondents reported staying in both settings) was greater than or equal to 365 nights. In these situations, no editing was done to the data. If respondents reported spending 366 nights in an inpatient/residential setting or in foster care for treatment or counseling for emotional or behavioral problems, however, the number of nights in the corresponding edited variable was "trimmed" back to 365. For example, if respondents reported staying overnight in a hospital for 366 nights for treatment or counseling for emotional or behavioral problems, the corresponding edited variable YUHOSPNM was set to a value of 365.

7.4.17 Consumption of Alcohol Module

The consumption of alcohol module covered a variety of topics related to respondents' use of alcohol. Administration of questions in this module was limited to respondents who had ever used alcohol, and a subset of questions was limited to respondents who reported alcohol use in the past 30 days. Respondents were asked questions about the following topics:

- number of drinks consumed on the last occasion of alcohol use in the past 30 days;
- the last occasion of alcohol use in the past 30 days by adolescents or young adults younger than the legal drinking age of 21, including who respondents were with, where they drank, and how they obtained the alcohol;
- use of other drugs in combination with alcohol (i.e., at the same time or within a couple of hours of using alcohol) in the past 30 days; and
- history of binge alcohol use (i.e., consumption of five or more drinks on a single occasion for males or four or more drinks on a single occasion for females).

This section documents the editing procedures for the consumption of alcohol module.

7.4.17.1 Skip Logic Issues in the Consumption of Alcohol Module

One of the important aspects of the processing of variables in the consumption of alcohol module consisted of assigning codes of 91, 93, and 99 to variables that had been skipped because the questions did not apply. In particular, certain questions may not apply to respondents for multiple reasons, such as if they never used alcohol or if they were male.

Therefore, Table B.26 in Appendix B discusses the prioritization of skip logic edits in this module. For example, if respondents had never used alcohol, variables in the consumption of alcohol module were assigned a code of 91, 991, or 9991 (NEVER USED ALCOHOL), regardless of whether questions might not be applicable for other reasons. In particular, male respondents who had never used alcohol were assigned codes of 91, 991, or 9991 in the questions about females' history of consuming four or more drinks on a single occasion, even though males were not eligible for these questions because of their gender. However, this edit kept a constant number of cases coded as indicating nonuse of alcohol for most variables in the module.

7.4.17.2 Edit Issues for the 30-Day Consumption of Alcohol Variables

Other than implementation of edits based on the skip logic (see Table B.26), only limited editing was done for the variables pertaining to the number of drinks that respondents consumed the last time they drank in the past 30 days (question CA01; edited variable CADRLAST) and the variables pertaining to alcohol use in the past 30 days among adolescents or young adults aged 12 to 20 (questions CA02A through CA07SP). In particular, no editing was done to CABUYFRE (corresponding to question CA03), CAGVMONY (corresponding to question CA04), CABUYWHO (corresponding to question CA05A), or CABPLACE (corresponding to question CA05B) if respondents reported in question CA04 that they gave money to someone else who bought their last alcoholic beverage for the respondents, but then they reported in CAFRESP (corresponding to CA07SP) that the "other" way that they got the last alcoholic beverage they drank was by purchasing it, such as with fake identification. The rationale for not editing these variables was that there may not be enough detail from what respondents specified to know how these variables should be edited. For example, question CA04 about buying alcohol was asked only if respondents reported in question CA03 that they paid for the last alcohol that they drank. Even if the respondent specified buying alcohol in CAFRESP, however, it would not be known whether the respondent bought the alcohol with his or her own money or whether the alcohol was purchased with someone else's money.

In question CA07 (edited variable CAFREWHO), respondents were asked to give one response for how they got the last alcoholic beverage that they drank, including the option to report that they got it "some other way." Respondents who reported that they got the alcohol some other way were asked to specify what this other way was (edited variable CAFRESP). If a response was given in CAFRESP that corresponded to one of the existing response options in question CA07, then CAFREWHO was assigned the relevant response plus a value of 10. For example, if CAFRESP indicated that respondents got their last alcoholic beverage from a nonrelative who was 21 or older who gave the alcohol to the respondent (i.e., equivalent to CA07 = 3), then CAFREWHO was assigned a code of 13.

Questions CA02B (where underage respondents drank alcohol the last time) and CA09 (the specific drugs that respondents of any age used at the same time or within a couple of hours of when they last drank alcohol) were "enter all that apply" questions. As in other modules, the responses were captured as discrete variables. The individual edited variables for these "enter all that apply" variables were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (Section 2.4.4).

For the edited variables corresponding to question CA02B, if respondents specified some other place where they drank alcohol the last time that corresponded to one of the existing response options in question CA02B, and they had not chosen that response, then the corresponding edited variable was assigned a code of 3, where 3 = Response entered LOGICALLY ASSIGNED. For example, if respondents reported that they drank alcohol in someone else's home the last time they drank but they had not chosen this response in question CA02B, then the edited variable CADROTHM was assigned a code of 3. In addition, codes of 94 and 97 (for "don't know" and "refused," respectively) were assigned to the entire list of variables corresponding to question CA02B if respondents did not know or refused to report where they were the last time they drank alcohol.

For the edited variables corresponding to question CA09, codes of 94 or 97 were retained only for those drugs that respondents used in the past 30 days (or were potentially past month users, such as if a core recency variable had been coded as 8, meaning that the respondent used the drug at some point in the past 12 months, which could include within the past 30 days). Otherwise, codes of 94 or 97 in the individual drug variables corresponding to question CA09 were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned) if respondents used some drug in the past 30 days but not the drug that corresponded to the variable of interest. Suppose, for example, that respondents used marijuana and inhalants in the past 30 days and did not use any other illicit drugs in that period. If question CA08 was answered as "yes" and CA09 was refused, then the edited variables CADRKMRJ and CADRKINH retained a code of 97 and the remaining drug variables were assigned a code of 89.

Special consideration also was given to the editing of the variables CADRKCOC (last use of alcohol in combination with cocaine), CADRKHER (last use of alcohol in combination with heroin), CADRKSTM (last use of alcohol in combination with prescription stimulants), and CADRKMTH (last use of alcohol in combination with methamphetamine). In particular, respondents who answered question CA08 as "yes" would have the response option "Cocaine or 'crack'" (captured in CADRKCOC) available to them if they previously reported past month use of cocaine in any of the following places: (1) in the core cocaine module, (2) in the core crack cocaine module, or (3) in the noncore special drugs module (i.e., use of cocaine with a needle in the past 30 days). Similarly, respondents who answered question CA08 as "yes" could choose the response option "Heroin" (captured in CADRKHER) if they reported past month use of heroin in the core heroin module or if they reported past month sniffing, smoking, or injection of heroin in the noncore special drugs module. The response option "Prescription stimulants" (captured in CADRKSTM) was available to respondents if they reported nonmedical use of stimulants in the past 30 days in the core stimulants module or if they reported that they used stimulants other than methamphetamine with a needle in the past 30 days. For methamphetamine, the response option "Methamphetamine" (captured in CADRKMTH) was available to respondents who indicated past month methamphetamine use in the core stimulants module or who indicated past month use in the noncore special drugs module, provided that these additional methamphetamine users from the noncore special drugs module did not indicate in question SD17A1 that their earlier report from the core that they never used methamphetamine was actually the correct response (Section 6.2.6).

Consequently, respondents would have the response options for CADRKCOC, CADRKHER, CADRKSTM, or CADRKMTH available to them if they reported past month use (or nonmedical use) of the relevant drug(s) in the noncore special drugs module, even if they had not reported past month use in the corresponding core module(s). In most situations, no further editing was done to these data, even though the corresponding core data did not indicate use in the past 30 days. In particular, no editing was done to data in CADRKMTH if respondents reported in the core stimulants module that they never used methamphetamine, but they reported past month use of methamphetamine in the special drugs module and they also reported in the follow-up question SD17A2 that they "made a mistake" (or a similar other reason for not reporting methamphetamine use) when they were answering the core question about methamphetamine.

The one exception to this editing principle concerned editing of CADRKSTM. Respondents who reported nonmedical use only of methamphetamine and of "some other stimulant" in the core stimulants module were routed to question ST09 (for the most recent nonmedical use of any stimulant) and to question ST19 (for the most recent use of methamphetamine). If respondents reported nonmedical use of stimulants in the past 30 days in question ST09 but they were logically inferred to be lifetime nonmedical users of only methamphetamine (i.e., the only "other" stimulant that they specified using nonmedically was methamphetamine), then any data in the response option corresponding to CADRKSTM were overwritten with a code of 89 as part of the data editing. That is, it was logically inferred that these respondents should not have been asked about their use of *prescription* stimulants while they were drinking alcohol or within a couple of hours of drinking alcohol in the past 30 days.

7.4.17.3 Edit Issues for the Binge Alcohol History Variables

Questions CA10 through CA14D captured information on respondents' history of having consumed five or more drinks on a single occasion (regardless of gender) or four or more drinks on a single occasion (for females). If respondents were lifetime alcohol users and other data from the core alcohol module or from CADRLAST did not indicate that respondents had consumed five or more drinks on a single occasion in the past 30 days, then they were asked in question CA10 (edited variable CABNGEV) whether they had ever had five or more drinks on a single occasion in their lifetime. Those respondents who had ever consumed five or more drinks on a single occasion (or had reported this in the past 30 days in the core alcohol module or in CADRLAST) were asked in question CA11 (edited variable CABNGAGE) to report how old they were the first time when they first had five or more drinks. If CABNGAGE was within 1 year of respondents' current age, they were routed to questions CA11A through CA11D (edited variables CABNGYFU and CABNGMFU) to report the year and month when they first had five or more drinks on a single occasion.

Similarly, female respondents were asked question CA12 (edited variable CA4FDEV) if they were lifetime alcohol users and other data from the core alcohol module or from CADRLAST did not indicate that they had consumed four or more drinks on a single occasion in the past 30 days, and CABNGEV did not indicate that they had ever consumed five or more drinks on a single occasion. In addition, question CA13 (edited variable CA4FDDYS) asked female respondents who used alcohol in the past 30 days to report the number of days they had four or more drinks on a single occasion in the past 30 days. Question CA14 (edited variable CA4FDAGE) asked female respondents to report how old they were the first time they had four or more drinks on a single occasion. If CA4FDAGE was within 1 year of female respondents' current age, they were routed to questions CA14A through CA14D (edited variables

CA4FDYFU and CA4FDMFU) to report the year and month when they first had four or more drinks on a single occasion.

As noted previously, the routing logic for these questions took into account respondents' answers in the core alcohol module. Specifically, the computer-assisted interviewing (CAI) program created a flag variable called FOURORMOREFLAG. The default value for this flag was 0. The flag was coded as 2 if there was some indication in the alcohol module that respondents had consumed five or more drinks on a single occasion in the past 30 days. Otherwise, if question AL07 indicated that respondents usually had four drinks on those days when they drank in the past 30 days, then this flag was coded as 1.

The edited variable corresponding to this flag was named ED4FLAG. If the variables NODR30A (corresponding to question AL07) and DR5DAY (corresponding to question AL08) indicated consumption of five or more drinks on a single occasion but ED4FLAG did not already have a value of 2, then ED4FLAG was set to a logically assigned value of 12. Otherwise, if NODR30A indicated usual consumption of four drinks in the past 30 days and ED4FLAG did not already have a value of 1, then ED4FLAG was set to a value of 11. If NODR30A and DR5DAY had been set to bad data as a result of edits in the core alcohol module (see Section 6.2.4.7 and Table B.7 in Appendix B), then ED4FLAG also was set to bad data. In turn, any questions that had been answered in CA11 through CA14D were set to bad data.

Revised routing logic for variables related to binge alcohol history has been in place since 2007 based on patterns that were identified during review of the edited data in the 2006 survey. Specifically, the routing logic was revised for question CA11 (edited variable CABNGAGE) to take into account FOURORMOREFLAG (edited variable ED4FLAG). In 2006, the logic took into account respondents' original answers about consumption of five or more drinks on a single occasion in the core alcohol question AL08. However, respondents in 2006 who changed their answer to indicate that they had five or more drinks on 0 days in the past 30 days in response to a consistency check between the 30-day frequency of any alcohol use and the frequency of binge alcohol use in that period were incorrectly asked question CA11 based on their original answer in AL08. The FOURORMOREFLAG variable reflected this consistency resolution between the overall 30-day frequency and the frequency of consumption of five or more drinks, such that these respondents have been correctly skipped out of question CA11 since 2007. Consequently, no data in CABNGAGE were overwritten with a code of 989 (LEGITIMATE SKIP Logically assigned) to indicate that respondents should not have been asked CA11.

In addition, updated routing logic for females for questions CA13 (edited variable CA4FDDYS) and CA14 (edited variable CA4FDAGE) has been in place since 2007 to include FOURORMOREFLAG in addition to questions CA01 (edited variable CADRLAST) and CA12 (edited variable CA4FDEV). This corrected for a logic error in 2006 in which female respondents were routed to CA13 and CA14 only if they answered question CA12 as "yes" or reported in question CA01 (CADRLAST) that they had four or more drinks the last time they drank alcohol in the past 30 days. Those female respondents in 2006 who were skipped out of question CA12 based on the flag variable from the core alcohol module (i.e., ED4FLAG) or who reported consumption of five or more drinks in question CA10 (CABNGEV) were not asked subsequent questions about their consumption of four or more drinks, unless they reported in question CA01 that they last had four or more drinks. Because of this change in 2007, no data in

CA4FDAGE were assigned a code of 990 (NOT ASKED THE QUESTION Logically assigned) due to female respondents being incorrectly skipped out of question CA12. Similarly, a code of 90 (which had the same meaning as that for 990) was not assigned to CA4FDDYS in 2007 if female respondents had used alcohol in the past 30 days.

Table B.27 discusses additional edits associated with the binge alcohol history variables in the consumption of alcohol module. The first set of edits pertains to the consistency checks between the age when respondents reported that they first had five or more drinks (or four or more drinks for females) in a single occasion, and the age at initiation based on the year and month when respondents reported that they first engaged in this behavior.

Table B.27 then presents edits for these binge alcohol variables relative to respondents' ages and to corresponding core alcohol use data. For example, if respondents reported in question CA11 that they first had five or more drinks on a single occasion at an age that was earlier than what they reported for when they first had a drink of any alcohol (from the core alcohol use variable ALCTRY), then CABNGAGE was set to bad data. Any nonmissing values in CABNGYFU and CABNGMFU also were replaced with codes for bad data. Similarly, suppose respondents first used any alcohol at their current age. If they had a lifetime history of consumption of five or more drinks on a single occasion (including any indication of this behavior in the past 30 days) but CA11 had a missing value (e.g., "don't know" or "refused"), then it could be logically inferred in CABNGAGE that the first episode of binge alcohol use occurred at the respondents' current age.

For female respondents, the data pertaining to consumption of five or more drinks on a single occasion and consumption of four or more drinks on a single occasion also were edited for consistency with one another. For example, if CABNGAGE (from question CA11) indicated consumption of five or more drinks at an age that was earlier than what respondents reported in CA14, then CA4FDAGE was assigned the earlier age from CABNGAGE.

In addition, "editing indicator" (EI) variables were created to indicate when nonmissing values were logically assigned to the binge alcohol variables in the consumption of alcohol module. The default value for these EI variables was 1 (i.e., questionnaire data). The EI variables were assigned a value of 2 (Logically assigned data) if a logical inference was made that involved assignment of a nonmissing value. In the above example, if CA4FDAGE was assigned an earlier age from CABNGAGE, then the corresponding variable EI4FDAGE was assigned a value of 2.

In keeping with the general principle of not using noncore data to edit core data, no editing was done to core data when a respondent's reported first consumption of five or more drinks on a single occasion (or four or more drinks for females) suggested more recent use of alcohol than was indicated in the core recency variable ALCREC. For example, ALCREC may not indicate use in the past month or past year, but respondents could report that they first had five or more drinks or four or more drinks on a single occasion at their current age. Although these data from the consumption of alcohol would suggest that these respondents used alcohol in the past 12 months, no editing was done to ALCREC based on these data.

8. Editing and Imputation for the NSDUH Roster Variables

8.1 Introduction

This chapter describes the techniques used to edit and impute variables associated with the household roster for the 2014 National Survey on Drug Use and Health (NSDUH). The variables described in this chapter can be divided into three groups:

- respondent-level detailed roster variables,
- roster-derived household composition variables, and
- proxy variables.

The respondent-level detailed roster variables included the age, gender, and relationship to the respondent for each household member. The introductory question for the household roster portion of the questionnaire (QD54) was interviewer administered. This question asked the respondent how many individuals lived in the household. The computer-assisted interviewing (CAI) instrument was set up to be able to collect data on up to 25 household members. If only one person lived in the household or the respondent did not know or refused to answer, then the household composition (roster) section was skipped. Otherwise, the respondent was asked questions about the age, gender, and relationship to the respondent of every member of the household, starting with the household's oldest member and including the respondent.

If a pair of respondents was selected in a household, the interviewer indicated which member of a respondent's household roster corresponded to the other selected pair member. The roster entry for the respondent was referred to as the "self" entry. In effect, the respondent completed a grid with the number of rows corresponding to the value entered in QD54. Table 8.1 shows an example grid where the number of individuals in the household is four. In this example, the roster of the respondent is shown, and the indicator variable shows that the respondent's son was selected as the other pair member. The possible relationship codes and specific relationship details between pair members are listed in Table 8.2.

Table 8.1 Roster Grid Example Where Number of Individuals in Household (QD54) Equals 4

Individual #	Relationship to Respondent	Age in Years	Other Member Selected for Pair[1]
1	Self	44	0 (No [Impossible])
2	Husband	42	0 (No)
3	Son	16	1 (Yes)
4	Boarder/Roomer	16	0 (No)

[1] This indicator variable applied only to respondents who were part of a pair selection. The other member selected could not have been the self because respondents were not interviewed twice. The other member selected was the roster member who had a value of "1" for this variable.

Table 8.2 Roster Relationship Codes

Relationship Code #	Relationship to Respondent	Details about Relationship
1	Self	
2	Parent	Biological, Step, Adoptive, or Foster
3	Child	Biological, Step, Adoptive, or Foster
4	Sibling	Full, Half, Step, Adoptive, or Foster
5	Spouse	
6	Unmarried Partner	
7	Housemate or Roommate	
8	Child-in-Law	
9	Grandchild	
10	Parent-in-Law	
11	Grandparent	
12	Boarder or Roomer	
13	Other Relative	
14	Other Nonrelative	

The second group of variables, the roster-derived household composition variables, was derived from the respondent-level detailed roster variables. These were mostly count variables reporting the number of individuals in the roster with various characteristics. These variables are listed in Table 8.3, which also shows that some of these variables underwent imputation. Among the three groups of variables described in this chapter, only the roster-derived household composition variables underwent imputation of any kind.

Table 8.3 Roster-Derived Household Composition Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Total number of rostered individuals	TOTPEOP	IRHHSIZE
Number of individuals in household aged 17 or younger	KID17	IRKID17
Number of individuals in household aged 65 or older	HH65	IRHH65
Indicator of whether the respondent had family members in household	FAMSKIP	IRFAMSKP
Number of respondent's family members in household (includes foster relationships)	FMLYSIZE	IRFMLYSZ
Number of respondent's family members in household aged 17 or younger (includes foster relationships)	KIDFMLY	IRKDFMLY
Number of respondent's family members in household (excludes foster relationships)	FAMSIZE	IRFAMSZE
Number of respondent's family members in household aged 17 or younger (excludes foster relationships)	KIDFAMSZ	IRKIDFAM
Number of respondent's children in household aged 2 or younger	NRBABIES	N/A
Number of respondent's children in household aged 3 to 5 years old	NRPRESCH	N/A

Table 8.3 Roster-Derived Household Composition Variables (continued)

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of respondent's children in household aged 6 to 11 years old	NRYUNGCH	N/A
Number of respondent's children in household aged 12 to 17 years old	NRTEENS	N/A
Number of respondent's children in household aged 17 or younger	NRCH0_17	N/A
Number of respondent's children in household aged 18 to 20 years old	NROLDRCH	N/A
Number of respondent's children in household aged 21 or older	NROLDCH	N/A
Number of roommates/housemates in household	NROOMATE	N/A
Indicator of presence of mother in household (12- to 17-year-olds)[1]	IMOTHER	N/A
Indicator of presence of father in household (12- to 17-year-olds)[1]	IFATHER	N/A
Indicator of presence of foster child in household (12- to 14-year-olds)[2]	FSTRCHLD	N/A

[1] The IMOTHER and IFATHER indicators were not 0/1 indicators because levels were provided for "unknown" and "18 or older."

[2] This variable was required for the creation of the POVERTY variable for the 2003-2005 survey years.

The third group of variables, the proxy variables, allowed for the selection and identification of a relative of the respondent who lived in the respondent's household (according to the household roster), was aged 18 years or older and answered the health insurance coverage and income questions for the respondent. In 2014, QP03a was added to the questionnaire to give respondents the opportunity to select another proxy if the previously selected proxy was unavailable. The edited versions of these variables and the questions to which they map are shown in Table 8.4.

Table 8.4 Proxy Variables

Unedited Variable	Text of Survey Question Associated with Unedited Variable	Edited Variable
QP01	Is there anyone else who lives here who is 18 or older who would be better able to give me the correct information about your health insurance coverage and the kinds of income you receive?	PRXABLE2
QP02	Who is the person you think can help us get the correct information for these questions?	PRXRELAT
QP03	Is your [QP02 fill] available right now?	PRXHOME2
QP04	Would you ask your [QP02 fill] to join us to help with these last questions about health insurance and income?	PRXJOIN2
HASJOIN	Has the person's [QP02 fill] joined R?	PRXYANS2
QP03a	Is there any other adult family member available who might be able to answer these questions?	PRXRETRY

8.2 Editing the Respondent-Level Detailed Roster Variables

This section describes the methods used to create edited versions of the respondent-level detailed roster variables: ROSAGE1-ROSAGE25 (roster age), ROSSEX1-ROSSEX25 (roster gender), ROSRLT1-ROSRLT25 (relationship to respondent), ROSMSL1-ROSMSL25 (0/1 indicator: other member selected, pair members only), PRNTYP1-PRNTYP25 (type of parent: biological, adoptive, etc.), SIBTYP1-SIBTYP25 (type of sibling: biological, adoptive, etc.), CHDTYP1-CHDTYP25 (type of child: biological, adoptive, etc.), and TWNTYP1-TWNTYP25 (type of twin: identical, fraternal, or neither). These variables describe up to 25 members of the household. The editing procedures for the respondent-level detailed roster variables began with consistency checks included in the Blaise program code, which were implemented to reduce the amount of editing required at the data processing stage. The consistency checks in the questionnaire were supplemented with other edits involving the respondent-level detailed roster variables outside the CAI instrument. These involved resolving cases where it was unclear which roster member was the self and cases where relationship codes were impossible (or very unlikely) given the age and gender in relation to the self.

Section 8.2.1 describes the consistency checks programmed into the questionnaire. Section 8.2.2 describes the creation of a roster-level dataset for further processing. Section 8.2.3 describes roster edits involving the self. Section 8.2.4 describes roster edits for other household members, after the self has been established. Finally, Section 8.2.5 describes the creation of the final edited respondent-level detailed roster variables.

8.2.1 Roster Consistency Checks

Two types of consistency checks were employed in the CAI instrument for the household roster section of the questionnaire. These checks (1) compared the roster entry corresponding to the respondent with previously entered questionnaire information and (2) compared a roster entry against other roster entries for internal consistency. With the exception of the check against the previously entered respondent's gender, the interviewer could override the consistency checks and explain why the response given was correct. Interviewers' explanations for overrides to consistency checks and evaluations of their validity are provided in Appendix G.

8.2.1.1 Comparisons with Previously Entered Questionnaire Information

Gender and age were the two consistency checks built into the household roster section of the CAI instrument that compared the roster entry with the previously entered questionnaire information. The check for gender was added in 2001 and was triggered if the respondent in the household roster entered a gender that was different from the one previously recorded in the interview (question QD01). If the gender did not match, the interviewer was required to change either the roster entry or the gender that had been entered at the beginning of the interview.

The consistency check for age was added in 2002 to compare the respondent's age in the roster with the age previously entered in the questionnaire (the Blaise variable CURNTAGE). The interviewer could either change the respondent's age entered in the roster or override the consistency check. If the interviewer chose to override the consistency check, then he or she provided an explanation as to why the roster age did not match CURNTAGE. Explanations given by the interviewer for overriding this particular consistency check were carefully reviewed. Interviewers' explanations for overrides to consistency checks and evaluations of their validity are provided in Appendix G. In rare cases, the final value for age (AGE) was set to the age of the self in the household roster (the "roster age") based on these explanations as well as other evidence. Additional details about how roster age was used for creating AGE are described in Chapter 4. Strategies for the more common situation, where the original value for AGE was not set to the roster age, are discussed in Section 8.2.3.

8.2.1.2 Internal Consistency Checks

Since the 2002 survey, internal consistency checks have been implemented in the CAI instrument to compare one roster entry with another. These checks were triggered if any of the following conditions occurred:

1. The interviewer reported that the respondent had more than one spouse or unmarried partner or reported a spouse and an unmarried partner.
2. The interviewer reported that a household member was a parent or grandparent of the respondent and the respondent was older than the household member.
3. The interviewer reported that a household member was a child or grandchild of the respondent and the respondent was younger than the household member.
4. The interviewer reported that a household member was a spouse or an unmarried partner of the respondent and the household member was 16 years old or younger.
5. The interviewer reported that the respondent had a spouse or unmarried partner and the respondent was 16 years old or younger.
6. The interviewer reported that the respondent was either a child-in-law or a parent-in-law and the respondent was 16 years old or younger.
7. The interviewer reported that a household member was a child-in-law of the respondent and the household member was the same age or older than the respondent.
8. The interviewer reported that a household member was a parent-in-law of the respondent and the household member was the same age or younger than the respondent.
9. The interviewer reported that a household member was a biological parent of the respondent and the household member was less than 13 years older than the respondent.
10. The interviewer reported that a household member was a biological child of the respondent and the household member was less than 13 years younger than the respondent.
11. The interviewer reported that a household member was a biological sibling of the respondent and the household member was more than 24 years older or younger than the respondent.
12. The interviewer reported that a household member was a grandparent or grandchild of the respondent and the age difference was less than 30 years.

13. The interviewer reported that only one member of a pair was available for interviewing, potentially due to a change in household size.

In most cases, if a consistency check was triggered, the interviewer changed either an age code or a relationship code in the roster to a more appropriate value. Any edit that was invoked because of an override to a consistency check was carefully scrutinized during the data processing stage. The relevant household roster, as well as the explanation given by the interviewer for the override, was carefully examined to determine whether the override was legitimate. If the override was deemed legitimate (e.g., a father marries a woman, listed as [step] mother, who is younger than the respondent), the original answer was allowed to remain and no edit was applied. If the interviewer's explanation was not considered legitimate, then an edit was applied. More details about roster edits are provided in Section 8.2.4. Explanations given by the interviewers for the overrides that were not considered legitimate are provided in Appendix G.

8.2.2 Roster-Level Dataset

To facilitate processing of the roster variables, a roster-level dataset was created in which the number of records per respondent was given by the household size in question QD54. For example, if a respondent indicated a household of three consisting of himself or herself, a mother, and a father, then there would be three records on the dataset associated with this respondent: one for the self, one for the mother, and one for the father. Even if the respondent did not start or complete the roster questions, records were created for each household member.

8.2.3 Roster Edits Involving the Self

The Blaise program code required the interviewer to identify exactly one self and a corresponding age and gender in the household roster. Moreover, the interviewer was required to confirm with the respondent that the respondent was in fact the identified self. Because the check involving gender was not allowed to be overridden, the gender for the self in the roster always matched QD01, which was equivalent to IRSEX (see Chapter 4). However, it was possible to have problems matching AGE (see Chapter 4 for a description of the methodology used to create AGE) with the age of the self in the roster, despite the consistency check comparing the respondent's roster age against CURNTAGE.

The interviewer was able to override the consistency check for age of the self for one of two reasons: (1) the self was misidentified and another roster member was the true self but the interviewer decided not to change the entries; or (2) the interviewer correctly identified the self but indicated that the correct age for the respondent was different than CURNTAGE, and other evidence did not support this claim (AGE was not set to the roster age, as discussed in Section 8.2.1.1). In the case of a misidentified self, a second roster member in the household was selected whose gender matched IRSEX and whose age was within 1 year of AGE. The second roster member who replaced the original self had a gender and age that matched IRSEX and AGE, respectively.

If the consistency check was overridden, a misidentified self was diagnosed if (1) the roster age of the self differed from AGE by more than 1 year and (2) another roster member of the same gender as QD01 (and IRSEX) had a roster age within 1 year of AGE.[134] If a misidentified self was diagnosed, it was assumed that the interviewer used the roster member identified as the self, rather than the respondent, as the point of reference. Using the example shown in Table 8.1, if the respondent's son was used as the reference point, the relationship for the respondent became "mother" instead of "self," and the "husband" became "father." Under these circumstances, the self code was set to missing, and the respondent's roster entries did not include a self. The remaining relationship codes in the roster also were set to missing. In some cases, the original relationship codes were salvaged, depending upon the roster member who was used as a reference point.

8.2.3.1 Original Self Misidentified: Identifying the Real Self

If the self was misidentified in the roster, an attempt was made to identify a self among the roster members corresponding to the respondent. A roster member was selected as the self under one of two possible circumstances: (1) the roster member's age, gender, and relationship data were missing; or (2) the roster member was of the respondent's gender and was within 1 year of the respondent in age. If more than one roster member met the above criteria, the roster members who met the criteria but were not assigned the self code, were given a bad data code; that is, the original relationship code would no longer make sense because the person of reference in the household had been changed.

8.2.3.2 Original Self Misidentified: Salvaging Relationship Codes

As stated earlier, if the self was misidentified, all other relationship codes were set to missing because the reference person was someone other than the respondent. In some cases, however, the original relationship codes were salvaged, depending upon the roster member who was used as a reference point. Relationship codes were salvaged under the following circumstances:

1. If the reference person was the respondent's sibling, the roster member listed as "self" was actually a sibling, and all other relationship codes were salvaged. (Generally, relationships between the respondent and other household members would be the same with a sibling. For example, the respondent's parents are also the respondent's sibling's parents.)
2. If the reference person was the respondent's spouse or unmarried partner, the roster member listed as "self" was actually a spouse or unmarried partner, and the children relationship codes were salvaged.
3. If all the roster members other than the misidentified self were either roommates, boarders, or other nonrelatives, then the reference person was the respondent's roommate, boarder, or other nonrelative. All other relationship codes were salvaged.

[134] A 1-year difference was allowed because the respondent's age might have changed during the interview. In this instance, the values of AGE and CURNTAGE may have differed by 1 year.

8.2.4 Roster Edits for Other Household Members

Relationship codes were edited if the relationship of the roster member to the self was logically impossible based on age and gender. Edits of household roster ages, genders, and/or relationship codes were performed that either changed the reported value to another value or changed the reported value to bad data. It is important to note that in some cases, two members were selected in a household, which greatly increased the ability to edit the roster for those respondents. Some edits were associated with consistency checks, and interviewers' explanations for overrides to these consistency checks were carefully examined to assess the legitimacy of the override as explained in Section 8.2.1. Some edits were "automatic" in the programming code, which meant that the interviewer was assumed to have been incorrect when the override was implemented. These edits were undone if the interviewer's explanation for the override was considered legitimate. In other situations, the default strategy was to assume that the override of the consistency check was correct and, therefore, that the edit was applied only if the interviewer's explanation appeared incorrect.

8.2.4.1 Edits to Roster Age, Gender, and Relationship Codes: Changes to Different Values (Correct Person of Reference)

The following edits were performed on the roster age, gender, and relationship code values when the recorded age, gender, and/or relationship code was either missing or internally inconsistent and replaced by internally consistent values. When typing on a computer keyboard, it was possible for a double-digit age to have been entered as a single-digit age ("5" instead of "55"), or vice versa ("55" instead of "5"). If the relationship code was still believable, even with the incorrectly entered age (e.g., "other relative"), this type of error was difficult to detect because no inconsistency check was triggered. On the other hand, if an age entered this way triggered one of the consistency checks discussed in Section 8.2.1.2, the interviewer had an opportunity to correct the entry error. On those occasions where the age did not trigger a consistency check, detection of the error was still possible among selected pairs by examining the roster entries of the other pair member. For example, if one pair member had an x-year-old and no xx-year-olds, and the other had an xx-year-old and no x-year-old, where x denoted a single-digit number, it was highly probable that an error had occurred. By comparing the number of children younger than 12 years old in each roster with the number of children on the screener roster, it was apparent how a correction should be made. In this instance, the incorrectly entered age was replaced with the value given by the pair member whose roster age and screener age agreed.

1. If two members were selected in a household, the roster age for the other member selected was commonly not the same as the questionnaire-edited age (AGE, defined in Chapter 4) of the other pair member. In this case, the roster age for the other member selected was changed to this questionnaire-edited age value.
2. If two members were selected in a household, the gender that one member selected for the other on the household roster was often not the same as the gender (IRSEX, defined in Chapter 4) reported by that other pair member in his or her interview. In this case, the roster gender was changed to match the gender value the other pair member reported in his or her interview.

3. In previous survey years, the relationship codes for grandchild (9) and grandparent (11) were commonly confused. The following edit, which was used in previous survey years, was maintained in case of overrides: If the age of the respondent was at least 20 years older than that of the roster member, but the roster member was identified as a grandparent, the relationship code was changed to grandchild. Conversely, if the age of the respondent was at least 20 years younger than that of the roster member, but the roster member was identified as a grandchild, then the relationship code was changed to grandparent.

8.2.4.2 Edits to Relationship Codes: Changes to Missing Codes

The following edits were performed on the roster relationship code values, where the relationship code given was internally inconsistent and no internally consistent value could be used to replace it. These edits were performed before the edits listed in Section 8.2.4.1 were completed. For respondents who had changes to their rosters that were due to the edits described below, the changes to age and gender that were due to the edits described in Section 8.2.4.1 were checked to make sure that they did not impact the decision to implement the edits below. The relationship code in these instances was set to a bad data code.

1. More than one roster member aged 15 years or older was listed as the respondent's unmarried partner or as the respondent's spouse. This situation should have been covered by consistency check #1 listed in Section 8.2.1.2.
2. A roster member aged 15 years or older was identified as a spouse and another was identified as an unmarried partner. In this case, the spouse code was maintained and the unmarried partner code was set to bad data. This situation should have been covered by consistency check #1 listed in Section 8.2.1.2.
3. The roster member was the respondent's parent, but was younger than the respondent. This situation should have been covered by consistency check #2 listed in Section 8.2.1.2. This edit would have been automatic for respondents younger than 15 years old.
4. The roster member was the respondent's child, but was older than the respondent. This situation should have been covered by consistency check #3 listed in Section 8.2.1.2. This edit would have been automatic for respondents younger than 15.
5. The roster member was the respondent's biological parent, but was less than 13 years older than the respondent. This situation should have been covered by consistency check #9 listed in Section 8.2.1.2.
6. The roster member was the respondent's biological mother, but was more than 60 years older than the respondent.
7. The roster member was the respondent's biological child, but was less than 13 years younger than the respondent. This situation should have been covered by consistency check #10 listed in Section 8.2.1.2.
8. A respondent had a biological sibling older than a biological parent, where the biological parent was at least 13 years older than the respondent. If this situation occurred, the relationship code of the "sibling" was set to missing. If the age

difference between the biological sibling and the respondent was more than 25 years, then a consistency check was triggered (consistency check #11 listed in Section 8.2.1.2).

9. A respondent had a biological parent younger than a biological sibling, where the biological parent was less than 13 years older than the respondent. If this situation occurred, the relationship code of the "parent" was set to missing. As with the previous edit, this edit was partially covered by consistency checks #9 and #11 listed in Section 8.2.1.2.
10. The roster member was the respondent's child-in-law, but was at least 10 years older than the respondent. This situation should have been covered by consistency check #7 listed in Section 8.2.1.2.
11. The roster member was the respondent's parent-in-law, but was at least 10 years younger than the respondent. This situation should have been covered by consistency check #8 listed in Section 8.2.1.2.
12. The roster member was the respondent's parent-in-law or child-in-law, but either the roster member or the respondent was younger than 15 years old. This situation should have been covered by consistency check #6 listed in Section 8.2.1.2.
13. The respondent had two or more children-in-law, but had no children in the household. The in-law codes were all set to missing.
14. The roster member was the respondent's grandchild, but the respondent or respondent's spouse (if applicable) was 25 years old or younger. This situation should have been covered by consistency check #12 listed in Section 8.2.1.2.
15. The roster member was the respondent's grandchild, but the respondent's parents lived in the household. Also, the respondent had no children in the household and was less than 24 years older than the roster member. As with the previous edit, if the grandchild was in fact older than the respondent, this error should have been covered by consistency check #3 listed in Section 8.2.1.2.
16. The roster member was the respondent's sibling and another roster member was a parent, but the roster member's age was within 4 years of the age of the parent. If the sibling was a half- or step-sibling, an additional requirement was that there was only one parent.
17. The roster member was the respondent's grandparent or grandchild, but the age difference between the respondent or the respondent's spouse (if applicable) and the roster member was less than 20 years. If the roster member was a "grandchild" who was older than the respondent, then this situation was covered by consistency check #3 listed in Section 8.2.1.2. Similarly, if the roster member was a "grandparent" who was younger than the respondent, then this situation was covered by consistency check #2 listed in Section 8.2.1.2. If the age difference was less than 30 years, this was covered by consistency check #12 in Section 8.2.1.2.
18. If the respondent had two parents, but both parents were listed as biological mothers or both parents were listed as biological fathers, the roster genders of both roster members were set to missing.

8.2.4.3 Edits to Relationship Codes: Changes to Different Values (Incorrect Person of Reference: Illogical Child Code)

In Section 8.2.4.2, illogical relationship codes were set to bad data. Often, this occurred because the interviewer used someone other than the respondent as the person of reference for one or more roster members. In some of these cases, the structure of the roster could have been used to determine the appropriate relationship code for that individual. Edits where the illogical code was "child" are listed below.

1. The interviewer might have put a roster member after the respondent's parent in the household roster. If the relationship code for that roster member was given as "child," the relationship code was illogical if the age made it impossible for the roster member to be the respondent's child (see #4 in Section 8.2.4.2). In fact, if more than one "child" was listed after the respondent's parent, each would be listed as illogical. However, it was likely that the interviewer was making the reference to the respondent's parent rather than the respondent. In this case, if the child relationship was not a stepchild and the age difference between the respondent's parent and the "child" was at least 12 years, then the relationship code was changed to sibling.
2. In some cases, the interviewer's entry for a roster member listed as "child" might simply be a typographical error, for example, where the "3" (child) should be a "2" (parent) (see Table 8.2 for all the relationship codes). Interviewers usually corrected such errors when a consistency check was triggered in cases where the child was older than the parent or the child was a biological child who was less than 12 years younger than the parent (Section 8.2.4.2). However, in cases where the interviewer insisted on the code, or where the child was younger than the respondent, but was less than 12 years younger than the respondent and was not biological, these typographical errors were more difficult to detect. If the respondent was living with parent(s) and unmarried and not living with an unmarried partner, and the roster member was not 12 or more years younger than the respondent, then the relationship code was changed to sibling.
3. Both sides in a selected pair[135] were respondents aged 18 or younger, both sides identified parents in the household, and one side had an illogical child code. When the number of illogical child codes was added to the number of siblings on one side, the sum was equal to the number of siblings on the other side. If the age of the roster member was younger than 25 years, then the relationship code was changed to sibling.
4. A roster member was listed as the respondent's child who was not more than 12 years younger than the respondent and the respondent was 25 or younger. The previous roster member was listed as "grandparent." The "child" was in reference to the respondent's grandparent and was considered either the respondent's parent or the respondent's uncle or aunt. If the roster member's age was at least 12 years older than the respondent and there were no nonimmediate family codes (7, 12, 13, or 14 as described in Table 8.2), then no uncles or aunts lived in the household. If a pair was

[135] A selected pair has two rosters where each respondent is from the same household. A "side" refers to one of the two rosters that make up a selected pair.

selected and no nonimmediate family codes were found in either pair member's roster, then in either of these cases the relationship code was set to parent. Otherwise, the relationship code was set to missing.

8.2.4.4 Edits to Relationship Codes: Changes to Different Values (Incorrect Person of Reference: Illogical Spouse Code)

The interviewer also could have used an incorrect person of reference with illogical spouse codes. This error occurred most frequently when a selected child had a parent with a spouse (the other parent) or unmarried partner. Rather than identifying this individual as a "parent" or "other nonrelative," the interviewer identified the roster member as a spouse or unmarried partner of the child, even though the interviewer intended that the point of reference be the child's parent rather than the child. This manifestation of the illogical spouse code, along with others, is described below. Many of these edits were covered by consistency checks #4 and #5 listed in Section 8.2.1.2, provided either the respondent or the roster member was 16 or younger.

1. Both sides in a selected pair identified that they had a spouse or unmarried partner, but the two respondents were not part of a spouse-spouse pair. This legitimately could have occurred only if there were multiple spouse-spouse pairs in the household. In this edit, an attempt was made to identify cases with a single spouse-spouse pair in the household, where one pair member had a correctly identified spouse or unmarried partner and the other pair member had an incorrectly identified spouse or unmarried partner. If the younger respondent, who was 21 years old or younger and at least 10 years younger than the older respondent, indicated a parent, and the older respondent indicated neither parents nor parents-in-law, then the older respondent should be considered either the younger respondent's parent or the parent's spouse or unmarried partner. If the misidentified code was "spouse," then the code was changed to "parent." However, if the misidentified code was "unmarried partner," then the roster member may or may not be considered the parent of the respondent. In most cases where the misidentified unmarried partner was the respondent's parent's unmarried partner, the code was changed to parent. The exception occurred when (1) the unmarried partner of this respondent's parent was the other respondent selected in a pair and (2) the unmarried partner did not indicate that the other pair member selected was his or her child in the parenting experiences question, FIPE3. In this instance, the relationship code was changed to a special code indicating that the roster member was an unmarried partner of the respondent's parent.
2. As in the previous edit, both sides in a selected pair identified a spouse or unmarried partner, but were not part of a spouse-spouse pair, and there was only a single spouse-spouse pair in the household. In this edit, both sides incorrectly identified the spouse or unmarried partner. In most cases, the pair was a sibling-sibling pair. If both respondents were younger than 21, both indicated a parent in the household, and the age difference between the respondents and their respective "spouse or unmarried partner" was unusually large, then on each side the misidentified spouse or unmarried partner should have been considered a spouse or unmarried partner of the respondent's parent. If both misidentified codes were "spouse," then both codes were

changed to "parent." As stated in the previous edit, if both misidentified codes were "unmarried partner," then it was not clear whether each misidentified code should have been "parent." The rules used to determine whether the roster member was the respondent's parent were the same as in edit #1. The same special code as in the previous edit was used to identify an unmarried partner of the respondent's parent. Hence, the incorrectly identified "spouse or unmarried partner" code was changed for each respondent in the pair to either "parent" or the aforementioned special code.

3. In this edit, only one side in a selected pair identified a spouse (not unmarried partner), but the spouse was identified even though (1) the respondent was younger than 15, (2) the spouse was younger than 15 and the other pair member did not have a spouse, or (3) the respondent was younger than 18 but responded that he or she was "never married" in the core part of the questionnaire, and the respondent did not have any parents-in-law in the household. If the respondent listed one parent, but the other pair member listed two parents, then the pair was a sibling-sibling pair and the relationship code was in reference to the parent. If the respondent listed one fewer sibling than the other pair member, then the pair was a sibling-sibling pair and the spouse code was a typographical error (meant to be a sibling, with a code of "4" instead of "5").
4. Only one side in a selected pair identified an unmarried partner, but the unmarried partner was identified even though (1) the respondent was younger than 15 or (2) the unmarried partner was younger than 15. If the respondent listed one parent, but the other pair member listed two parents, then the pair was a sibling-sibling pair and the relationship code was in reference to the parent's unmarried partner. In this case, the relationship code was changed to parent. If the respondent listed one fewer sibling than the other pair member and the age difference between the respondent and the roster member identified as the unmarried partner was less than 15 years, then the pair was a sibling-sibling pair and the unmarried partner code was changed to sibling.
5. Both sides in a pair identified the same household member as spouse or unmarried partner. If the previous roster member on one of the sides was a sibling, then the spouse or unmarried partner should be considered the sibling's spouse or unmarried partner. The spouse or unmarried partner relationship code was changed to bad data. If both sides had a previous roster member who was a sibling, then it was not clear to which pair member the spouse or unmarried partner belonged. To maintain proper counts, the spouse or unmarried partner code for the youngest pair member was changed.
6. A spouse or unmarried partner was identified even though (1) the respondent had one parent in the household who was the roster member listed before the spouse or unmarried partner, (2) the respondent either was younger than 17 years old or was between 17 and 20 years old and the spouse or unmarried partner was older than the respondent's parent, and (3) the respondent was more than 15 years younger than the spouse or unmarried partner. In the case of the misidentified spouse, the "spouse" of the respondent was considered the respondent's other parent. In the case of the misidentified unmarried partner, the "partner" of the respondent was considered the unmarried partner of the respondent's parent. The code was changed to "parent." For a

household member with a spouse code who was aged 16 years or younger, this edit should have been covered by consistency check #4 listed in Section 8.2.1.2.

7. In cases where the respondent was younger than 15 years old, he or she identified a spouse or unmarried partner, and the above edits did not apply, the relationship code was set to bad data. In cases where the roster member was younger than 15, the roster member was identified as a spouse or unmarried partner, and the above edits did not apply, the relationship code and roster member's age were set to bad data. This should have been covered by consistency checks #4 and #5 listed in Section 8.2.1.2.

8.2.4.5 Edits to Relationship Codes: Changes to Different Values (Incorrect Person of Reference: Illogical Sibling Codes)

If the relationship code indicated that one of the other roster members was the respondent's sibling, but the age difference between the sibling and the respondent was at least 20 years, then the sibling relationship code was suspicious. If the previous roster entry was either the respondent's child or another sibling with the same characteristics, and either the respondent did not have parents in the household or the parent was a mother and the age difference between the mother and the sibling was more than 50 years, then the sibling relationship codes were referencing the respondent's children's relationships to each other. The relationship codes were therefore changed to "child." Age differences greater than 25 years among biological siblings would have been covered by consistency check #11 in Section 8.2.1.2.

8.2.4.6 Edits to Relationship Codes: Changes to Different Values (Incorrect Person of Reference: Illogical Grandchild Codes)

If the relationship code indicated that one of the other roster members was the respondent's grandchild, but the respondent was too young to have a grandchild (25 or younger), it was possible that the roster member was a grandchild of a previous roster member. If two young respondents were selected where both identified the same grandparents and the same parents, and the respondent on the other side had siblings, then the grandchild should be considered the respondent's sibling. If this was not established, then the roster member could be the respondent's sibling or the respondent's cousin, and the code was set to bad data. If the grandchild was older than the respondent, then this edit would have been covered by consistency check #3 listed in Section 8.2.1.2. If the age difference between the grandchild and the respondent was less than 30 years, then this edit would have been covered by consistency check #12 listed in Section 8.2.1.2.

8.2.4.7 Edits to Relationship Codes: Changes to Different Values (Incorrect Person of Reference: Illogical In-Law Codes)

In some situations, the in-law code was incorrectly used because the respondent was not using himself or herself as the person of reference. In such cases, either the child-in-law was the child of someone else in the roster other than the respondent or the respondent was referring to himself or herself as the parent-in-law of the roster member. An in-law code was deemed incorrect if a roster member was listed as the respondent's child-in-law who was not more than 12 years younger than the respondent and the respondent was 25 or younger. If the relationship code was listed as child-in-law, and the previous roster member was listed as grandparent, then the child-in-law was in reference to the respondent's grandparent and should have been considered either the respondent's parent or the respondent's uncle or aunt. If the roster member's age was at least 12 years older than the respondent and there were no nonimmediate family codes (7, 12, 13, or 14 as described in Table 8.2), then no uncles or aunts lived in the household. If a pair was selected, no nonimmediate family codes were found in either pair member's roster. In either of these cases, the relationship code was set to parent. Otherwise, no certainty was associated with the relationship code, and this code was set to missing.

8.2.5 Final Edited Respondent-Level Detailed Roster Variables

The unedited roster variables contained information for each roster member: age, gender, relationship to respondent, and a 0/1 variable that indicated whether the roster member was the other member selected in a pair (Table 8.1 provides an example). This information could be captured for up to 25 members of a household. Within the CAI instrument, separate variables were created to collect this information for male and female household members and for household members with ages reported in years as opposed to months. When the edited versions of these variables were created, this information was combined for each household member into four variables, one for each attribute (i.e., age, gender, relationship to respondent, and pair status). The edits listed in Section 8.2 were incorporated into the values of the detailed roster variables: ROSAGE1-ROSAGE25 (roster age), ROSSEX1-ROSSEX25 (roster gender), ROSRLT1-ROSRLT25 (relationship to respondent), and ROSMSL1-ROSMSL25 (0/1 indicator: other member selected, pair members only). Additional variables were also created: PRNTYP1-PRNTYP25 (type of parent: biological, adoptive, etc.), SIBTYP1-SIBTYP25 (type of sibling: biological, adoptive, etc.), CHDTYP1-CHDTYP25 (type of child: biological, adoptive, etc.), and TWNTYP1-TWNTYP25 (type of twin: identical, fraternal, or neither).

Final edited versions of the respondent-level detailed roster variables were used to derive (or, at a minimum, to calculate bounds when data were missing) the household composition variables described in Section 8.3.

8.3 Editing and Imputation Procedures for Roster-Derived Household Composition Variables

8.3.1 Creation of Edited Roster-Derived Household Composition Variables

This section discusses the creation of edited versions of the roster-derived household composition variables. After replacing apparently erroneous information in the roster with missing values, the number of individuals with various characteristics in each roster was determined. These counts were recorded in the edited roster-derived household variables shown in Table 8.3. If any information in the roster was missing, the roster-derived variable was set to missing. However, if some of the roster records for a respondent's household had missing data, then roster records with nonmissing data for that household were used to limit the possible values to which the missing roster-derived variable could have been imputed. Details on the imputation of the roster-derived household variables are provided in Section 8.3.2. If two respondents were selected in a single household as part of a pair, then the information from one pair member was not used to directly edit the household composition variables of the other pair member.

The respondent's household size was assumed to equal the total number of rostered individuals in the household, TOTPEOP, as shown in Table 8.3. The value of TOTPEOP was expected to equal the value of QD54 in most cases. However, in some cases, the original self was misidentified and no other roster members were close to matching the respondent's age and gender. In these cases, an extra roster member was added to correspond to the respondent (the self) so that the value of TOTPEOP was 1 greater than the value of QD54. For other cases, the respondent did not enter a value for QD54, and thus TOTPEOP and all the roster-derived variables were missing. Finally, it was possible that duplicate entries were put into the household roster so that the value of TOTPEOP would be determined by excluding the duplicates from the roster. This latter situation was usually impossible to detect, unless the respondent had two biological fathers or two biological mothers of exactly the same age. In this instance, the extra biological parent of the same gender was dropped from the roster, and the value of TOTPEOP was reduced to 1 less than the value of QD54.

The variables KID17 (number of individuals in the household aged 17 or younger) and HH65 (number of individuals in the household aged 65 or older) were simple counts based on the roster ages and did not account for the relationships of the individuals to the respondent. If some of the roster members had missing ages, the values of KID17 and HH65 also were missing, regardless of whether some of the roster members were eligible to be part of the count. In these instances, the imputed values for KID17 and HH65 were restricted based on the nonmissing information available in the roster, as explained in Section 8.3.2.2. However, if the roster member was missing a relationship code, but not an age, then that roster member was still eligible to be included in these variables.

The variable FAMSKIP was an indicator of whether the respondent's household contained any other family members. It was created based on the relationship codes of the roster members. If one or more of the roster members had a missing relationship code, and no other family members were in the respondent's household, then the value of FAMSKIP was set to missing. However, if one of the nonmissing roster member's relationship codes indicated that the household contained one of the respondent's family members, then the value of FAMSKIP was not missing, even if other roster members had missing relationship codes.

The variables FMLYSIZE (number of respondent's family members in the household, including foster relationships), FAMSIZE (number of respondent's family members in the household, excluding foster relationships), KIDFMLY (number of respondent's family members in the household aged 17 or younger, including foster relationships), and KIDFAMSZ (number of respondent's family members in the household aged 17 or younger, excluding foster relationships) were simple counts based on the relationships of the individuals to the respondent and the ages in the respondent's household roster. FMLYSIZE and KIDFMLY were created to determine appropriate measures of poverty levels, using Federal poverty definitions starting in 2006. FAMSIZE and KIDFAMSZ were used in the 2003-2005 surveys. The definition of "family" for FAMSIZE and KIDFAMSZ was slightly different from that used for other roster variables; foster relationships were not considered family relationships. If some of the roster members had missing ages or missing relationship codes, the values of FMLYSIZE, FAMSIZE, KIDFMLY, and KIDFAMSZ were set to missing, even though some of the roster members might have been eligible to be part of the count. In these instances, the imputed values were restricted based on the nonmissing information available in the roster, as explained in Section 8.3.2.2.

Eleven other roster-derived variables were created that used both the age and relationship codes of the roster members. All of the roster-derived variables and their definitions are summarized in Table 8.3. Except for FAMSKIP, each of these variables was missing if the age or relationship codes for at least one roster member in a respondent's household were missing. FAMSKIP could be coded despite missing values if there was at least one nonmissing family relationship code in the roster. Edited versions of the roster-derived variables were also used in the editing procedures applied to the creation of the edited proxy variables, described in Section 8.4.

8.3.2 Imputation for Roster-Derived Household Composition Variables

Of the three groups of edited roster variables described in the introduction to this chapter, the only group that underwent any imputation at all was a subset of the roster-derived household composition variables. Each of the eight variables in this subset formed its own single-member imputation set[136] and tended to have item response rates of more than 99 percent. Table A.28 in Appendix A has details on the rates of missingness for the variables that were imputed. The single RP/single PRD type of PMN, described in Section 3.4.1, was used to impute nonmissing values among these eight variables in the order shown in Table 8.3. The order was important, as imputation-revised variables from earlier in the sequence were frequently used to assist with imputation-revised variables later in the sequence.

Section 8.3.2.1 describes the imputation process applied to the first edited variable, TOTPEOP. Section 8.3.2.2 summarizes the imputation processes applied to the other seven variables. Since the processes applied to the other seven variables are very similar to the process applied to TOTPEOP, Section 8.3.2.2 will only list divergences from the process that was applied to the TOTPEOP variable.

8.3.2.1 Imputation for TOTPEOP (Imputation Set 1)

The first imputation set included a single variable, TOTPEOP. The analogous imputation-revised variable IRHHSIZE was created using the single RP/single PRD type. There were no noteworthy deviations from this general approach. Section 8.3.2.1.1 describes the RP step, Section 8.3.2.1.2 describes the PRD step, and Section 8.3.2.1.3 describes the hot-deck step. As is true for all the roster-derived household composition variables that underwent imputation, the item response rate was very high (more than 99 percent).

8.3.2.1.1 Response Propensity Step

The response propensity model for imputation set 1 utilized the preliminary analysis weight, PANALWT. All respondents were in the domain for the TOTPEOP variable. A domain member was considered an item respondent if and only if TOTPEOP was nonmissing. See

[136] An imputation set is a set of variables for which a single donor is used in the final hot-deck step. Chapter 3 describes this concept more fully.

Tables D.49, D.50, D.51, and D.52 in Appendix D for details of the covariates used in the RP models for this variable.

8.3.2.1.2 Prediction Step

TOTPEOP was a count variable. It was assumed to have a Poisson distribution, and the parameters for the models were estimated using the adjusted weights that are outputs of the RP step and using Poisson regression as implemented by the LOGLINK procedure in SUDAAN software.[137] The single predicted mean used in the subsequent hot-deck step was the predicted number of people in the household.

8.3.2.1.3 Hot-Deck Step

The hot-deck step for the TOTPEOP variable was the simplest one used in the NSDUH. There were no logical constraints, and the only likeness constraint was the delta constraint. The predictive mean vector was actually a scalar. Every item nonrespondent was handled on the first attempt to find a donor. Additional details on the hot-deck step for TOTPEOP are available in Appendix E.

8.3.2.2 Imputation for Other Roster-Derived Household Composition Variables That Underwent Imputation (Imputation Sets 2 through 8)

Like TOTPEOP, the remaining seven roster-derived household composition variables that underwent imputation (from Table 8.3: KID17, HH65, FAMSKIP, FMLYSIZE, KIDFMLY, FAMSIZE, and KIDFAMSZ) were handled separately using the single model type of PMN and utilized the preliminary analysis weight, PANALWT, in the RP step. The methods were very similar, with only a few exceptions, as follows:

- FAMSKIP was a dichotomous variable, not a count variable. Therefore, its PRD model was a logistic regression model as implemented by the RLOGIST procedure in SUDAAN. The single predicted mean used in the later hot-deck step was the predicted probability that the respondent did not have any other family members in his or her household.
- Bounds were determined for every other variable except FAMSKIP. These bounds were based both on nonmissing information in the roster and on previously imputed variables. For each of these variables, a single logical constraint was used in the hot-deck step, which required the donor to have a value within the bounds.
- Previously imputed roster-derived household composition variables were frequently used in likeness constraints in the hot-deck steps.

Tables D.49, D.50, D.51, and D.52 in Appendix D provide details of the covariates used in the RP models for these variables.

[137] Details about the LOGLINK procedure and additional references are provided in the *SUDAAN® Language Manual, Release 11.0.1* (RTI International, 2013).

8.4 Editing the Proxy Variables

This section describes the creation of edited proxy variables, as listed in Table 8.4. Section 8.4.1 describes the creation of an indicator variable, EDFAM18, which was used to determine skip patterns and missing codes for the five edited proxy variables. Sections 8.4.2 and 8.4.3 describe the editing processes for each value of EDFAM18.

All survey respondents were allowed to choose someone from the household to be their proxy as long as the following conditions were met:

1. There was more than one person in the household.
2. The eligible person was a relative (not a boarder, roommate, or some other nonrelative).
3. The eligible person was aged 18 or older.

Table 8.4 shows the correspondence between the six questionnaire items in the proxy section of the questionnaire and the corresponding edited variables. Except for QP02 and its edited variable PRXRELAT, the valid questionnaire responses were "1 = Yes" and "2 = No." QP02 and PRXRELAT had multiple responses ranging from 1 to 21, with each level representing the relationship of the proxy to the respondent. The levels of PRXRELAT are shown in Table 8.5. Depending on household size and composition, QP02 gave respondents a maximum of nine possible proxy choices out of the 21 relationship types detailed by PRXRELAT. In addition to the nine possible answer choices presented by QP02 based on reported household size and composition, respondents were also given the final choice of "other adult relative" to select a proxy who is related to the respondent but may not live in the household.

Table 8.5 Levels of PRXRELAT

PRXRELAT	Relationship of Proxy Member	Gender of Proxy Member
1 = Father	Parent	Male
2 = Mother	Parent	Female
3 = Son	Child	Male
4 = Daughter	Child	Female
5 = Brother	Sibling	Male
6 = Sister	Sibling	Female
7 = Husband	Spouse	Male
8 = Wife	Spouse	Female
9 = Male Unmarried Partner	Unmarried partner	Male
10 = Female Unmarried Partner	Unmarried partner	Female
11 = Son-in-law	Child-in-law	Male
12 = Daughter-in-law	Child-in-law	Female
13 = Grandson	Grandchild	Male
14 = Granddaughter	Grandchild	Female

Table 8.5 Levels of PRXRELAT (continued)

PRXRELAT	Relationship of Proxy Member	Gender of Proxy Member
15 = Father-in-law	Parent-in-law	Male
16 = Mother-in-law	Parent-in-law	Female
17 = Grandfather	Grandparent	Male
18 = Grandmother	Grandparent	Female
19 = Other Male Relative	Other relative	Male
20 = Other Female Relative	Other relative	Female
21 = Other Adult Relative	Other relative	Male or Female

8.4.1 Edited Indicator of Potential Proxies in Household (EDFAM18)

As described in Section 8.3.1, a binary variable (FAMSKIP) was created that indicated whether the respondent's household roster included other family members. If the presence or absence of other family members was unknown because of a missing household size or missing values in the roster, FAMSKIP could not be determined. A similar variable was created to identify households where the respondent's household roster included other family members aged 18 years or older ("adult" family members), any one of whom could potentially serve as a proxy for the respondent. The edited indicator was called EDFAM18, where "1" indicated that no potential proxy existed in the respondent's household, "0" indicated otherwise, and "98" indicated unknown.

8.4.2 Editing the Proxy Variables when EDFAM18 = 1

In most cases, a value of EDFAM18 = 1 implied that the respondent was skipped out of the proxy questions because no potential proxy existed in the household. In these cases, all of the proxy variables were given a legitimate skip code (99). Two situations could occur, however, where adult family members were incorrectly identified in the household roster: (1) the respondent had not identified any adult family members in the household but had nonfamily members in the household whose ages were not known; and (2) the unedited household roster indicated that one potential proxy existed in the household but editing changed the age of this single potential proxy to younger than 18. In these cases, the respondent was allowed to answer the proxy questions even though the value of EDFAM18 was 1 (i.e., the final edited household roster indicated that no potential proxy existed in his or her household). Moreover, in these situations, the interviewer indicated that none of these household members who were incorrectly identified as adult family members were proxies. However, the "no" value in the first unedited proxy variable (QP01) was replaced by a logically assigned legitimate skip (89) in the corresponding edited variable (PRXABLE2). For cases where PRXABLE2 was set to 89, all of the edited proxy variables corresponding to the unedited proxy variables, which followed QP01, were given legitimate skip codes (99). These were cases in which the respondent answered the proxy module (questions about a proxy) but the interviewer indicated that they were not proxies, so no proxy should have actually answered the health insurance and income modules.

8.4.3 Editing the Proxy Variables when EDFAM18 = 0

If EDFAM18 was 0, the proxy variables were edited as follows:

1. If the unedited proxy variables had legitimate nonmissing values (i.e., not replaced by a logically assigned legitimate skip), the edited proxy variables (except PRXRELAT and QP03A) were set to those nonmissing values.
2. If any of the unedited proxy variables (except PRXRELAT and QP03A) had a value of 2 ("no"), then all of the variables that followed were edited to legitimate skips.
3. If any of the unedited proxy variables (except QP03A) had a value of "don't know" or "refused," then the corresponding edited variable and all the edited variables that followed were given a "don't know" or "refused" code (94 or 97).
4. If any of the unedited proxy variables did not have a value and a legitimate skip code could not be applied, then the corresponding edited variable and all the variables that followed were given a "no answer" code (98).

8.4.4 Additional Editing for PRXRELAT

In addition to these, more detailed rules were used to assign values to PRXRELAT. The value of QP02, which identified the proxy for the respondent, was chosen directly from the respondent's household roster with one caveat. A list of adult family members (a proxy roster) was shown to the respondent, and the respondent was asked to select the family member who could best answer the health insurance and income modules. In the cases where the proxy roster included a large number, only the first nine adult family members were listed. Regardless of roster size and composition, respondents were given the final choice to select an "other adult relative" as a proxy, even if that person was not part of the household roster. Prior to 2014, this response choice was "other adult/relative"; it was changed in 2014 to require the selected proxy to have a familial relation to the respondent. Once the proxy roster was established, the number selected in QP02 was matched to the corresponding person in the proxy roster.

8.4.5 Additional Editing for QP03A

Item QP03A allowed respondents to select another proxy from the list of potential proxies in QP02 if the previously selected proxy was not available to join the interview. This question was added in 2014. Respondents were redirected to question QP03A if they answered "no" to any of the unedited proxy variables; otherwise, QP03A was edited to a skip code. The edited variable of QP03A is PRXRETRY.

This page intentionally left blank

9. Editing and Imputation for the NSDUH Income Variables

9.1 Introduction

This chapter describes procedures for editing and imputing the income variables in the 2014 National Survey on Drug Use and Health (NSDUH) interview. Unlike other sections of the interview that asked about behaviors or situations in the 12-month period prior to the interview, the reference period for the income module on the 2014 NSDUH was the previous calendar year (i.e., January 1, 2013, through December 31, 2013). Respondents (or other household members serving as proxies[138]) were asked whether they (or other family members living in the dwelling unit, if applicable) received income or benefits from specific sources, such as social security or wages earned at a job or business, during the previous calendar year. These questions are subsequently referred to as "source-of-income" questions in this chapter. In addition, the income section of the interview included questions about the dollar amounts of total personal income and total family income (if applicable) during the previous calendar year.

As had been the case in the 2008-2013 surveys, separate questions to ascertain personal-level and other-family-level responses for binary (i.e., yes/no) source-of-income variables were not asked in the 2014 survey, nor were there separate questions about income from child support, interest/investment income, and other income. However, since 2008, respondents have been asked questions about binary (i.e., $20,000 or more or less than $20,000) and finer categories of total annual income at both the personal and family levels. A comparison between the 1999-2007 and 2008-2014 sets of income questions is shown in Table 9.1. See Section 3.4 of the 2008 imputation report (Ault et al., 2010) for a more detailed explanation of the changes to the income questions over the years.

Consistent with most of the imputation-revised variables that were discussed in the previous chapters of this report, imputations for the income variables in the 2014 NSDUH were performed using the predictive mean neighborhood (PMN) methodology detailed in Chapter 3. The edits that were applied to the income variables prior to imputation are described in Section 9.2.

Table 9.1 Comparison between 1999-2007 and 2008-2014 Sets of Income Questions

Income Questions Included in NSDUH	1999-2007 Surveys*			2008-2014 Surveys	
	Personal Level	Other Family Member Level	Family Level	Personal Level	Family Level
Social Security	Yes	Yes	No	No	Yes
Supplemental Security Income	Yes	Yes	No	No	Yes

[138] Reference is made only to "respondents" in the remainder of this chapter. However, readers are advised that the income information for a respondent may have been provided by another adult household member who was serving as a proxy for the respondent because the proxy was considered to be better able to answer the income questions for the respondent. See Chapter 8 for more information about proxy variables.

Table 9.1 Comparison between 1999-2007 and 2008-2014 Sets of Income Questions (continued)

Income Questions Included in NSDUH	1999-2007 Surveys* Personal Level	1999-2007 Surveys* Other Family Member Level	1999-2007 Surveys* Family Level	2008-2014 Surveys Personal Level	2008-2014 Surveys Family Level
Welfare Payments	Yes	Yes	No	No	Yes
Other Welfare Services	Yes	Yes	No	No	Yes
Investment Income	Yes	Yes	No	No	No
Child Support Payments	Yes	Yes	No	No	No
Wages	Yes	Yes	No	No	Yes
Other Income	Yes	Yes	No	No	No
Food Stamps	No	No	Yes	No	Yes
Months on Welfare	No	No	Yes	No	Yes
Binary Total Income	Yes	No	Yes	Yes	Yes
Finer Category Total Income	Yes	No	Yes	Yes	Yes

* In the 2006-2007 surveys, 5 percent of the selected sample was asked a reduced set of income questions, which are identical to the set of questions asked in the 2008-2014 surveys. For detailed explanations of these subsamples, consult the 2006 and 2007 Methodological Resource Books.

9.2 Editing for Binary and Finer Category Income Variables

9.2.1 Binary Income Variables

As stated previously, the source-of-income variables recorded binary (i.e., yes/no) responses. Information for most of these variables was captured through single questions about a given source of income. The exception to this single-question format was for measurement of the number of months on welfare (questions QI12AN and QI12BN and the edited variable WELMOS). Question QI12AN was asked if respondents did not report any family members' receipt of food stamps in question QI07N; otherwise, question QI12BN was asked. In asking about the number of months on welfare, question QI12BN included the additional phrase, "not including food stamps."

Creation of WELMOS also was dependent on respondents' answers in the questions corresponding to the relevant edited variables for family members' receipt of cash assistance from a state or county welfare program (FAMPMT) and receipt of other (noncash) welfare or public assistance (FAMSVC). If FAMPMT or FAMSVC indicated that respondents received welfare payments, WELMOS was assigned the values from questions QI12AN and QI12BN. Otherwise, if family members in the household (including the respondent) definitely did not receive welfare payments or noncash benefits (e.g., job training, help with child care), WELMOS was assigned a legitimate skip code.

In addition to the source-of-income variables, binary variables were created from a pair of questions that asked whether the respondent's personal total income or the respondent's total family income was $20,000 or more. For this pair of questions (QI20N and QI22, corresponding to edited variables PINC1 and FINC1), the second question in the pair applied to the entire family. For answering these questions about total personal or family income, respondents were shown a list of other sources of income that they had not been previously asked about (e.g., child support payments). Respondents were instructed in answering these questions about total income to consider both these other sources of income and the sources of income they had been asked about previously.

If a respondent had no other family members living in his or her dwelling unit, FINC1 was assigned a legitimate skip code (Section 2.4.2). QI22 also was skipped and FINC1 was assigned a legitimate skip code if respondents reported in QI20N that they had personal incomes of $20,000 or more. A third binary total family income variable, FAMINC1, was created and was equal to either PINC1 or FINC1, depending on whether other family members were present in the household.

9.2.2 Finer Category Income Variables

Respondents also were asked to identify, both for themselves and for their families, finer categories of income within the two general categories of "$20,000 or more" or "Less than $20,000." If respondents answered binary total income questions as "Less than $20,000," they were asked to report a finer category of income from incomes less than $1,000 (including a net loss of income) up to $20,000 in increments of $1,000. Similarly, respondents who answered binary total income questions as "$20,000 or more" were asked to report a finer category of income from $20,000 up to $50,000 in increments of $5,000. If respondents' or their families' incomes were greater than $50,000, they could select the following additional categories: "$50,000 to $74,999," "$75,000 to $99,999," or "$100,000 or more." (Prior to 2004, the maximum income category was "$75,000 or more.")

The variable PINC2 for finer categories of personal income was created from responses to questions QI21A (for personal incomes of less than $20,000) and QI21B (for personal incomes of $20,000 or more). The variable FINC2 for finer categories of family income was created from questions QI23A (for family incomes of less than $20,000) and QI23B (for family incomes of $20,000 or more). Income categories in PINC2 and FINC2 ranged from 1 (less than $1,000 [including loss]) to 29 ($100,000 or more).

Questions QI23A or QI23B were skipped if the respondent had no other family members in the household. Therefore, when the imputed household roster variable IRFAMSKP indicated that no other family members were living in the household, FINC2 was assigned a legitimate skip code of 99; this included situations in which data existed in questions QI23A or QI23B but IRFAMSKP subsequently indicated that no other family members lived in the household.

A third total family income variable with finer categories of income, FAMINC2, also was created. If other family members were living in the household, FAMINC2 was set to be equal to the response from the mutually exclusive pair of total family income questions QI23A and QI23B. If no other family members lived in the household, FAMINC2 was set to be equal to the responses from questions QI21A or QI21B. Thus, unlike FINC2, FAMINC2 was not assigned legitimate skip codes if no other family members lived in the household. Since 2010, respondents who reported in question QI21B that their total personal income was $100,000 or more were not asked question QI23B about their total family income. FAMINC2 was assigned a code of 29 ($100,000 or more) in this situation.

9.2.3 Additional Edits to Income Variables for Inconsistencies or Logical Inferences

Relatively little additional editing was done to the income variables, aside from assigning legitimate skip codes (Section 2.4.2). This section describes the additional edits that were implemented because of inconsistencies in the data or to make logical inferences.

As noted in Section 4.2.3, interviewers entered information at the beginning of the interview about the state where the dwelling unit was located. This information was used in the income module to fill in the state-specific names for Temporary Assistance for Needy Families (TANF) programs in question QI08N. If the edited variable STATELOC for the state where the dwelling unit was located had been set to bad data because of interviewer errors, any nonblank values in the variable corresponding to question QI08N about cash assistance from a state or county welfare program (FAMPMT) were set to bad data. This edit was implemented because the incorrect state residence information could result in respondents not being asked about cash assistance from the TANF program in the states where they actually lived. In turn, if FAMPMT had been set to bad data because STATELOC had been set to bad data, and FAMSVC did not indicate welfare assistance, then nonblank values in WELMOS also were set to bad data.

Two types of edits were implemented for overall family income variables based on data about personal income:

- If the binary family income variable FINC1 indicated that the family income was less than $20,000, and the corresponding personal income variable PINC1 had a missing value, then the respondent's personal income (PINC1) was logically inferred to be less than $20,000. Similarly, if the binary personal income variable PINC1 was $20,000 or more, and the family income variable FINC1 was skipped, then the total family income variable FAMINC1 indicated that the family income was $20,000 or more.
- If the finer category for total personal income (PINC2) was greater than the corresponding income category in FINC2, then FINC2 and FAMINC2 were set to bad data. For example, if PINC2 indicated that the respondent had an income of $25,000 to $29,999, and the respondent reported a total family income of only $20,000 to $24,999, then FINC2 and FAMINC2 were set to bad data.

9.3 Imputation for the Income Variables

The imputation of income was separated into two imputation sets. The first set involved the imputation of all the binary income variables (i.e., "yes/no" questions about the following sources of income: social security, supplemental security income, welfare cash assistance, welfare noncash assistance, wages, and food stamps), the number of months on welfare (the only variable that was not binary in this imputation set), and a yes/no question regarding whether the respondent's income or the respondent's household family income was $20,000 or more (including income from the sources referenced in the previous questions). This first set was processed using the single response propensity (RP)/multiple prediction (PRD) PMN type described in Section 3.4.3. The second imputation set for finer categories of income consisted of imputing more specific income categories for the respondent and the respondent's family in the household. This set was processed using the single RP/single PRD PMN type as described in Section 3.4.1.

The income variables for total personal income, total family income, and number of months on welfare have lower item response rates compared with most of the other variables that undergo imputation. See Table A.27 in Appendix A for details on the rates of missingness for these variables.

Imputations for all income variables were conducted separately within four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. The segregation into age groups was done to exploit the correlation between the income variables and age and to allow parallel processing of the data (thus reducing the time it takes to implement the procedures).

All income RP and PRD models used a state-rank variable as a covariate. States were ranked by the weighted proportion of respondents whose personal incomes during the prior calendar year were greater than or equal to $20,000. See Section 3.5 for a general discussion of state-rank variables.

9.3.1 Imputation for Binary Income Variables (Imputation Set 1)

The PMN imputation type used for the binary income variables was single RP/multiple PRD. The RP model is described in Section 9.3.1.1. The PRD model for the first binary income variable imputed each year, family social security (FAMSOC), is described in Section 9.3.1.2, and the provisional hot-deck step for FAMSOC is described in Section 9.3.1.3. The remaining PRD and provisional hot-deck steps described in Section 9.3.1.4 list deviations from the analogous steps for family social security. The final hot-deck step applied to all binary income variables is described in Section 9.3.1.5. Finally, a recode for the GOVTPROG variable, made from four imputation-revised binary income variables, is described in Section 9.3.1.6.

For the binary income models that predict whether a respondent had a given source of income, other sources of income were useful covariates. Therefore, provisionally imputed values were used as covariates in subsequent models within the set. The order in which missing values for the binary income variables were provisionally imputed is listed in Table 9.2.

Table 9.2 Imputation Order for Binary Income Variables (Imputation Set 1)

Income Type	Edited Variable Name(s)	Imputation-Revised Variable Name(s)
Family Social Security	FAMSOC	IRFAMSOC
Family Supplemental Security Income	FAMSSI	IRFAMSSI
Family Welfare Payments	FAMPMT	IRFAMPMT
Family Other Welfare Services	FAMSVC	IRFAMSVC
Family Wages	FAMWAG	IRFAMWAG
Family Food Stamps	FSTAMP	IRFSTAMP
Family Months on Welfare	WELMOS	IRWELMOS
Total Personal/Family Income	PINC1, FINC1, FAMINC1	IRPINC1, IRFINC1, IRFAMIN1

9.3.1.1 Response Propensity Step

The response propensity models for the binary income variables utilized the analysis weight, ANALWT. All respondents were in the domain for the binary income variables (i.e., eligible to have a valid value for these variables). For the single RP model, a domain member was considered an item respondent if and only if all of the variables listed in Table 9.2 were nonmissing. See Table D.53 in Appendix D for details of the covariates used in the RP models for these variables.

9.3.1.2 First Prediction Step (FAMSOC)

FAMSOC is a binary variable; after imputation, its only values are "yes" and "no" since all other missing data codes are replaced with imputed values. Using the adjusted weights that are outputs of the RP step, FAMSOC was modeled using logistic regression and, in particular, the RLOGIST procedure in SAS-callable SUDAAN. The single predicted mean used in later hot-deck steps was the predicted probability that the respondent's family-in-household received income from social security in the preceding calendar year.

9.3.1.3 First Provisional Hot-Deck Step (FAMSOC)

The provisional hot-deck applied to the FAMSOC variable is a simplified version of the final hot-deck step for the binary income variables, because its only purpose is to fill in missing values so that FAMSOC can be used as a covariate in the PRD models for binary income variables imputed later in the sequence. Section 3.4.2 describes the concept of a provisional hot-deck in more detail. The final hot-deck step is described in Section 9.3.1.5 and in tabular form in Tables E.98, E.99, and E.100 in Appendix E. This provisional hot-deck step included the following:

- The predictive mean vector included only one element: the predicted mean from the preceding step.
- No logical constraints were used.
- The following likeness constraints were used:
  - IRFAMSKP of donor = IRFAMSKP of recipient. This is a likeness constraint in this hot-deck step, not a logical constraint. The creation of IRFAMSKP is described in Section 8.3.2 (LogC1 in Table E.98).
  - Donor's predicted mean must be within 5 percent of recipient's predicted mean (delta constraint; LikC3 in Table E.99).
  - Donor must match recipient with respect to whether there are adults aged 65 or older in the household. The variable used in this constraint was IRHH65, whose creation is described in Section 8.3.2 (LikC5 in Table E.99).
- In the first attempt to find a donor, all three likeness constraints were applied. In the second attempt, the delta constraint (#2 above) was removed.

The provisionally imputed version of FAMSOC was called INTFAMSOC. INTFAMSOC was used as a covariate in the rest of the binary income PRD models.

9.3.1.4 Analogous Prediction and Provisional Hot-Deck Steps for Remaining Binary Income Variables (FAMSSI, FAMPMT, FAMSVC, FAMWAG, FSTAMP, WELMOS, and FAMINC1)

PRD models were fit for FAMSSI, FAMPMT, FAMSVC, FAMWAG, FSTAMP, and FAMINC1[139] in the same manner as for FAMSOC, as described previously. Only the PRD model for WELMOS was different. The domain of the WELMOS model included only respondents who reported that their family-in-household received welfare payments and/or other welfare services during the preceding calendar year, as defined by FAMPMT and FAMSVC. Least squares regression (not logistic regression) was used, where the dependent variable was a standard logit,[140] such that $Y = \text{logit}(p)$ and p = number of months on welfare divided by 12. The REGRESS procedure in SAS-callable SUDAAN was used to fit the model. The predicted mean from the WELMOS model was the predicted probability of receiving welfare in a given month in the previous calendar year, given that the respondent received welfare payments and/or welfare services in the previous calendar year.

The provisional imputation steps for the other variables were implemented in the same manner as for FAMSOC, as described previously, except that the likeness constraints sometimes differed slightly. The following deviations are noted:

- Only the provisional hot-deck step for FAMSOC included the constraint involving the IRHH65 variable.
- For FAMPMT and FAMSVC, LikC6 in Table E.99 was used: donor must match recipient with respect to whether there are children younger than 18 in the household. The variable used in this constraint was IRKID17, whose creation is described in Section 8.3.2. This likeness constraint was used in both the first and second attempts to find a donor.
- For FAMWAG, LikC7 in Table E.99 was used: donor must match recipient with respect to whether there are adults aged 18 to 64 in the household. The variables used in this constraint were IRHHSIZE, IRHH65, and IRKID17. The creation of these variables is described in Section 8.3.2. This likeness constraint was used in both the first and second attempts to find a donor.
- Also for FAMWAG, donor must match recipient with respect to whether he or she was employed. The variable used in this constraint was EMPSTATY, whose creation is described in Section 5.3.1. If the recipient was employed full time or part time, then the donor must also have been employed full time or part time. If the donor was not employed full time or part time, then the donor must also not have been employed full time or part time. This likeness constraint was used in both the first and second attempts to find a donor.
- FAMINC1 did not undergo a provisional imputation step since it was the last variable in the set.

---

[139] The response variable for the model for total income was total family income (not total personal income), and the provisionally imputed version of WELMOS was not included as a covariate in the model.

[140] The Cox empirical logit was used when an individual was on welfare for all 12 months.

9.3.1.5 Final Hot-Deck Step

Details on the missingness patterns, constraints, and predictive mean vectors for the binary income variables' final hot-deck step are available in Appendix E. This section explains the more general ideas behind the hot-deck step applied to these variables.

Because 10 imputation-revised variables are created in this step, and almost any subset of them can be missing, there are 639 missingness patterns for the binary income variables. The first 255 cover patterns where WELMOS is nonmissing; these are summarized as pattern 5 in Table E.100. The remaining 384 cover patterns where WELMOS is missing; these are summarized as patterns 1 through 4 in Table E.100.

- When WELMOS is nonmissing, any subset of FAMSOC, FAMSSI, FAMPMT, FAMSVC, FAMWAG, FSTAMP, PINC1, and FINC1 may be missing. However, if *none* of them are missing, then nothing is missing and no imputation is required. This leads to $2^8 - 1 = 255$ missingness patterns.[141]
- When WELMOS is missing, the elements of the predictive mean vector related to FAMPMT, FAMSVC, and WELMOS depend on the values of FAMPMT and FAMSVC. There are six relevant combinations of FAMPMT and FAMSVC (Table 9.3). For each of the six, any subset of FAMSOC, FAMSSI, FAMWAG, FSTAMP, PINC1, and FINC1 may be missing. This leads to $6 \times 2^6 = 384$ missingness patterns.
  - For the combinations labeled 1 in Table 9.3, the recipient is known to have received welfare payments and/or welfare services and neither is missing. Here, no predictive mean vector elements are required for FAMPMT and FAMSVC,[142] and there is no need to manipulate the element for WELMOS, which is already conditional on receipt of welfare as defined by FAMPMT and FAMSVC.
  - For combinations 2 and 3, the recipient is also known to have received welfare payments or welfare services, but one is missing. Here, a predictive mean vector element (i.e., a predicted mean) is required for the missing one of the two, but again there is no need to manipulate the corresponding element for WELMOS.
  - For combination 4, the recipient is known not to have received welfare payments, but welfare services is missing. Here, a predictive mean vector element is required for FAMSVC. The WELMOS predicted mean must be made conditional on the receipt of welfare services in the previous calendar year. Combination 5 is similar, but FAMPMT is the missing one.
  - For combination 6, both FAMPMT and FAMSVC are missing. Here, predictive mean vector elements are required for both. The WELMOS predicted mean must be made conditional on the receipt of welfare payments and/or welfare services. This probability is unknown, but can be crudely approximated by assuming the two are independent. Under that assumption, the probability of receiving either or

[141] Technically, if FAMPMT and FAMSVC were both missing, then WELMOS was necessarily missing. Therefore, $2^6 = 64$ out of these 255 missingness patterns are logically impossible and need not be set up.

[142] In general, in either the single RP/multiple PRD or multiple RP/multiple PRD types of PMN, only the elements of the predictive mean vector corresponding to missing responses were used (see Chapter 3).

both is one minus the probability of receiving neither, as expressed by (1 – PMT) × (1 – SVC).[143]

- If FAMPMT = "No" and FAMSVC = "No," then WELMOS is assigned a skip code and is therefore nonmissing. This combination is irrelevant for the purposes of this discussion, which only covers cases where WELMOS is missing.
- Combinations 1 through 3 in Table 9.3 correspond to pattern 1 in Table E.100, combination 4 corresponds to pattern 2, combination 5 corresponds to pattern 3, and combination 6 corresponds to pattern 4.

Table 9.3 Cross-Classification of FAMPMT and FAMSVC; Relevant Combinations for Binary Income Predictive Mean Vector when WELMOS Is Missing

		FAMSVC		
		Yes	**No**	**Missing**
FAMPMT	Yes	1	1	2
	No	1	N/A	4
	Missing	3	5	6

N/A = not applicable.

Both the likeness and logical constraints are written in such a way that they can be summarized using the five broad missingness patterns listed in Table E.100. For example, likeness constraint 4 states, "If recipient is missing months on welfare, then donor must match recipient with respect to personal welfare payments (if nonmissing) and welfare services (if nonmissing)." This constraint does not apply to the first 255 missingness patterns, nor does it apply to the 64 missingness patterns in combination 6 where both FAMPMT and FAMSVC are missing. However, instead of carefully listing the exact missingness patterns to which each constraint applied, the constraints were written using conditions that allowed them to be applied only where necessary.

The few logical constraints are due to relationships between (1) FAMPMT, FAMSVC, and WELMOS and (2) PINC1, FINC1, and FAMINC1. A key likeness constraint exploited the high degree of association among welfare payments, welfare services, food stamps, binary total income (at the personal and family level), and months on welfare. If a recipient required imputation for one or more of these six variables, but had information on at least one of these variables, the donors were restricted so that donors and recipients had the same values for these nonmissing variables.

As stated in the introduction to Section 9.3, the variables for total personal income, total family income, and number of months on welfare have lower item response rates compared with most of the other variables that undergo imputation. The constraints are fairly complex and extensive, especially when several of the other dichotomous income variables are missing. Also,

[143] A reasonable alternative method that requires no assumption of independence would be to model FAMPMT and FAMSVC together as a categorical variable with four levels: both, only FAMPMT, only FAMSVC, and neither. Then, the probability of receiving welfare payments and/or other welfare services is simply the sum of the first three predicted means. The assumption of independence is certainly questionable.

when several variables are missing, it is difficult to find a donor on the first try because the donor must be close to the recipient for all predicted means corresponding to the missing variables.

9.3.1.6 Recodes for Additional Analyses

A dichotomous recoded income variable GOVTPROG indicated whether the respondent participated in any government assistance programs. It was created from four imputation-revised variables: family Supplemental Security Income (IRFAMSSI), family food stamps (IRFSTAMP), family welfare payments (IRFAMPMT), and family welfare services (IRFAMSVC). Although a variety of recoded variables were created, only GOVTPROG is described here because it was used as a covariate in subsequent health insurance imputation models. See Chapter 10 for details on the imputation of missing values in the health insurance variables.

9.3.2 Imputation for Finer Category Income Variables (Imputation Set 2)

Three income variables resulted from editing the questions in the finer category income phase: personal total income (PINC2), total family income if there are other family members (FINC2), and total family income (FAMINC2). These edited variables are described in Section 9.2.2. All three imputation-revised variables derived from their edited counterparts were created using the single RP/single PRD PMN type described in Chapter 3. The single PRD model was fit for total family income (FAMINC2), but the item nonrespondent also received values from the donor for PINC2 and FINC2 if those were missing. There were no deviations from this general approach.

9.3.2.1 Response Propensity Step

The response propensity models for the finer category income variables utilized the analysis weight, ANALWT. All respondents were in the domain for this imputation set. Item respondents were those with nonmissing values for PINC2 and FINC2. See Table D.58 in Appendix D for details of the covariates used in the RP models for these variables.

9.3.2.2 Prediction Step

Each of the three finer category income variables was ordinal with 29 levels. Each response category covers an interval of income, with levels as follows:

1	LESS THAN $1,000 (INCLUDING LOSS)
2	$1,000 - $1,999
3	$2,000 - $2,999
4	$3,000 - $3,999
5	$4,000 - $4,999
6	$5,000 - $5,999
7	$6,000 - $6,999
8	$7,000 - $7,999
9	$8,000 - $8,999
10	$9,000 - $9,999
11	$10,000 - $10,999

12	\$11,000 - \$11,999
13	\$12,000 - \$12,999
14	\$13,000 - \$13,999
15	\$14,000 - \$14,999
16	\$15,000 - \$15,999
17	\$16,000 - \$16,999
18	\$17,000 - \$17,999
19	\$18,000 - \$18,999
20	\$19,000 - \$19,999
21	\$20,000 - \$24,999
22	\$25,000 - \$29,999
23	\$30,000 - \$34,999
24	\$35,000 - \$39,999
25	\$40,000 - \$44,999
26	\$45,000 - \$49,999
27	\$50,000 - \$74,999
28	\$75,000 - \$99,999
29	\$100,000 OR MORE

The FAMINC2 variable was modeled using the LIFEREG procedure in SAS/STAT® software.[144] This procedure was used for regression modeling of continuous nonnegative random variables, such as survival times and income, by fitting models that are sometimes referred to as "failure time models." This particular type of model, which was assumed for the response variable representing income, can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where **y** is a vector of observed responses, **X** is the matrix of covariates, $\boldsymbol{\beta}$ is the parameter vector, and $\boldsymbol{\varepsilon}$ is a vector of error terms. In particular, the error terms are assumed to come from a known multivariate distribution, such as the logarithm of a three-parameter generalized gamma model, or a more common two-parameter distribution, such as gamma, Weibull, lognormal, or log-logistic. Although the underlying random variable **y** is assumed to be continuous, the LIFEREG procedure allows the variable to be reported in interval categories, consistent with the 29 NSDUH income intervals defined previously for these finer category income variables. The contribution of an individual with covariates in the matrix **X** to the overall likelihood is simply the probability mass assigned by the model to the interval (l, u) containing the actual continuous income for that individual. For this interval, l represents the lower bound and u represents the upper bound. This contribution has the form $F(u|\mathbf{X},\boldsymbol{\beta},\sigma^2) - F(l|\mathbf{X},\boldsymbol{\beta},\sigma^2)$, where F is a cumulative distribution function and σ^2 represents the variance of the individual responses. The LIFEREG procedure uses standard likelihood methods of inference and incorporates the survey weights.

LIFEREG allows several choices for the functional form of the parametric model that correspond to the error distribution, including the two-parameter log-logistic, lognormal, gamma,

---

[144] Details about the LIFEREG procedure are discussed in the *SAS/STAT® 13.2 User's Guide: High-Performance Procedures* (SAS Institute Inc., 2014).

and Weibull and the three-parameter generalized gamma. Compared with the other models, the gamma distribution provided a better overall fit, as indicated by likelihood techniques. Because the three-parameter generalized gamma did not significantly improve on its two-parameter special cases, when using the likelihood ratio tests as criteria for comparison, it was decided to use a two-parameter model.

The predicted mean used in the subsequent hot-deck step was the term $\mathbf{X\beta}$, which was the predictive mean value. This value was a monotonic function of the conditional mean of the modeled income distribution at a given individual set of values of the regression covariates. Specifically, $\mathbf{X\beta}$ was a translation of the estimated mean of log income.

9.3.2.3 Hot-Deck Step

The hot-deck step for the finer category income variables is an example of univariate matching, multivariate assignment (see Table 3.4 in Chapter 3). The only predicted mean used in the hot-deck step was related to FAMINC2, but the recipient received values for PINC2 and FINC2 if either or both were missing. The imputation-revised version of FAMINC2 (IRFAMIN2) was created from the imputation-revised versions of PINC2 and FINC2, called IRPINC2 and IRFINC2, respectively. The constraints ensure consistency with existing information such as IRPINC1 and IRFINC1 from the preceding imputation set, and PINC2 and FINC2 if nonmissing.

The finer category income variables have among the lowest response rates of any NSDUH variables that undergo imputation. As a result of the absence of many constraints, the single predicted mean, and the large domain, most respondents are handled in the first attempt to find a donor.

9.3.2.4 Recodes for Additional Analyses

The recoded variable INCOME classified the families of respondents into four income levels: less than \$20,000; \$20,000 to \$49,999; \$50,000 to \$74,999; and \$75,000 or more. Another recoded variable (INCOME5) was created to take advantage of an extra level of income. This variable had five levels: the first three levels were equivalent to INCOME, but the last level of INCOME was separated into two levels: \$75,000 to \$99,999 and \$100,000 or more. Both INCOME and INCOME5 were recodes of the variable IRFAMIN2. A variety of recoded variables were created but are not discussed in this report. However, as with GOVTPROG, the variable INCOME is discussed here because it was used as a covariate in subsequent health insurance models (see Chapter 10 for details on the imputation of missing values in the health insurance variables). INCOME5, which is currently used for special requests, also is discussed because it is similar to the INCOME variable.

10. Editing and Imputation for the NSDUH Health Insurance Variables

10.1 Introduction

This chapter discusses procedures for logically editing and imputing data from the interviewer-administered sections pertaining to health insurance coverage. Respondents (or other household members serving as proxies because they were better able to provide information about the respondent's health insurance coverage) were asked whether they were currently covered by different types of health insurance.[145] If private health insurance coverage was reported, respondents were asked whether that included coverage for substance abuse treatment or mental health services. Data also were collected on periods when respondents never had health insurance coverage, most recent coverage (if currently without health insurance), and reasons for losing health insurance coverage or for never having had coverage (if applicable).

On the 2002-2012 National Surveys on Drug Use and Health (NSDUHs), missing values for the health insurance variables were replaced with valid values using two different methods: the "old" method and the "constituent variables" method. Both are predictive mean neighborhood (PMN) imputation methods. The "old" method consists of edited and imputed values for three health insurance variables in a way that is consistent with iterations of the NSDUH questionnaire prior to 2002. The "constituent variables" method edits and imputes for specific health insurance variables (e.g., Medicaid or Medicare coverage) using all relevant questions in the questionnaire from 2002 to the present and was first implemented for the health insurance module in 2002.

Starting with the 2013 NSDUH, however, the creation of the edited and imputed "old" method variables was discontinued, and the "constituent variables" method was thereafter referred to as simply the health insurance method. There were two main reasons for this decision. First, analysts have long been cautioned against performing trend analyses that combine pre-2002 survey years with 2002 and beyond. Second, including both sets of health insurance variables had the potential to confuse analysts who might use the incorrect health insurance variable set to produce estimates.

As with other variable groups (e.g., demographics, drugs), the health insurance variables were imputed in sets as follows:

- Imputation Set 1: IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT (types of health insurance)
- Imputation Set 2: IROTHHLT (any other health insurance)

[145] Reference is made only to "respondents" in the remainder of this chapter. However, readers are advised that the health insurance information for a respondent may have been provided by another adult household member who was serving as a proxy for the respondent because the proxy was considered to be better able to answer the health insurance questions for the respondent. See Chapter 8 for more information about proxy variables.

The five base variables (IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, IROTHHLT) that undergo imputation are simple recodes created from six "source" variables, each of which maps to a single question in NSDUH. In the 2014 survey, the source variables had the same values as the raw variables created from the responses to the questions, except that missing values were replaced by standard NSDUH missing value codes (Sections 2.2.1 and 2.4.3). With the exception of MEDICAID and CHIPCOV, the variables remain separate to form the base variables for imputation sets 1 and 2. Section 10.2.5 discusses the creation of base variables used in imputation sets 1 and 2.

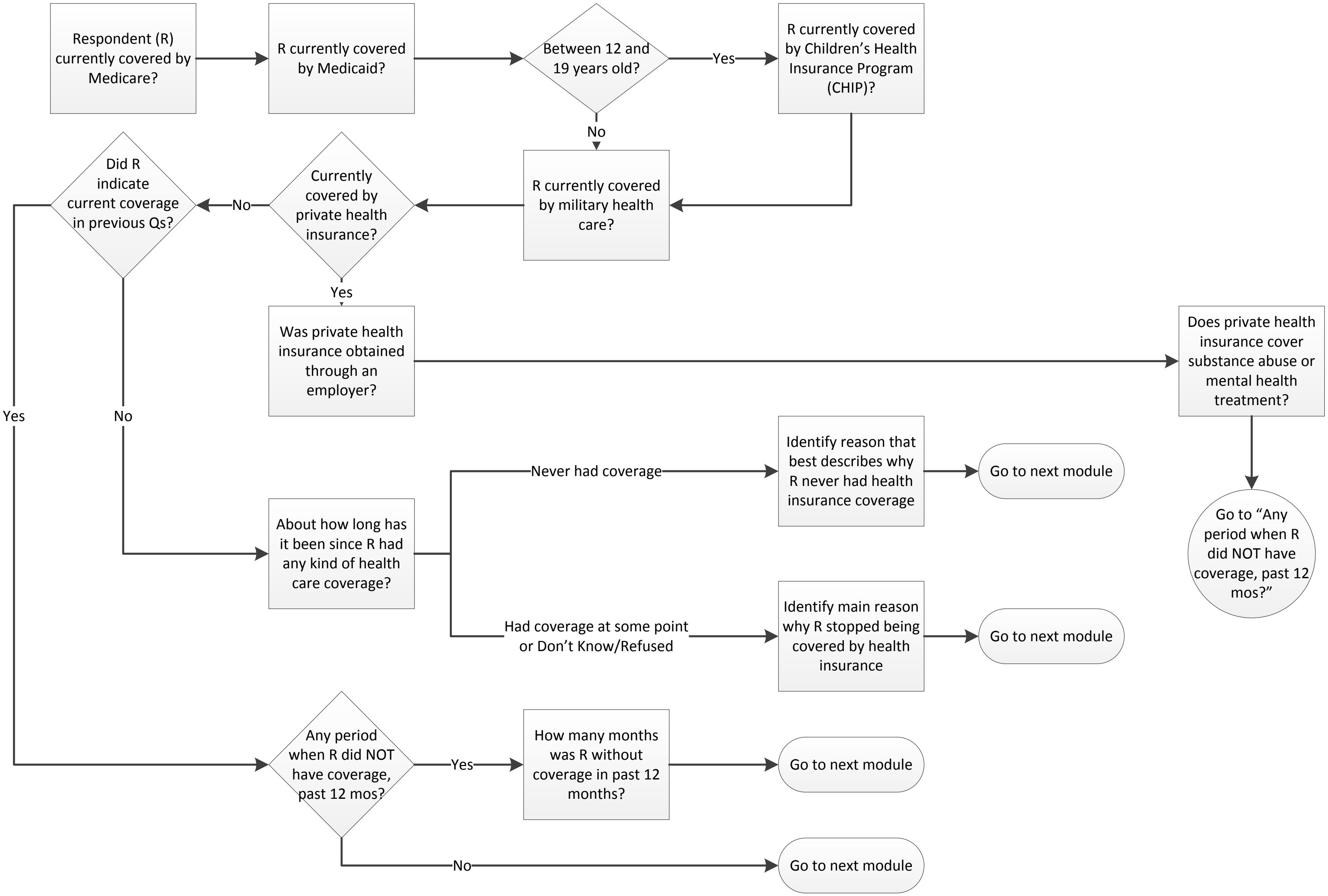
10.2 Editing the Health Insurance Variables

Because the health insurance questions are in a noncore section of the interview (Section 1.1), the content of these questions can change across years to improve the measurement of health insurance coverage. In practice, the content of this module has undergone relatively little change since the start of the new baseline in 2002. Documentation of historical changes to the health insurance questions prior to the new baseline in 2002 can be found in the report on editing of interviewer-administered data that was prepared for the 2011 Methodological Resource Book (Kroutil, Chien, Handley, & Bradshaw, 2013).

An important aspect of editing the health insurance variables consisted of assigning legitimate skip codes based on the skip logic in this section (Section 2.4.2). Figure 10.1 shows the logic for asking questions in the health insurance module. For example, if respondents answered "no" (where applicable) to questions QHI01 through QHI06 and then reported in QHI11 that they were not currently covered by any kind of health insurance (QHI11 = 2), legitimate skip codes were assigned to HLCNOTYR (anytime in the past 12 months that respondents were without health insurance, corresponding to question QHI13) and HLCNOTMO (number of months that respondents were without health insurance in the past 12 months, corresponding to question QHI14). Similarly, if respondents reported some type of current health insurance coverage in QHI01 through QHI06, edited variables corresponding to questions QHI15 through QHI18 were assigned legitimate skip codes (i.e., HLCLAST through HLNVSOR).

As was the case in prior years, question QHI18 (reasons that the respondent never had health insurance) was an "enter all that apply" question. Therefore, the edited variables corresponding to question QHI18 (HLNVCOST through HLNVSOR) were assigned a code of 1 (Response entered) if the corresponding response category was chosen from QHI18. The variables were assigned a code of 6 (Response not entered) if the corresponding response category was not chosen but at least one response had been entered in QHI18 (Section 2.4.4).

Figure 10.1 Routing Logic for Respondents in the Health Insurance Module



*Respondents immediately go to the next module (i.e., without further health insurance questions being asked) if current health insurance coverage status is unknown.

10.2.1 Editing Because of Incorrect State Information

Question QHI02 asked whether respondents were currently covered by Medicaid, and question QHI02A asked whether respondents aged 12 to 19 were covered by the Children's Health Insurance Program (CHIP). The CAI program filled in names for state-specific Medicaid program or CHIP names to aid respondent identification. If the interviewer's report of the state where the dwelling unit was located (edited variable STATELOC) had been set to bad data because of inconsistencies in the state information for the respondent (Section 4.2.3), then the edited variable CAIDCHIP for coverage by Medicaid or CHIP (see Section 10.2.2) was usually assigned a bad data code as well. The rationale for this edit was that the CAI logic would supply an incorrect name for the state's Medicaid program or CHIP if the state information that the interviewer reported was incorrect. Consequently, the respondent would be answering QHI02 or QHI02A based on a version of the question that did not correctly correspond to where the respondent would be eligible for publicly funded health insurance coverage. For example, if a respondent was aged 12 to 19 and was living in California (FIPE4 = 5), the respondent should have been asked in QHI02A whether he or she was covered by "Healthy Families." However, if a value of 6 had been entered in FIPE4 (i.e., for Colorado), the respondent would be asked whether he or she was covered by "Child Health Plan Plus, or CHP+."

An exception to this assignment of bad data codes concerned the special situation in which respondents were routed to questions QHI15 (time since the respondent last had health insurance) and QHI17 (main reason for losing health insurance coverage). If responses to QHI15 or QHI17 indicated that the respondent did not currently have (or never had) health insurance coverage, CAIDCHIP retained a code of 2 (i.e., "no"), even if STATELOC had been set to bad data, for consistency with information from QHI15 or QHI17 that the respondent was not currently covered by any type of health insurance.

10.2.2 Determination of Any Health Insurance Coverage

Determining whether respondents were currently covered by any type of health insurance was critical for determining the subsequent questions that respondents would be asked about periods without health insurance coverage in the past 12 months, most recent health insurance coverage (if not currently covered), reasons for losing health insurance, or reasons for never having health insurance. The following content issues have been relevant since the new NSDUH baseline in 2002 for determining whether respondents had health insurance coverage:

- As noted previously, respondents who were aged 12 to 19 are asked in question QHI02A whether they were covered by CHIP. Government experts in the health insurance field advised SAMHSA that it would be virtually impossible to produce separate estimates of coverage by the Medicaid program (question QHI02) and coverage by CHIP. For this reason, the variable CAIDCHIP has been created from responses to questions QHI02 and QHI02A.
- The state-specific Medicaid or CHIP names that are used to customize questions QHI02 and QHI02A may be updated for the CAI program across years to reflect any changes to these names at the state level.

- If STATELOC had a valid value (Sections 4.2.3 and 10.2.1), CAIDCHIP was assigned a code of 1 (i.e., "yes") if an affirmative response occurred in either QHI02 or QHI02A (if applicable). CAIDCHIP was coded as 2 ("no") if QHI02 was answered as "no" and (a) QHI02A also was answered as "no" (for respondents who were aged 12 to 19) or (b) QHI02A had been legitimately skipped (for respondents aged 20 or older). Otherwise, CAIDCHIP was coded as 97 ("refused") if a code of 97 occurred in either QHI02 or QHI02A, or 94 ("don't know") if a code of 94 (and no code of 97) occurred in these items. Remaining cases that did not meet any of these criteria were coded as 98 (i.e., blank).
- Respondents who answered "no" to all questions about Medicare, Medicaid, CHIP (if applicable), military health coverage, and private health insurance were asked a follow-up question (QHI11) to determine if they were covered by *any* type of health insurance. Responses to this question were used to determine subsequent routing in the health insurance section depending on whether respondents currently had or did not have health insurance. The variable HLTINNOS was created from QHI11.
- A recoded "any health insurance" variable, ANYHLTIN, was created from responses to MEDICARE (from QHI01), CAIDCHIP (from QHI02 and QHI02A), CHAMPUS (from QHI03), PRVHLTIN (from QHI06), and HLTINNOS (from QHI11).
  - If any affirmative response was reported in any of the above variables, ANYHLTIN was coded as 1 ("yes").
  - Otherwise, if HLTINNOS (from QHI11) had been answered as "no" (and by definition, preceding questions had been answered as "no"), ANYHLTIN was coded as 2 ("no").
  - If ANYHLTIN was not already coded as 1 or 2, it was coded as 97 ("refused") or 94 ("don't know"), as follows: (a) if a code of 97 (i.e., "refused") occurred in any of the above health insurance items, ANYHLTIN was coded as 97; or (b) ANYHLTIN was coded as 94 (i.e., "don't know") if a code of 94 (but no code of 97) occurred in the above items.
  - For remaining cases (e.g., if variables had been set to bad data, or a breakoff had occurred), ANYHLTIN retained a code of 98 (OTHER MISSING).

In addition, because the names for Medicare and Medicaid sound similar, follow-up questions have been included in the health insurance section of the interview since 2003 for respondents who were (1) younger than 65 and reported being covered by Medicare (which is for people aged 65 or older or certain disabled individuals) or (2) aged 65 or older and reported being covered by Medicaid (which is for low-income individuals or disabled individuals). In these situations, respondents were asked to confirm their previous answer. If respondents indicated on follow-up that their previous answer to Medicare or Medicaid coverage was incorrect, it was determined that they were not covered by that particular type of health insurance. However, if respondents did not know or refused to confirm whether their previous answer to the Medicare or Medicaid coverage question was correct, the editing procedures continued to classify them as having that type of coverage. Stated another way, the only response on follow-up that overruled a previous report of coverage by Medicare or Medicaid was for respondents to indicate that they did not have that type of coverage.

Despite these changes in 2003 for questions about Medicare or Medicaid coverage, the names for the edited variables did not change. The edited variable for Medicare coverage retained the name MEDICARE. The edited variable for coverage by Medicaid or CHIP retained the name CAIDCHIP (see previous discussion in this section).

10.2.3 Editing of Health Insurance Follow-Up Variables Based on Current Coverage

Information from ANYHLTIN and a preliminary edited variable SKHLCCOV[146] were used to determine whether variables pertaining to periods without health insurance coverage in the past 12 months, most recent health insurance coverage (if not currently covered), reasons for losing health insurance, and reasons for never having health insurance should be assigned "legitimate skip" codes (Section 2.4.2).

- If respondents reported some health insurance coverage in QHI01 to QHI11, legitimate skip codes were assigned to the edited variables corresponding to questions QHI15 to QHI18. Questions QHI15 to QHI18 applied to respondents who currently were without coverage and pertained to situations in which respondents formerly or never had coverage.
- If respondents currently were without any health insurance coverage (i.e., applicable questions in QHI01 to QHI11 all were answered as "no"), legitimate skip codes were assigned to the edited variables corresponding to questions QHI13 and QHI14. Questions QHI13 and QHI14 applied to respondents who currently had coverage and pertained to periods when these respondents were without coverage in the past 12 months.
- If it was unknown from questions QHI01 to QHI11 whether or not respondents currently had health insurance coverage, the CAI program skipped respondents out of all remaining health insurance questions. The edited variables corresponding to questions QHI13 to QHI18 had missing values.

10.2.4 Miscellaneous Issues for Editing the Health Insurance Variables

Table 10.1 discusses edit specifications for the health insurance variables. For example, the data could indicate that respondents were covered currently by Medicare, Medicaid, CHIP (for respondents who were aged 12 to 19), some type of military health coverage (e.g., CHAMPUS or the VA), or private health insurance. If respondents were reported to have been *currently* covered by all of the types of insurance they were asked about, a flag was set and included on the data file. The original data were retained, but this flag was designed to alert analysts to the presence of this unlikely data pattern.

---

[146] SKHLCCOV was a precursor to the final variable ANYHLTIN but was used in some places in the editing program to determine whether respondents were legitimately skipped out of certain follow-up questions about health insurance coverage. Both variables indicate whether respondents currently had health insurance coverage. SKHLCCOV was coded as 3 if respondents reported in question QHI02A that they were covered by CHIP but they did not report coverage by Medicaid, Medicare, military health insurance (e.g., CHAMPUS), or private health insurance. These values of 3 in SKHLCCOV were recoded to 1 in ANYHLTIN.

In addition, the only types of current health insurance coverage that were asked about in 1999 were Medicare, Medicaid, some type of military health coverage, or private health insurance. Therefore, a second flag was set for comparability with a similar flag set in the 1999 data. This second flag indicated when respondents reported that they were covered currently by all four of these types of health insurance that were asked about in 1999, even if they did not report being covered by CHIP (if aged 12 to 19) or they were aged 20 or older and were skipped out of question QHI02A.

Table 10.1 Edits Pertaining to the Health Insurance Section

Issue	Edits Implemented
The respondent (R) reported being currently covered by Medicare, Medicaid, Children's Health Insurance Program (CHIP) (if aged 12 to 19), military coverage, and private health insurance.	A flag (HLCALLFG) was provided to indicate that this pattern occurred, but no further editing was done to the data.
The R reported being currently covered by Medicare, Medicaid, military coverage, and private health insurance, the only types of current coverage that were asked about in 1999.	A flag (HLCALL99) was provided to indicate that this pattern occurred, but no further editing was done to the responses. This HLCALL99 variable was comparable to the HLCALLFG variable in 1999.
The R's only indication of current health insurance coverage came from reports of coverage by Medicaid or CHIP, but the state location variable STATELOC had been set to bad data.	Nonblank values in the variables pertaining to any period in the past 12 months when the R was without health insurance (HLCNOTYR, corresponding to question QHI13) and the number of months that the R was without health insurance in the past 12 months (HLCNOTMO, corresponding to question QHI14) were replaced with bad data codes. This edit was not done if the R indicated current coverage by Medicare, the military, or private health insurance.
The R had some indication of current coverage from at least one of the five sources of insurance listed above. However, the R also was reported to have had a period in the past 12 months when he or she was without health insurance. Further, it was reported that the R had been without health insurance for 12 of those months.	No editing was done when this pattern occurred. The rationale for not doing any editing was that the R may just recently have gotten insurance or have become qualified for insurance.
The R had no indication of current coverage from any of the five sources of insurance listed above. If the R (or proxy) answered "don't know" or "refused" when asked when the R last had coverage, the R was routed to questions about what coverage the R last had, and why the R lost health insurance coverage. That is, the skip logic assumed that the R had some prior history of coverage, but that may not necessarily have been the case.	If the R was reported to have had previously some form of health insurance or medical coverage, or if some reason was given why the R lost insurance coverage, legitimate skip codes were assigned to the variables pertaining to reasons that the R never had coverage. That is, the implicit assumption made in the computer-assisted interviewing (CAI) skip logic was verified by an answer indicating some prior history of health insurance coverage. However, if nothing was reported to indicate that the R previously had health insurance, the skipped variables pertaining to reasons for never having had insurance retained codes of blank.
The R was male but reported in QHI17 that he lost health insurance coverage because he "received Medicaid or medical insurance only while pregnant."	The edited variable HLLOSRSN (corresponding to QHI17) was set to bad data.

10.2.5 Editing Types of Health Insurance and Any Other Health Insurance Variables, Imputation Sets 1 and 2

The editing process for health insurance combined the source variables (MEDICAID and CHIPCOV) to create the variable CAIDCHIP (whether someone was covered by Medicaid or one of the state children's health insurance programs). This CAIDCHIP variable and all the other source variables in Table 10.2, except HLTINNOS, were used directly as base variables for imputation.

Table 10.2 Mapping of Raw Health Insurance Variables to Edited Variables

Question Number[1]	Question Text[2]	Source Variable[3]	Used to Create CAIDCHIP	Used to Create MEDICARE	Used to Create CHAMPUS	Used to Create PRIVATE	Used to Create ANYOTHER
QHI01, QHI01v	Is the respondent covered by Medicare?	MEDICARE (1 = yes, 2 = no)	No	Yes	No	No	Yes
QHI02, QHI02v	Is the respondent covered by Medicaid or Medical Assistance?	MEDICAID (1 = yes, 2 = no)	Yes	No	No	No	Yes
QHI02A	Is the respondent currently covered by a Children's Health Insurance Program operated by your state of residence?[4]	CHIPCOV (1 = yes, 2 = no)	Yes	No	No	No	Yes
QHI03	Is the respondent currently covered by CHAMPUS or TRICARE, CHAMPVA, the VA, or military health care?	CHAMPUS (1 = yes, 2 = no)	No	No	Yes	No	Yes
QHI06	Is the respondent currently covered by private health insurance?	PRVHLTIN (1 = yes, 2 = no)	No	No	No	Yes	Yes

Table 10.2 Mapping of Raw Health Insurance Variables to Edited Variables (continued)

Question Number[1]	Question Text[2]	Source Variable[3]	Used to Create CAIDCHIP	Used to Create MEDICARE	Used to Create CHAMPUS	Used to Create PRIVATE	Used to Create ANYOTHER
QHI11	Is the respondent currently covered by any kind of health insurance, that is, any policy or program that provides or pays for medical care?	HLTINNOS (1 = yes, 2 = no, 99 = legitimate skip5)	No	No	No	No	Yes

[1] The "v" questions were asked to verify the answer given in the previous question for respondents who were younger than 65 and a Medicare recipient or older than 65 and a Medicaid recipient.
[2] The questions provided in this table are abbreviated versions of those given in the questionnaire.
[3] Missing values in these edited values were represented by standard missing value codes. CHIPCOV was replaced in the final analytic file by CAIDCHIP, a combination of MEDICAID and CHIPCOV. See Section 10.2.2 for details.
[4] The questionnaire did not ask the question exactly in this way. It identified the specific program, depending upon the state of residence entered by the respondent and was asked only of respondents aged 12 to 19.
[5] Respondents were assigned a legitimate skip for HLTINNOS if they answered "yes" or gave no answer to at least one of the other health insurance questions.

A respondent was routed to QHI11 (whether the respondent was covered by any kind of health insurance at the time of the survey) if they answered "no" to all the other health insurance questions. All other respondents were given a legitimate skip value to the variable HLTINNOS, as shown in Table 10.2. Therefore, it was possible that the imputation-revised versions of the four specific health insurance variables would all have had a value of "no," and the value of HLTINNOS would have been a legitimate skip, if one or more of the "no" values was imputed. In this instance, another variable was needed to reflect the fact that a respondent could have had a valid "yes" or "no" imputed value for any other health insurance, even though the respondent was never asked QHI11 and was assigned a legitimate skip code. The ANYOTHER variable was created using the preliminary edited variable, SKHLCCOV (Section 10.2.3), which indicated whether a respondent was covered by any health insurance. SKHLCCOV and ANYOTHER were defined as follows:

SKHLCCOV =

- 1 (or 3[147]) if CAIDCHIP = 1, MEDICARE = 1, CHAMPUS = 1, PRVHLTIN = 1, or HLTINNOS = 1; else
- 2 if CAIDCHIP = 2, MEDICARE = 2, CHAMPUS = 2, PRVHLTIN = 2, and HLTINNOS = 2; else

[147] See Section 10.2.3 for discussion of the assignment of a code of 3 to SKHLCCOV. Respondents with SKHLCCOV = 3 were treated in the same manner as those with SKHLCCOV = 1.

- missing value code if the nonmissing values of CAIDCHIP, MEDICARE, CHAMPUS, PRVHLTIN, and HLTINNOS are all "2," and at least one of these variables had a missing response.

ANYOTHER =

- legitimate skip code (99) if CAIDCHIP = 1, MEDICARE = 1, CHAMPUS = 1, or PRVHLTIN = 1; else
- SKHLCCOV.

10.3 Imputation for the Health Insurance Variables

Beginning in the 2013 NSDUH, missing values for the health insurance variables were replaced with valid values using the health insurance method,[148] whereby indicator variables for more specific types of health insurance (i.e., coverage by Medicare or by Medicaid) were imputed in two imputation sets. Imputation set 1 included the variables IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. These are indicators of coverage for Medicaid/CHIP, Medicare, CHAMPUS or similar coverage for military personnel, and private health insurance, respectively. Section 10.3.1 describes the creation of the four imputation-revised health insurance variables that comprise imputation set 1. Within imputation set 2, the "any other health insurance" variable, IROTHHLT, was created (Section 10.3.2). Together, these five imputation-revised health insurance variables were then used to create a third indicator of "any health insurance" coverage, IRINSUR4.

The final health insurance imputations were performed using the PMN methodology. The health insurance variables that undergo imputation tend to have item response rates of more than 99 percent. See Chapter 3 for details on PMN; see Table A.28 in Appendix A for details on rates of missingness for the health insurance variables.

10.3.1 Types of Health Insurance Variables, Imputation Set 1

The first four imputation-revised variables were created using the single RP/multiple PRD PMN type. There were no noteworthy deviations from this general approach. The four imputation-revised variables thus created were yes/no indicators of whether the respondent had health insurance from Medicaid or a state children's health insurance program (IRMCDCHP), Medicare (IRMEDICR), CHAMPUS (IRCHMPUS), or private health insurance (IRPRVHLT). The health insurance indicators were imputed in the following order: CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. Since the response rates were high for these variables, the single RP/multiple PRD type of PMN was appropriate for this imputation set.

---

[148] The health insurance method varies slightly from other methods used to impute variables for the NSDUH. First, it uses some uncommon constraints and covariates. For example, from the core demographics module, the SERVICE variable, an indicator of service in the United States armed forces, was used both as a covariate and in a likeness constraint related to CHAMPUS. Second, age groups were created to ensure reasonable domain sizes for both modeling and hot-deck steps. Finally, provisional imputation steps in imputation set 1 were not actually hot-deck steps but simple stochastic imputations based on predicted means, and in the final hot-deck step for these variables, different constraints were applied to different age groups.

10.3.1.1 Response Propensity Step

The response propensity models for imputation set 1 utilized the final analysis weight, ANALWT. All respondents were in the domain for this stage. No respondents received skip codes for the final imputation-revised variables IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. For the RP model, a domain member was considered an item respondent if all four base variables (CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN) were nonmissing. See Table D.60 in Appendix D for details of the covariates used in the RP models for these variables.

10.3.1.2 First Prediction Step (CAIDCHIP)

CAIDCHIP was a dichotomous variable; after imputation, its only values were "yes" and "no." Therefore, it was modeled using logistic regression and, in particular, using the RLOGIST procedure in SAS-callable SUDAAN. The single predicted mean used in the later hot-deck step was the predicted probability that the respondent received Medicaid or was covered by a state children's health insurance program.

10.3.1.3 First Provisional Stochastic Imputation Step (CAIDCHIP)

The provisional imputation step used for CAIDCHIP was not a hot-deck step, but a stochastic imputation. Each recipient was assigned a "yes" response for the intermediate imputation-revised variable INTCDCH with probability equal to the predicted mean from the model. This method was based on the idea of a centered PMN, as described in Singh, Grau, and Folsom (2004). As applied to single dichotomous variables like CAIDCHIP, the centered PMN approach reduces to a simple stochastic imputation that requires no donors and no constraints. This approach to imputation for categorical variables is further explored in the evaluation of imputation methods for the NSDUH report (Center for Behavioral Health Statistics and Quality, in press).

10.3.1.4 Analogous Prediction and Provisional Hot-Deck Imputation Steps (MEDICARE, CHAMPUS, and PRVHLTIN)

PRD models were fit for MEDICARE, CHAMPUS, and PRVHLTIN in the same manner as for CAIDCHIP, as described above, with a few notable deviations:

- For the MEDICARE variable, a single PRD model was fit for the 18-64 age group. This was done because (1) only a small proportion of respondents in these age groups had Medicare, particularly for the 18-25 age group, and (2) a respondent of working age could have received Medicare only if he or she was not working because of disability. This was true regardless of whether the respondent was aged 18 to 25 or 26 to 64.
- The CHAMPUS PRD models used a covariate that underwent a simplified imputation procedure. For respondents aged 18 or older, models included an indicator of whether the respondent had ever been in the military service, designated by an imputation-revised version of the edited variable SERVICE. The variable SERVICE generally had a very low level of missingness (one missing value in the 2014 survey). Because covariates in these models were not supposed to have any missing values, the missing

values in the SERVICE variable were randomly imputed as "yes" responses if the random number was greater than the mean value of SERVICE across all the other respondents, and imputed as "no" otherwise.

- PRVHLTIN did not undergo a provisional imputation step since it was the last variable in the set.

10.3.1.5 Final Hot-Deck Step

Details on the missingness patterns, constraints, and predictive mean vectors for the final hot-deck step of this first stage of imputation are available in Appendix E. This section explains the general ideas behind the hot-deck step applied to these variables.

There are 15 missingness patterns for the first stage of health insurance imputation.[149] No logical constraints were applied, because there were no logical relationships among the variables (or between the variables and any other NSDUH variables). The predictive mean vectors were also made simple because of the absence of logical relationships. Nonetheless, the likeness constraints were fairly sophisticated and sometimes varied across age groups, exploiting auxiliary variables from the demographics and income modules. The SERVICE variable was also used, but not an imputation-revised version of it; the constraint only applied when the item nonrespondent had a nonmissing value. The vast majority of item nonrespondents in imputation set 2 are typically handled on the first attempt to find a donor because (1) the item response rates are high, (2) the domain includes all unit respondents, and (3) the constraints are not very restrictive.

10.3.2 Any Other Health Insurance Variable, Imputation Set 2

The final imputation-revised variable, IROTHHLT, indicated whether respondents had any type of health insurance, even though they reported or were imputed to have none of the four types of specific health insurance, as recorded by IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. IROTHHLT was created from the base variable ANYOTHER. The PMN imputation type used for this variable was single RP/single PRD, and there were no noteworthy deviations from this general approach.

For this second stage, the age groups were 12 to 17, 18 to 25, and 26 or older. Three age groups were used instead of four because of the small number of respondents who would have otherwise comprised the domain for the age group of 65 or older.

10.3.2.1 Response Propensity Step

The response propensity models for imputation set 2 utilized the final analysis weight, ANALWT. The domain for the ANYOTHER variable included respondents who had either a reported or imputed "no" value to all four imputation-revised specific health insurance variables from the first stage (IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT). For a domain member to be considered an item respondent, he or she had to have complete data for the

---

[149] Each of the four variables in the set can be missing or nonmissing. This leads to 2 × 2 × 2 × 2 = 16 possible combinations, but the combination where none are missing is one that does not require imputation.

variable ANYOTHER. See Table D.60 in Appendix D for details of the covariates used in the RP model for this variable.

10.3.2.2 Prediction Step

Like all the other health insurance variables which undergo imputation, ANYOTHER was a dichotomous variable; after imputation, its only values are "yes" and "no." Like the variables in the first stage, it was modeled using logistic regression, as implemented by the RLOGIST procedure in SAS-callable SUDAAN. The single predicted mean that was output by the RP step and used in the later hot-deck step was the predicted probability that the respondent received other health insurance, given that he or she was not covered by Medicaid/CHIP, Medicare, CHAMPUS, or private health insurance.

10.3.2.3 Hot-Deck Step

The hot-deck step for the ANYOTHER variable was the simplest one used in the NSDUH. There were no logical constraints, and the only likeness constraint was the delta constraint. The predictive mean vector was actually a scalar. Approximately 1 percent of the records in the domain underwent imputation, and nearly every item nonrespondent was handled on the first attempt to find a donor.

10.3.2.4 Recodes for Additional Analyses

The overall health insurance variable, IRINSUR4, was created by combining IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT. If a respondent had a reported or imputed "yes" value for any of these five variables, the respondent was considered to have health insurance. Otherwise, he or she did not have health insurance. Though IRINSUR4 was technically a recoded variable created from other variables, an imputation indicator (IIINSUR4) was nevertheless created. IIINSUR4 was set to "1" if the respondent had a reported "yes" value for any of the five health insurance variables or a reported "no" for all five of them; and "3" otherwise.

This page intentionally left blank

11. Editing and Imputation for the NSDUH Pair Variables

11.1 Introduction

Each household selected for the National Survey on Drug Use and Health (NSDUH) has zero, one, or two household members selected for interviewing. When two members of the same household, a pair, are selected, the data can be used to study outcome variables based on the relationship between the pair (i.e., the pair relationship). The outcome variables can be at either the person level or the pair level. The most common type of analysis is the person-level analysis, where the inferential population is defined by one of the pair members (called the focus pair member). An example of an outcome at the person level is the proportion of children who use drugs and whose parents report talking to them about drugs. The focus is on the child in a parent-child pair. By contrast, an example at the pair level is parent-child drug behavior for all possible parent-child pairs (within the child's age group). This chapter describes the techniques used to edit and impute the pair variables from the household roster (hereafter, "pair variables") for the 2014 NSDUH.[150] The variables described in this chapter can be divided into three stages:

- Stage One: Pair Relationships,
- Stage Two: Multiplicity Counts, and
- Stage Three: Household-Level Person Counts.

The creation of the edited and imputation-revised pair variables was conducted in three stages because the variables from earlier stages were needed for the creation of variables in later stages. Stage one consisted of the creation and imputation of the variables that identify the pair relationships. The multiplicity and household-level person counts were created and imputed in stages two and three, respectively. Missing values in all three stages were imputed using the predictive mean neighborhood (PMN) imputation procedure, which uses predicted means from models to find donors in a nearest neighbor hot deck. Chapter 3 of this report provides background information about imputation in general (including hot-deck imputation) and details about the PMN methodology in particular.

Though this chapter presents first the editing procedures applied to the variables in each stage and then the imputation procedures for each stage, it is important to note that the actual order of processing was by stage; that is, both editing and imputation were completed for the variables in each stage before moving on to the next stage.

[150] Edits to the variables for the parenting experiences module, which also involves selection of a pair, are discussed in Section 7.4.12 in Chapter 7. The parenting experiences module was administered if (1) an adult and an adolescent pair were selected for an interview, and (2) the adult was the parent of the adolescent who also was selected for an interview.

11.1.1 Pair Relationship Variables

The pair relationship variables were derived from the household composition (roster) variables, described in Chapter 8 of this report. Pair relationship variables include the edited and imputation-revised pair relationship variables PAIRREL[151] and IRPRREL, as well as the imputation indicator IIPRREL, which summarizes how the data in IRPRREL were obtained. PAIRMEM was created to indicate that a respondent was part of a respondent pair. RELMATCH was created to detail how well the pair members matched on their relationship to one another. The variable PRNTIND identified whether the respondent was a parent in a parent-child relationship. Finally, three additional variables were created to aid in pair analyses by giving details about the other pair member: AGEOTHER, SEXOTHER, and PAIRID. AGEOTHER contained the age of the other respondent in the pair; SEXOTHER contained the gender of the other respondent in the pair; and PAIRID contained the questionnaire ID (QUESTID) of the other pair member.

11.1.2 Multiplicity Count Variables

"Multiplicity" is the term used to describe the complication that arises in the analysis of pair data at the person level for a given pair domain and several pairs in the household could be associated with the same person. Whereas, when analyses are at the pair level, the pair domain is completely and uniquely defined by the pair relationship. For example, to tabulate the number of sibling-sibling pairs where both siblings have used marijuana, it is necessary to know only whether a pair of respondents contains two siblings. By contrast, for person-level analyses, the pair domain depends upon which pair member is the focus. In this case, to analyze the influence that older siblings have on younger siblings in terms of drug use, it is necessary to know which pair member is the older sibling. The multiplicity count is a count of the number of pairs in the household that can be associated with the person of focus because—to continue the example above—a child may have more than one older sibling. If the multiplicity is not considered, then individuals of focus who are members of more pairs receive a weight that is too large (i.e., they are counted too many times in the target population), and individuals of focus who are members of fewer pairs receive a weight that is too small. Adjustment for multiplicity was not done a priori because that would lead to the addition of 12 weights to the file, one for each of the 12 possible pair relationships that require this adjustment (e.g., excluding the spouse-spouse and spouse-spouse with children relationships). Note that the multiplicity problem does not arise if there is only one inclusion possibility (e.g., a single-parent household, if the child is the focus) or if the analysis is a pair-level analysis (e.g., parent-child pair drug behavior).

To illustrate how multiplicities appear in the definitions of parameters and estimates, consider estimating the total number of children who used drugs in the past year, where a parent reported talking to them about drugs. Let $y_{hip}(d)$ be defined as the drug-related behavior outcome for pair *p* containing the individual *i* belonging to domain *d* in household *h*. Now, for the population of all individuals who belong to the domain *d*, the total parameter is defined as

[151] The levels of PAIRREL are provided in Table 11.3.

$$\tau_y(d) = \sum_{h=1}^{H} \sum_{i=1}^{N_h(d)} \sum_{p=1}^{M_{hi}(d)} \frac{y_{hip}(d)}{M_{hi}(d)}$$

(i.e., total of averages over pairs (*p*) associated with the individual *i*, over all *i* in domain *d* and in the household *h*) (Chromy & Singh, 2001). Here $M_{hi}(d)$ denotes the multiplicity (i.e., the number of pairs associated with the individual *i* in domain *d*), and $N_h(d)$ can be thought of as the multiplicity count for the household *h* (i.e., the number of people in the household that are in domain *d*).

Multiplicity count variables were only created for specific relationships of interest. These variables are listed in Table 11.1. With the exception of the spouse-spouse multiplicity variables listed at the bottom of this table, these variables all underwent imputation.

Table 11.1 Edited and Imputation-Revised Multiplicity Count Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of parents of a child aged 12-14 who is a member of a parent-child pair	MCPCC14	IRMPCC14
Number of children aged 12-14 belonging to a parent who is a member of a parent-child pair	MCPCP14	IRMPCP14
Number of parents of a child aged 15-17 who is a member of a parent-child pair	MCPCC57	IRMPCC57
Number of children aged 15-17 belonging to a parent who is a member of a parent-child pair	MCPCP57	IRMPCP57
Number of parents of a child aged 12-17 who is a member of a parent-child pair	MCPCC17	IRMPCC17
Number of children aged 12-17 belonging to a parent who is a member of a parent-child pair	MCPCP17	IRMPCP17
Number of parents of a child aged 12-20 who is a member of a parent-child pair	MCPCC20	IRMPCC20
Number of children aged 12-20 belonging to a parent who is a member of a parent-child pair	MCPCP20	IRMPCP20
Number of siblings aged 12-14 for a respondent aged 15-17 who is a member of a sibling-sibling pair	MCS1417	IRMS1417
Number of siblings aged 15-17 for a respondent aged 12-14 who is a member of a sibling-sibling pair	MCS1714	IRMS1714
Number of siblings aged 12-17 for a respondent aged 18-25 who is a member of a sibling-sibling pair	MCS1725	IRMS1725
Number of siblings aged 18-25 for a respondent aged 12-17 who is a member of a sibling-sibling pair	MCS2517	IRMS2517
Pair relationship is spouse-spouse with no children younger than 18	MCSPSP	N/A
Pair relationship is spouse-spouse with children younger than 18	MCSPSPWC	N/A

N/A = not applicable.

11.1.3 Household-Level Person Count Variables

The household-level person count variable is the number of individuals of focus in the household for a given pair domain, regardless of which pair (or whether a pair) was actually selected. For example, if two parents were in the household with three children aged 12 to 14, the household person count for the *parent* focus of the parent-child (12-14) domain would be 2, and the household person count for the *child* focus would be 3. If the parents indicated that they had a spousal relationship, the household person count for the spouse-spouse with children domain would be 2. The rest of the household counts would be zero (even the sibling focus) because they do not fall in the age groups of the domains that are edited and imputed; that is, counts were not created for siblings that were both aged 12 to 14. Also, note that household person counts for all domains were calculated for every respondent, even when only one respondent was selected in the household or when a selected pair did not fall in a particular domain. The household-level person count variables are listed in Table 11.2, and all underwent imputation.

Table 11.2 Edited and Imputation-Revised Household-Level Person Count Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of children aged 12-14 in the household with at least one parent living with them	HCPCC14	IRHPCC14
Number of parents in the household with at least one child aged 12-14 living with them	HCPCP14	IRHPCP14
Number of children aged 12-17 in the household with at least one parent living with them	HCPCC17	IRHPCC17
Number of parents in the household with at least one child aged 12-17 living with them	HCPCP17	IRHPCP17
Number of children aged 12-20 in the household with at least one parent living with them	HCPCC20	IRHPCC20
Number of parents in the household with at least one child aged 12-20 living with them	HCPCP20	IRHPCP20
Number of household members aged 15-17 with a sibling aged 12-14 living with them	HCS1417	IRHS1417
Number of household members aged 18-25 with a sibling aged 12-17 living with them	HCS1725	IRHS1725
Number of spouse-spouse pairs without children younger than 18	HCSPSP	IRHCSPSP
Number of spouse-spouse pairs with children younger than 18	HCSPSPWC	IRHCSPWC

11.2 Stage One Editing: Pair Relationships

11.2.1 Editing the Household Roster of Each Pair Member

Prior to identifying the relationships between selected pair members, a key step is editing the questionnaire household rosters for each pair member. This involves identifying situations where the relationship listed in the roster for a particular roster member was not possible given the roster member's age and relationship to the respondent. In the majority of cases where the relationships could not be determined, this resulted in setting the relationship code to bad data and sometimes setting the roster member's age to bad data as well. In general, no effort was made to try to match the values of roster-derived household composition variables between pair members, because interviews of the different members of the same household could have taken place at different times. However, information from other pair members was sometimes used to change a relationship code from one value to another, instead of setting the relationship code to bad data.

11.2.2 Creating the Pair Relationship Variable (PAIRREL)

The process of identifying the pair relationships involved two steps: (1) match the household rosters of the pair members, and (2) determine the pair relationships using the relationship codes and ages of the matched rosters, if they could be determined. The first step is described in Section 11.2.2.1 and Appendix H, and the second step is described in Section 11.2.2.2 and Appendix H. Any relationships that could not be determined by this process were imputed as described in Section 11.5.2.

No restrictions were placed on the types of pairs that could be selected for inclusion in the NSDUH sample. However, the variable PAIRREL includes a limited number of relationships. This is because the identification of the particular relationships between a given pair was limited by the relationship codes that were available: parent, child, grandparent, grandchild, sibling, spouse, unmarried partner, roommate, parent-in-law, child-in-law, boarder, other relative, and other nonrelative. (This precluded the possibility of identifying an uncle-nephew relationship, for example.)

Furthermore, because the creation of the multiplicity factors required complicated programming logic, multiplicities could not be created for all possible pair relationships. Only certain pair relationships considered to be of higher analytic interest were included in the creation of multiplicities:

- parent-child, child aged 12 to 14;
- parent-child, child aged 12 to 17;
- parent-child, child aged 15 to 17;
- parent-child, child aged 12 to 20;
- sibling-sibling, younger sibling aged 12 to 14, older sibling aged 15 to 17;
- sibling-sibling, younger sibling aged 12 to 17, older sibling aged 18 to 25;

- spouse-spouse (includes partner-partner), with children younger than 18;[152] and
- spouse-spouse (includes partner-partner), with or without children.

The levels in PAIRREL, summarized in Table 11.3, do not correspond exactly with those given above, but the relevant pair relationships can be derived from the value of PAIRREL. For example, a value of PAIRREL = 3 indicates that, among the pair relationships given above, the pair relationship was a parent-child pair with a child aged 18 to 20.

Table 11.3 Levels of the Variable PAIRREL

Value of PAIRREL	Interpretation	Domain of Interest
1	Respondent is part of a parent-child (12–14) pair	Yes
2	Respondent is part of a parent-child (15–17) pair	Yes
3	Respondent is part of a parent-child (18–20) pair	Yes, indirectly
4	Respondent is part of a parent-child (21+) pair	No
5	Respondent is part of a sibling (12–14)-sibling (15–17) pair	Yes
6	Respondent is part of a sibling (12–17)-sibling (18–25) pair	Yes
7	Respondent is part of another sibling-sibling pair	No
8	Respondent is part of a spouse-spouse[1] pair, with children in the household younger than 18	Yes
9	Respondent is part of a spouse-spouse pair, with no children in the household younger than 18	Yes
10	Respondent is part of a spouse-spouse pair, but it is unclear whether children younger than 18 in the household belong to the pair	Yes
11	Respondent is part of a grandparent-grandchild pair	No
12	Respondent is part of another clearly identifiable pair	No
13	Respondent is part of a pair that is not clearly identifiable, but it is clear from the relationship codes that it is not within codes 1 through 11	No
14	Respondent is part of a pair that is not clearly identifiable, and it could be any pair relationship	Maybe
15	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 1 or 12	Maybe
16	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 2 or 12	Maybe
17	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 3 or 12	Maybe
18	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 4 or 12	No

[152] The spouse-spouse pair relationship included respondents who were legally married, as well as respondents who lived together as though married (partners). Although the questionnaire distinguished between "spouses" and "partners," the pair relationship variable being described here did not distinguish between the two. In rare instances, a spouse-spouse pair included one pair member who identified the second pair member as a spouse, whereas the second pair member identified the first as a partner.

Table 11.3 Levels of the Variable PAIRREL (continued)

Value of PAIRREL	Interpretation	Domain of Interest
19	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 5 or 12	Maybe
20	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 6 or 12	Maybe
21	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 7 or 12	No
22	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 8 or 12	Maybe
23	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 9 or 12	Maybe
24	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 8, 9, or 12	Maybe
25	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 11 or 12	No
99	Respondent is not a member of a pair	No

[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships.

11.2.2.1 Matching the Household Rosters

For the purpose of discussing how to match the household rosters in this report, the pair members are identified as pair member "A" and pair member "B." For the household roster of pair member A, it was necessary to determine which household member listed in A's roster corresponded to the other selected pair member. The same had to be done for pair member B. This was accomplished using the age and gender of the pair members, in addition to the variable MBRSEL, which was used to identify the roster member corresponding to the other selected pair member.

In a perfect setting, the questionnaire age and gender of pair member B (AGE and IRSEX, respectively) would have corresponded exactly to the age and gender entered for one of the members of pair member A's household roster (RAGE and RSEX). Moreover, the value of MBRSEL for this matched roster member would have been 1, and the value of MBRSEL for all other roster members would have been zero or missing. Furthermore, in this perfect setting, matches with exactly one MBRSEL = 1 correctly identifying the other pair member also would have been found with pair member B's roster. This did not always occur, however, so some effort was required to determine the roster member most likely to correspond to the other selected pair member.

The quality of the match between pair members varied depending upon the quality of the roster entries and the time between interviews. A number of if-then-else conditions, called priority conditions (because of the hierarchical nature of the conditions), gave a pair match that was considered valid in the vast majority of cases. These conditions are provided in Appendix H. In general, the conditions matched IRSEX and AGE for the one pair member against the gender and age of the roster members in the other pair member's roster, using MBRSEL to help identify the appropriate roster member. For a match to be considered valid, it was necessary that at least one of the two pair members have a match.

In the cases where a single roster member had to be selected among duplicates and where the duplicates had the same relationship code, it was necessary to limit the relationship codes to child or sibling. In some cases, because of the poor quality of the rosters of the pair members, it was not possible to locate the household member listed in A's roster that corresponded to pair member B, and vice versa. The determination of the pair relationships for these cases was left to imputation. Even when a pair of roster members was successfully identified, it was not always possible to successfully determine the pair relationship, as described in the next section.

11.2.2.2 Determining the Pair Relationship Using the Relationship Codes of the Matched Rosters

Once the pair was identified, two observations per household resulted, each with a relationship code corresponding to the other selected pair member. The relationship codes for these two observations had to be matched to determine the pair relationship. For example, suppose a 15-year-old and a 38-year-old were selected to be interviewed. If the 38-year-old was subsequently identified as the parent on the 15-year-old's roster, and the 15-year-old was identified as the child of the 38-year-old on the 38-year-old's roster, then the pair relationship would be identified as PAIRREL = 2, according to the levels of PAIRREL provided in Table 11.3. Thus, these two individuals would belong to the following pair relationships of interest: parent-child with child aged 15 to 17, parent-child with child aged 12 to 17, and parent-child with child aged 12 to 20. As noted earlier, the pair relationship of interest was derived from the values of PAIRREL. In particular, the parent-child with child aged 12 to 17 and parent-child with child aged 12 to 20 domains were derived from the parent-child pair relationships created using 12- to 14-year-olds, 15- to 17-year-olds, and 18- to 20-year-olds, the levels referenced in PAIRREL. Moreover, the overall spouse-spouse domain was derived from the two spouse-spouse pair relationships with and without children.[153]

As with the procedure used to match the household rosters, a series of conditions was used to identify the relationship between pair members. These priority conditions used ages and relationship codes to identify the pair relationships. In a perfect setting, the relationship codes would be nonmissing and in agreement between the pair members, as in the example given in the previous paragraph. In some instances, however, either the relationship codes were missing or they did not agree across the pair members. The priority conditions offer a method for interpreting the relationship codes in such cases.

Below are the strategies used to identify a pair relationship in an imperfect setting:

1. If a relationship code was missing on one side of the pair but not on the other, then the pair relationship was assumed to be identified by the nonmissing relationship code. The exception to this rule occurred if the identified relationship was parent-child with a child younger than 18, the "parent" was less than 10 years older than the child, and the "parent" answered the parenting experiences question (FIPE3) by

[153] The spouse-spouse pair relationship includes partner-partner pair relationships.

saying that the other respondent was not his or her child. In this case, the nonmissing relationship code was considered spurious, and the relationship was left missing.

2. If it was not possible to definitively determine the relationship between the pair members using the relationship codes, but the relationship codes on both sides indicated that the unknown pair relationship was not a relationship of interest, then the pair relationship was identified as such and no imputation was required. For example, if pair member "A" identified pair member "B" as a "boarder," but pair member "B" identified pair member "A" as "other relative," then the relationship was not a relationship of interest and the variable PAIRREL would have been assigned the value "13" (Table 11.3).
3. If it was not possible to definitively determine the relationship between the pair members using the relationship codes, but a parent-child relationship was possible given the relationship code in one of the pair member's rosters, then the FIPE3 variables were used to assist in the determination of a pair relationship. For example, consider a pair member who was a stepparent and refers to his or her stepchild as "child," but the child refers to the stepparent as "other nonrelative." Membership in a parent-child relationship where the child was younger than 18 was indicated if the stepparent answered FIPE3 affirmatively, thereby proceeding to the parenting experiences module. On the other hand, if the stepparent answered FIPE3 negatively, then the stepparent was not considered the parent. A third scenario arose if the FIPE3 answer was not given. In this case, a parent-child relationship was assumed if the stepparent was legally married and the child identified the spouse of the other pair member as "parent."

The quality of the match for PAIRREL levels 1 through 25 is indicated by the variable RELMATCH, the levels of which are summarized in Table 11.4. In general, imputation was required for values of RELMATCH of 0 or 4, or if PAIRREL = 10. PAIRREL = 10 was a special case because it was clear that a relationship "of interest" always would have been involved. For this value of PAIRREL, the value of RELMATCH was equal to 1 or 2. However, imputation was still required because it was not clear whether children were in the household. The number of cases that were matched or not matched, as indicated by the RELMATCH variable (or PAIRREL = 10), for the 2014 survey is provided in Table 11.5. The amount of imputation required was dependent upon the quality of the rosters. The attributes of the roster are described in Chapter 8.

Table 11.4 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH

Value of RELMATCH	Values of PAIRREL	Interpretation
0	14	FAILURE: The relationship was not identifiable and could have been a relationship of interest.
1	1–9, 11–13	SUCCESS: The relationship was clearly identifiable using information from both pair members or was unmistakably not a relationship of interest.
1	10	FAILURE: A spouse-spouse[1] relationship was definitively established using information from both pair members, but it was unclear whether the pair had children in the household.

Table 11.4 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH (continued)

Value of RELMATCH	Values of PAIRREL	Interpretation
1.5	8	SUCCESS: A spouse-spouse relationship was definitively established using information from both pair members, and children younger than 18 were in both rosters. Relationship codes on one side of the pair indicated children belonged to the pair, and on the other side of the pair, the relationship codes corresponding to the children were missing.
2	1–9, 11–13	SUCCESS: The relationship was clearly identifiable using information from one pair member, while the relationship code from the other pair member was missing.
2	10	FAILURE: A spouse-spouse relationship was definitively established using information from one pair member, while the relationship code from the other pair member was missing. It was unclear whether the pair had children in the household.
3	1–8, 12, 13	SUCCESS: Relationship information was conflicting between the pair members, but conclusions were drawn anyway for some parent-child pairs, some sibling-sibling pairs, and some spouse-spouse pairs using either information outside the household roster or logical reasoning.[2]
4	15–25	FAILURE: Relationship information was not identifiable. Information was in conflict between the pair members, where one pair member indicated relationship of interest and the other did not. However, ages supported a relationship of interest (may be used to limit imputation).

[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships.

[2] In the case of potential parent-child pairs, further evidence that a parent-child relationship was involved or not involved was obtained by looking at the FIPE3 variable to see whether a stepparent had a spouse who corresponded to a child's parent or to see the ages of the respondents. For spouse-spouse relationships, two situations occurred. In the case where the respondents were not legally married, the children of one pair member were considered the children of the pair in the household, even though they were not identified as such by the other pair member. In the case where only one pair member referred to the other as a "married" or "unmarried partner," and if both had the same children, they were considered "spouse-spouse-with-children." The other pair member was usually referred to as a "roommate" or "other nonrelative."

Table 11.5 Frequencies of the Levels of the Variable RELMATCH: 2014

RELMATCH	Frequency (Percent)	Requires Imputation
0	32 (0.18)	Yes
1 (PAIRREL ≠ 10)	17,798 (97.64)	No
1 (PAIRREL = 10)	26 (0.14)	Yes
1.5	1 (0.01)	No
2 (PAIRREL ≠ 10)	101 (0.55)	No
2 (PAIRREL = 10)	0 (0.00)	Yes
3	119 (0.65)	No
4	152 (0.83)	Yes

11.3 Stage Two Editing: Multiplicity Counts

As stated earlier, multiplicities were required to account for analyses that were performed at the person level, even though the pair weights were calculated at the pair level. Because the multiplicities were relevant only at the person level, the definition of multiplicity required the identification of the focus member of the pair. Using the pair relationships determined in Section 11.2, the following domains were considered:

1. parent-child (child 12 to 14), parent focus;
2. parent-child (child 12 to 14), child focus;
3. parent-child (child 15 to 17), parent focus;
4. parent-child (child 15 to 17), child focus;
5. parent-child (child 12 to 17), parent focus;
6. parent-child (child 12 to 17), child focus;
7. parent-child (child 12 to 20), parent focus;
8. parent-child (child 12 to 20), child focus;
9. sibling (12 to 14)-sibling (15 to 17), sibling (15 to 17) focus;
10. sibling (12 to 14)-sibling (15 to 17), sibling (12 to 14) focus;
11. sibling (12 to 17)-sibling (18 to 25), sibling (18 to 25) focus;
12. sibling (12 to 17)-sibling (18 to 25), sibling (12 to 17) focus;
13. spouse-spouse (includes partner-partner) with children younger than 18; and
14. spouse-spouse (includes partner-partner).

Determining the multiplicity entailed finding the number of roster pairs in the domain of interest that contained the focus member in the pair. In broad terms, the process of determining the multiplicity count involved two steps: (1) calculate the multiplicity count for each pair member, and (2) use the screener, quality of roster, and other means to determine the appropriate count if each pair member's counts did not match. The first step is described in Section 11.3.1, and the second step is described in Section 11.3.2 and Appendix I. Multiplicities that could not be determined through these steps were left to imputation as described in Section 11.5.3.

Because the pair weights reflect selection done at the time of screening, the multiplicity count should reflect the household makeup at that time. However, this was not possible in all cases, because for some households the screener roster was not as complete as the questionnaire roster, and recorded relationships in the screener roster were relative to the head of the household rather than to each pair member. No account was made for cases where a change in the household makeup occurred between the screening date and the dates of both interviews. In other words, due to the passage of time only, the observed change in household makeup could have occurred because of an intervening birthday or because a roster member left or entered the household after screening. Because an adjustment to the multiplicity counts would have been extremely complicated to implement for the small number of cases to which it applied, no such adjustment was made. Nevertheless, when possible, the screener was used to resolve cases where there were disagreements between pair members on the value of the multiplicity count.

11.3.1 Determining the Multiplicity Count for Each Pair Member

The multiplicity counts for each pair member consisted of a direct count and an indirect count. The direct count was obtained by looking at the focus pair member. It was a count of the roster members who could have been selected, where the same pair domain would have resulted. The indirect count was obtained by looking at the pair member who was not the focus. It was a count of the pair member himself or herself, plus other roster members who, by virtue of their relationship code, would have had the same pair relationship had they been selected.

A summary of the ways of determining the direct count and indirect count for each pair domain are provided in Table 11.6. For these domains, neither the direct nor the indirect count could be zero, because the pair member who was not the focus had to be part of the count. For spouse-spouse counts, no work was necessary to determine multiplicity counts. If a respondent was in a spouse-spouse pair, the multiplicity count was necessarily 1 in almost all cases, because only one spouse-spouse pair could have been selected that included that pair member. If the true multiplicity count exceeded 1, then the multiplicity count was set to 1.[154] Note that other spouse-spouse pairs in the household (e.g., one spouse's parents) would have been of interest in the household counts discussed in subsequent sections.

Table 11.6 Multiplicity Counts for Each Pair Member

Pair Relationship	Focus Member	Direct Count	Indirect Count
Parent-Child	Child	From child: number of parents	From parent: self + spouse/partner
Parent-Child	Parent	From parent: number of children in appropriate age range	From child: self + number of siblings in the appropriate age range
Sibling-Sibling	Older sibling	From older sibling: number of siblings in younger age range	From younger sibling: self + number of siblings in younger age range
Sibling-Sibling	Younger sibling	From younger sibling: number of siblings in older age range	From older sibling: self + number of siblings in older age range

11.3.2 Determining the Final Multiplicity Count

Once the counts were determined for each pair member, it was necessary to resolve any differences between these counts across pair members. In most cases, the direct and indirect counts agreed, with no bad relationship codes for either pair member, resulting in a straightforward determination of the final multiplicity count. This was usually possible if one pair member had bad relationship codes or had a count of zero, which meant that the final

[154] In rare cases, it was possible for a respondent to have two or more spouses who lived in the same household. Determining the appropriate multiplicity count in these cases required knowledge of which spouse was the focus, which would be arbitrary. Because having multiple spouses was an extremely rare occurrence, and because of the complexity of determining the appropriate multiplicity count, these situations were not explicitly addressed during data processing.

multiplicity count came from the pair member with good data.[155] For some cases, both pair members had bad relationship codes, which meant that the final multiplicity was left to imputation. Among the remaining cases, some could be reconciled and some could not. In the cases where reconciliation was possible, many of the disagreements between the pair members were resolved by going to the screener. The method used to reconcile differing counts depended upon the domain. For the parent-child domains, for example, in addition to the screener, the parent/legal guardian variable, FIPE3, was used to help reconcile differences. Detailed rules for reconciling differences between pair members are provided in Appendix I.

If reconciliation between the counts from the two pair members in the household and the screener was not possible, upper and lower bounds within which the imputed value had to reside were determined from the counts for each pair member and the counts for the screener. The amount of imputation required for the multiplicity counts for the 2014 survey is shown in Table 11.7. From this table, it is apparent that the greatest degree of uncertainty came with the determination of the number of parents in the child-focus parent-child domains. This occurred because, even though the parent-child pair relationship had been established, it often was unclear whether there was a second "parent" in the household.

Other domains had very little uncertainty. The counts of the number of children in the parent-focus parent-child domain, for example, were almost always definitively determined.

Table 11.7 Amount of Imputation Required for Multiplicities in Various Pair Domains: 2014

Pair Domain	Multiplicity	Missing Cases
Parent-Child (12–14), Child Focus	Number of parents	162
Parent-Child (12–14), Parent Focus	Number of children	58
Parent-Child (15–17), Child Focus	Number of parents	150
Parent-Child (15–17), Parent Focus	Number of children	66
Parent-Child (12–17), Child Focus	Number of parents	248
Parent-Child (12–17), Parent Focus	Number of children	56
Parent-Child (12–20), Child Focus	Number of parents	292
Parent-Child (12–20), Parent Focus	Number of children	62
Sibling (12–14)-Sibling (15–17), Older Sibling Focus	Number of younger siblings	64
Sibling (12–14)-Sibling (15–17), Younger Sibling Focus	Number of older siblings	66
Sibling (12–17)-Sibling (18–25), Older Sibling Focus	Number of younger siblings	76
Sibling (12–17)-Sibling (18–25), Younger Sibling Focus	Number of older siblings	74

11.4 Stage Three Editing: Household-Level Person Counts

In order to improve the quality of the estimates from the pair data through poststratification of the appropriate weights (Center for Behavioral Health Statistics and Quality, 2016b), it was necessary to identify the household-level person counts for each domain. This

[155] There were some exceptions to this rule. If the bad relationship codes were only within the relevant age ranges, then the count from the good side was used only if the age ranges in the good side matched the screener.

entailed finding the number of individuals in the household who belonged to a particular domain, given that one member of a domain was selected as the focus. These counts were more difficult to derive than the multiplicity counts because all households were considered. Within each household, counts for any of the domains of interest were derived, regardless of whether the pair belonged to that domain or even whether a pair was selected at all. The counts were derived for 10 of the 14 pair domains described in Section 11.3. For the parent-child counts where the child was between 15 and 17, calculating the household counts was unnecessary.[156] For the other two remaining sibling-sibling domains, the reason is historical: They were added after the procedures were first developed, and there was insufficient time to develop the household counts for those domains. The domains where these counts were created are listed below:

1. parent-child (child 12 to 14), parent focus;
2. parent-child (child 12 to 14), child focus;
3. parent-child (child 12 to 17), parent focus;
4. parent-child (child 12 to 17), child focus;
5. parent-child (child 12 to 20), parent focus;
6. parent-child (child 12 to 20), child focus;
7. sibling (12 to 14)-sibling (15 to 17), sibling (15 to 17) focus;
8. sibling (12 to 17)-sibling (18 to 25), sibling (18 to 25) focus;
9. spouse-spouse (includes partner-partner) with children younger than 18; and
10. spouse-spouse (includes partner-partner).

Determining the household-level person counts was a two-step process: (1) calculate the household count for each respondent, whether a member of a pair or a single respondent; and (2) use the screener, quality of roster, and other means to determine the appropriate final count either by attempting to reconcile differing counts between pair members or by attempting to determine the appropriate count when information from only one roster was available. For households where only one respondent was selected, the matching step (step 2) was unnecessary. The first step is described in Section 11.4.1, and the second step is described in Section 11.4.2 and Appendix J. Household counts that could not be determined by this process were left up to imputation as described in Section 11.5.4.

Because the pair weights reflected selection at the time of screening, the household-level person counts should have reflected the household makeup at that time. As with the multiplicity counts, however, this was not entirely possible, so no account was made for cases where a change in the household makeup occurred between the screening time and the time of each interview. An explanation for why this was not possible for the multiplicity counts is described in Section 11.3. Moreover, as stated in that section, to implement such an adjustment would have

[156] Because household counts were defined for everybody, it was possible to derive these counts using the counts for the parent-child domains where the child was between 12 and 14 and where the child was between 12 and 17. However, the multiplicity counts for the parent-child (15 to 17) domain had to be calculated and could not have been derived in as straightforward a way. This was because the multiplicity counts were only defined if the pair relationship corresponded to the pair domain of interest.

been extremely complicated for the household-level person counts. Nevertheless, in cases where there were disagreements between pair members on the value of the household-level person count, the screener was used to resolve those disagreements.

11.4.1 Determining the Household-Level Person Count for Each Respondent

11.4.1.1 Parent-Child Domains

When obtaining household-level person counts for parent-child domains, the six parent-child domains previously listed were considered. In any household, the household-level person counts for parent-child domains were nonzero if at least one parent was present in the household with children within the relevant age range. In this instance, the child-focus count would have been the number of children in the household within that age range that belonged to the parent in the household, and the parent-focus counts would have been the number of parents. If more than one "family unit" (mother and/or father with children) lived within the household, the child-focus counts should have counted children from more than one set of parents, and the parent-focus counts should have counted two or more parents, at least one for each set of children.

One situation where this could occur was where three generations lived within the same household, with children in both the youngest and the second generations within the relevant age range. Using the youngest generation as the reference point, some of the parent's siblings (the grandparents' other children) were within the relevant age range. In this instance, the parent-child domains of the number of children would have included both the children of the parents and the children of the grandparents who were in that age range. The count of the number of parents included both the parents and the grandparents (and exceeded 2). Identifying more than one family unit in a household with children within the relevant age range under other scenarios (e.g., two sisters both with children within the relevant age range, both living within the same household) could not be determined from the data and had to be disregarded. Regardless of how many family units were in the household, counts had to be determined in different ways depending upon whether a parent-child pair "of interest" was selected or not.

Descriptions of how to obtain the household-level person counts are provided below for the parent-child domains outlined above. Parent-child pairs of interest with parent-focus and child-focus domains considered together are described first. In this instance, the pair actually belonged to a pair relationship where analysis using one or more of the domains listed was possible. This is followed by descriptions for other pairs and single respondents with parent-focus and child-focus domains considered separately.

11.4.1.1.1 Obtaining Counts for Parent-Child Domains (Parent-Focus and Child-Focus): Parent-Child Pairs, Child Younger than 21

If the pair was identified as parent-child and the three-generation situation described above was not apparent, the household-level child-focus person count was given by the parent-focus multiplicity count. Similarly, the household-level parent-focus person count was given by the child-focus multiplicity count. If a three-generation situation was identified and the grandparent also had children within the relevant age range, the number of children and the number of parents were adjusted appropriately. The final household count in this instance was greater than the imputation-revised multiplicity count, which did not include all of the children in the household within the relevant age range.

11.4.1.1.2 Obtaining Counts for Child-Focus Parent-Child Domains: Other Pairs and Single Respondents

For other pairs[157] and single respondents, the following conditions were required to determine the household count for the number of children of parents in the household:

1. If the age of the respondent was within the relevant age range and that respondent had at least one parent, then the child-focus counts were determined in the same way as the parent-focus multiplicity counts: The count was of the "self" plus the respondent's siblings within the relevant age range. If the respondent's parents were not identified as living with him or her in the household, then the count was set to zero.
2. If the respondent had children within the relevant age range, then the count was of the respondent's children within that range. If the respondent also had older children who had children of their own within the relevant age range, then the count was of the respondent's children and grandchildren within the relevant age range.
3. If the age of the respondent was outside the relevant age range, but the respondent had parents living with them in the household and had siblings within the relevant age range, then the count was of the number of the respondent's siblings.
4. If the respondent had grandchildren within the relevant age range and the respondent also had children older than 25 or children-in-law living with them, then the count was the number of the respondent's grandchildren. The assumption was that the respondent's children or children-in-law were the parents of the respondent's grandchildren. The likelihood of this not being the case was small. In the case where a pair was selected, this was resolved by looking at the count of the other pair member.

11.4.1.1.3 Obtaining Counts for Parent-Focus Parent-Child Domains: Other Pairs and Single Respondents

For other pairs and single respondents, the following conditions were required to determine the household count for the number of parents of children in the household:

1. If the age of the respondent was within the relevant age range, then the count was of the number of the respondent's parents (which could be zero).
2. If the age of the respondent was outside the relevant age range but the respondent had siblings within the relevant age range, then the count was of the number of the respondent's parents (again, this could be zero).

---

[157] "Other pairs" included pairs that were not within a domain of interest because the age of at least one of the pair members was outside the relevant age range. For parent-child pairs, this applied to a pair with a child who was 21 or older. For sibling-sibling pairs, this applied to siblings where both were within the same age range (both were 12 to 14, 15 to 17, or 18 to 25) or at least one of the siblings was older than 25. "Other pairs" also are referenced in Sections 11.4.1.1.3 and 11.4.1.2.2.

3. If the respondent had children within the relevant age range, then the parent-focus counts were determined in the same way as the child-focus multiplicity counts: The count was of the self plus the spouse or unmarried partner. If the respondent also had older children (older than 25 and living with him or her) who had children of their own (identified as grandchildren) within the relevant age range, then the count was at least two. If the respondent had a spouse or unmarried partner in the household, then the count was incremented by one, and if a child-in-law was in the household, then the count also was incremented by one. (Note that, under these scenarios, the number of parents could range between two and four.)
4. If the respondent had grandchildren within the relevant age range but no children in that range, and the respondent had a child older than 25 or a child-in-law living with them, then the count was 2 if both the child older than 25 and the child-in-law were living in the household, and the count was 1 if not.

11.4.1.2 Sibling-Sibling Domains

When obtaining household-level person counts for sibling-sibling domains, only the two sibling-sibling domains previously listed were considered. As with the parent-child counts, the household-level person counts for sibling-sibling domains were nonzero if at least one sibling-sibling pair was present in the household within the relevant age ranges, in which the count was the number of appropriately aged siblings. If sets of siblings from more than one "family unit" (sets of siblings from different parents) resided within the same household, the sibling-sibling counts should have counted possible pairs from within each set. However, sets of siblings that did not involve the respondent's family unit could not have been identified from the data. Regardless of how many sets of siblings were in the household, counts had to be determined in different ways depending upon whether a sibling-sibling pair "of interest" was selected or not.

Descriptions of how to obtain the household-level person counts are provided below for the sibling-sibling domains outlined above. Sibling-sibling pairs of interest are described first. In this instance, the pair actually belonged to a pair relationship where analysis using one or more of the domains listed was possible. This is followed by descriptions for other pairs and single respondents. In each case, the descriptions apply regardless of which sibling-sibling domain was considered.

11.4.1.2.1 Obtaining Counts for Sibling-Sibling Domains: Sibling-Sibling Pairs of Interest

If the pair was identified as sibling-sibling within a relevant domain, the multiplicity count was the number of younger siblings because the older sibling was the focus. The household-level sibling-sibling person counts were determined in a similar manner to the multiplicity count, except that the count of interest was for the number of older siblings. If the pair member was the older sibling, then the household count was the self plus the number of siblings in the older age range. The count for the younger sibling pair member was the number of siblings within the same older age range. Unlike the case with the parent-child household-level counts, inconsistencies in the sibling-sibling counts when the pair selected was sibling-sibling still needed to be resolved. However, the rules for resolving inconsistencies followed directly from those used for the multiplicity counts when counting the number of younger siblings

(Appendix I). Note that a pair that was within one sibling-sibling pair domain had to be outside the other sibling-sibling pair domain.

11.4.1.2.2 Obtaining Counts for Sibling-Sibling Domains: Other Pairs and Single Respondents

For other pairs and single respondents, the following conditions were required to determine the household count for the number of siblings within the older age ranges of the domains of interest in the household:

1. If the age of the respondent was within the age range of the older sibling and that respondent had at least one sibling in the younger age range, then the count was the self plus the respondent's siblings within the older age range. If the respondent did not have any siblings within the younger age range, then the count was set to zero.
2. If the age of the respondent was within the age range of the younger sibling and that respondent had at least one sibling in the older age range, then the count was the number of the respondent's siblings in the older age range.
3. If the age of the respondent was outside the age range of the older or younger sibling but had at least one sibling in each of the older and younger age ranges, then the count was the number of siblings in the older age range.
4. If the age of the respondent was outside the age range of the older or younger sibling but the respondent had children within both the older and the younger age ranges, then the count was set to the number of respondent's children in the older age range.
5. If the age of the respondent was outside the age range of the older or younger sibling but the respondent had grandchildren within both the older and younger age ranges, then the count was the number of grandchildren in the older age range. If the respondent's grandchildren were cousins rather than siblings, then there was no way of deciphering this from the data. This had to be resolved by looking at the information from the other pair member, if another pair member was selected.

11.4.1.3 Spouse-Spouse Domains

What is referred to as a "spouse-spouse domain" was actually derived from spouse-spouse and partner-partner pair relationships. The following conditions were required for the number of spouse-spouse (including partner-partner) pairs to be incremented by 1. Some of these conditions were applied to the same household:

1. The respondent was part of a spouse-spouse (or partner-partner) pair.
2. The respondent was not part of a spouse-spouse pair but had a spouse (or unmarried partner).
3. The respondent had two parents living in the house.
4. The respondent had two parents-in-law living in the house.

5. The respondent had two grandparents living in the house.
6. The respondent had a child and a child-in-law living in the house.

The following conditions were required for the number of spouse-spouse pairs with children younger than 18 to be incremented by one. (These also include partner-partner pairs with children younger than 18.) Some of these conditions were applied to the same household:

1. The respondent was part of a spouse-spouse (or partner-partner) pair with children younger than 18.
2. The respondent was not part of a spouse-spouse pair[158] but had a spouse (or unmarried partner) and children younger than 18.
3. The respondent had two parents living in the house and was either younger than 18 or had siblings younger than 18.
4. The respondent had a child and a child-in-law living in the house and had grandchildren younger than 18.

11.4.2 Determining the Final Household-Level Person Count

For a particular type of household-level person count, there are three types of households from a sample selection perspective. For the first type, a pair was selected where the pair relationship corresponded directly to the pair domain being counted and both pair members responded. In this case, the household-level person count was usually easy to obtain using the multiplicity counts, although an adjustment was sometimes required if more than one family unit was in the household. For example, if a parent-child pair was selected where the child was 12 years old, the household-level person counts for the parent-focus parent-child (12-14) domain could usually be obtained from the multiplicity count that was calculated in stage two. In the second type of household, a pair also was selected and both pair members responded, but in this type the pair relationship did not correspond directly to the pair domain being counted. In this case, determining the final count was sometimes more difficult, particularly if one or more of the counts was a count of zero. A count of zero from a roster with good data did not necessarily mean that the final count should be zero. For example, suppose a household consisted of a man, his wife, brother, and two sons, and suppose one of the sons and his uncle (the man's brother) were selected. If the uncle's roster had a count of zero for all domains of interest—because all of the household members were "other relatives" from his perspective—then no nonzero parent-child count could be obtained. The final count would have to be determined from imputation. In the third type of household, only one respondent was selected. In this case, it was not possible to match counts from different pair members, but determining the final count could still be difficult if the count was zero for a household where the value was not truly zero.

For situations where a pair was selected and both pair members had good roster data, if the counts agreed between the pair members and were not zero, then a straightforward determination of the final household-level count was possible. This occurred in a majority of

[158]All spouse-spouse pairs were excluded here because spouse-spouse pairs with children were already accounted for, and spouse-spouse pairs without children had already been defined (possibly by imputation) not to have children younger than 18.

cases. If one pair member had a bad roster with no information in it and the other had a good roster, this was treated in the same way as if a single respondent was selected with a good roster. In either of these cases, the final count could be determined, provided a considerable number of conditions were satisfied. The conditions used to accept a good roster's count, when either the other pair member's roster was bad or no pair was selected, are provided in Appendix J. If these conditions were not met, the final household-level person count was left to imputation. Imputation also was required if two pair members were selected, both with bad rosters.

Among the remaining cases, some could be reconciled and some could not. In the cases where reconciliation was possible, some of the disagreements were caused by the pair members' rosters having different age and gender compositions. In these cases, many of the disagreements between the pair members were resolved by going to the screener. However, the screener did not provide much help if the age and gender composition of the pair members' rosters were identical, yet the counts still disagreed, as was the case with the uncle-nephew pair described above. In that example, one count was zero and the other was nonzero. Under conditions set out in Appendix J, it was possible to determine that the disagreement in this case was due to the uncle not being able to identify the parent-child domains, and the nonzero count was used. More detailed rules for reconciling differences between pair members are described in Appendix J.

If the attempt to reconcile differences in the household-level person counts between pair members was unsuccessful, upper and lower bounds within which the imputed value must reside were determined from the counts for each pair member and the counts for the screener.

11.5 Imputation for the Pair Variables

Imputation was required for variables in all three stages, and the imputation models for the pair variables required the inclusion of covariates at the pair level. The creation of these covariates is described in Section 11.5.1. The imputation process for stages one, two, and three are described in Sections 11.5.2 through 11.5.4, respectively. The final covariates for the fitted models are listed in Appendix D.

11.5.1 Creation of Covariates for Imputing Pair-Level Variables

Imputation was performed at the household level rather than at the respondent level. Thus, it was necessary to model covariates defined at the household level. Segment-level covariates were used for this purpose because they were automatically defined at the household level, using external information that was constant regardless of when the interviews were conducted. In addition to these segment-level covariates, information from the questionnaire would also have been useful as modeling variables. The logical choices for questionnaire-derived variables include the household composition variables IRHHSIZE (household size), IRKID17 (number in household younger than 18), IRHH65 (number in household aged 65 or older), and IRFAMSKP (presence of other family members in household indicator).

However, because interviews between pair members could have been conducted at different times, these variables were not necessarily consistent across pair members. Therefore, new count variables were needed that were consistent across the pair members (i.e., used screener information to reconcile disagreements between them) within a household. These variables were created in a two-step process: (1) create the count variables for each pair member, and (2) attempt to reconcile disagreeing values between pair members. The following sections describe these two steps in the creation of household size, household composition age count variables, and household composition age count variables "for males only," each of which were consistent across pair members. Note that household composition age count variables were not created for females because, for a given age range, the number of females could be obtained by subtracting the number of males from the total number within that age range. These variables also had to be created for respondents who were not part of a pair, for the purposes of creating and imputing the household-consistent person counts of various domains.

11.5.1.1 Household Size

The new variable created to represent a household size that was consistent across selected pair members was called HHSIZE and constructed as follows. First, the edited household size, TOTPEOP, was compared between pair members. If the values for TOTPEOP agreed across pair members and were both nonmissing and greater than 1,[159] then HHSIZE was set to that value. If the values of TOTPEOP disagreed across pair members because the count for one pair member was missing and the count for the other was not missing and was greater than 1, then a natural choice for HHSIZE was the valid, nonmissing value contributed by one of the pair members. If the values of TOTPEOP either (1) disagreed across pair members without a clear indication of which one was valid or (2) were both missing or equal to 1, then the tools used to determine the final value of HHSIZE included the reported and edited household size variables, QD54 and TOTPEOP, as well as other measures of household size and "quality of roster" measures. These "other measures" included the screener household size and two sums of total valid ages within a pair member's roster.

The first sum was a simple total count of the number of roster members with valid ages, obtained by summing the counts within age groups. The second sum counted the ages of the pair members as reported during questionnaire administration. The two sums differed if a roster count (the first sum) was less than the number of pair members in a given age category. For example, if a household roster had one 12- to 17-year-old, but two 12- to 17-year-olds were selected, then the value of HHSIZE would be increased by 1. An additional situation occurred where the household size counts could not be easily determined by looking at both pair members. If the counts for both pair members were missing, then the screener household size was used to define HHSIZE.

In still other cases, disagreement between pair members with regard to the true household size could not be easily resolved. The screener household size did not support either household size in these cases, and the age counts mentioned above also did not resolve the disagreement. A decision had to be made as to which pair member's household size should be believed. This decision depended upon the "quality of the roster," where the household size was determined by the pair member with a "better" roster quality. One obvious way to measure roster quality was by noting the number of cases where the ages, relationship codes, or genders were missing in the roster. If a roster was missing one or more of these three variables for some of the roster

---

[159] In households where a pair of respondents was selected, the household size had to be greater than or equal to two.

members, the roster was considered to be of "poorer quality" than a roster with these variables nonmissing for all roster members.

If only one household member was selected as a respondent, referred to as a "nonpair household," the rules for creating HHSIZE were the same as those that were used if two household members were selected in a pair but only one of the pair members had a nonmissing, acceptable value for a reported household size. Note that if only one household member was selected as a respondent, it was permissible to have a reported household size of one, whereas in a selected pair, a reported household size of one was considered "bad data."

In summary, the variables used to determine HHSIZE included (for each pair member) the reported and edited household sizes, the number of cases with valid ages in the roster, the number of cases with valid ages with the count in some age categories replaced by the minimum possible in that age category, and a roster quality count of the number of roster members with missing information. The screener household size, which was the same for each pair member, also was used. Using all of these tools, HHSIZE did not have any missing values in the 2014 survey, nor did it have any in surveys from previous years. General points about the creation of the household size variable are provided in Appendix K.

11.5.1.2 Household Composition Age Count Variables

It would seem logical to assert that the ages of other household members would be good covariates for the domain to which a pair might belong. Such variables also would be important for imputing multiplicity and household-level domain counts. The household-consistent age counts were limited to the following age ranges: younger than 12, 12 to 14, 15 to 17, 12 to 17, 12 to 20, 18 to 25, 26 to 34, 35 to 49, and 50 or older. These variables were called AGE011, AGE1214, AGE1517, AGE1217, AGE1220, AGE1825, AGE2634, AGE3549, and AGE50P, respectively.

The first step in this process was to count the nonmissing ages for roster members in the household for each pair member. In some cases, it was necessary to adjust the count because the ages could not be matched exactly. For example, suppose a 38-year-old and a 17-year-old were interviewed, and the 17-year-old was interviewed first. Suppose also that the 17-year-old turned 18 (i.e., had his or her 18^{th} birthday) before the 38-year-old was interviewed. Hence, the 17-year-old would have had an age of 18 in the 38-year-old's roster. However, because the younger pair member was 17 at the time of his or her interview, the ages of interest for this pair domain were defined to be 17 and 38. Hence, it was necessary to account for this by creating a new roster age variable that matched the age provided in the other pair member's questionnaire. The age counts using this new roster age variable were equivalent to subtracting 1 from the previously obtained 18-25 count and adding 1 to the previously obtained 12-17 count in the 38-year-old's roster. These adjustments were made for all cases where a match was made between one pair member's roster and another pair member's interview age and gender and the ages did not match exactly.

If no roster ages were missing, the sum of these counts was equal to the edited household size TOTPEOP. Note that the reported household size was not considered here, because the counts were obtained from an edited roster. As with household size, a series of priority conditions was used to obtain the most likely count within each age group. If the appropriate count was ambiguous because of disagreement between the pair members, the quality of the roster and the age of the respondent (in that order) were used to determine the appropriate count. The roster quality was determined by the number of bad or missing roster entries and the quality of the match between the pair member's roster and the other pair member's questionnaire age and gender.

If only one household member was selected as a respondent, the rules were the same as when two household members were selected in a pair but only one of the pair members had nonmissing data for the roster ages. One important exception to these rules was that when determining minimum possible counts for various age groups, it was not necessary to incorporate information from another pair member to increment the minimum for that pair member. General points about the creation of the age variables are provided in Appendix K.

11.5.1.3 Household Composition Age Counts of Males

For some pair variables, particularly spouse-spouse pairs, knowledge of the gender of the roster member was important for imputation purposes. In a similar manner to that used in the creation of the household composition age count variables, variables counting the number of males within the given age ranges were created. Disagreements between pair members were resolved in a similar manner to what was done with the household composition age count variables. The names of the male age counts were MALE011, MALE1214, MALE1517, MALE1217, MALE1220, MALE1825, MALE2634, MALE3549, and MALE50P.

11.5.2 Stage One Imputation: Pair Relationships

Missing pair relationships were filled in using the single response propensity (RP)/single prediction (PRD) type of the PMN method. In this stage, the imputation set involved only the edited variable PAIRREL, with its corresponding imputation-revised variable IRPRREL. Because the pair relationship varies according to the ages of the respondents, modeling and imputation were done independently within each of 11 age group pairs. Table 11.8 presents these 11 age group pairs, as well as the pair relationships prevalent within each age group pair.

Table 11.8 Age Group Pairs with Associated Possible Pair Relationships

Age Group Pair Number	Age Group Pair	Pair Relationships Appearing in Age Group Pair (in Order of Prevalence)[1]	
		≥ 10% Prevalence[2]	< 10% Prevalence
0	12–14/12–14	Sibling-sibling	Other relationship
1	12–14/15–17	Sibling-sibling	Other relationship
2	12–14/18–25	Sibling-sibling	Other relationship; parent-child; spouse-spouse**
3	15–17/15–17	Sibling-sibling	Other relationship; spouse-spouse*
4	15–17/18–25	Sibling-sibling	Other relationship; spouse-spouse; parent-child*
5	18–20/18–25	Other relationship; sibling-sibling; spouse-spouse	Parent-child**

Table 11.8 Age Group Pairs with Associated Possible Pair Relationships (continued)

Age Group Pair Number	Age Group Pair	Pair Relationships Appearing in Age Group Pair (in Order of Prevalence)[1] ≥ 10% Prevalence[2]	< 10% Prevalence
6	21–25/21–25	Spouse-spouse; other relationship; sibling-sibling	Parent-child**
7	12–14/26+	Parent-child	Other relationship; grandparent-grandchild; sibling-sibling*
8	15–17/26+	Parent-child	Other relationship; grandparent-grandchild; sibling-sibling; spouse-spouse**
9	18–20/26+	Parent-child	Other relationship; sibling-sibling; spouse-spouse; grandparent-grandchild
10	21+/26+	Spouse-spouse; parent-child; other relationship; sibling-sibling	Grandparent-grandchild*

* Pair relationships occur in less than 1 percent of the overall total number of pair relationships.
**The pair relationship is so rare that it does not appear in the age group pair in every survey year.
[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships. The spouse-spouse domain as listed here actually consists of two domains (spouse-spouse-with-children and spouse-spouse-without-children) that have been collapsed for the purposes of making the table easier to read. "Other relationship" refers to a relationship other than sibling-sibling, parent-child, grandparent-grandchild, or spouse-spouse.
[2] The pair relationships each form at least 10 percent of the overall total number of pair relationships within the given age group pair, and the total is at least 85 percent of the overall total.

11.5.2.1 Response Propensity Step

For a respondent pair to be considered complete, the pair relationship must be definitively established. In terms of the variable PAIRREL, this meant that the pair had to have a value of PAIRREL within the range of 1 to 9 or equal to 11 or 12. A value of PAIRREL equal to 13 also was considered complete, even though the pair relationship was not definitively established, because it was known that the pair relationship was not a relationship of interest. Response propensity adjustments then were calculated for each age group pair in order to make the respondent pair weights representative of the entire sample of pairs. Because the modeling of the final pair weight adjustments was not completed at the time of the pair imputations, the pair-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment. These adjustments were calculated using an item response propensity model, which is a special case of the generalized exponential model. See Appendix A of the questionnaire dwelling unit-level and person pair-level sampling weight calibration report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2016b) for technical details about this procedure.

11.5.2.2 Prediction Step

After the weights were adjusted using the item response propensity model within each age group pair, logistic regression models were fitted using the adjusted weights. Preferred covariates in these models included the age count variables described in Section 11.5.1.2. However, variables with missing values for some observations cannot be used as covariates in the models. To allow for better models whenever possible, the data were partitioned into two groups: those with nonmissing values for all age count variables, and those where one or more values of the age count variables were missing. When all of the age count variables were nonmissing, the predicted mean model was fitted using these variables. Otherwise, the model was fitted using the overall household count. This resulted in two predicted mean models for each of the 11 age group pairs.[160] These groups were combined again before the hot-deck imputation step. Further details about the variables used in these models can be found in Appendix D.

All modeling was done using the SUDAAN procedure MULTILOG; however, the number of levels in the response variable varied based on the age group pair.[161] For age group pairs 0 through 4 and 7 through 9, dichotomous logistic regression models were built. Because there were three outcomes with age group pairs 5, 6, and 10, polytomous logistic models were fitted for these age group pairs. All the models incorporated the weights and were calibrated to account for item nonresponse (where a pair responded to the survey but the pair relationship was unknown), using the item response propensity models, as described in Section 11.5.2.1.

Ideally, each type of pair relationship within an age group pair would constitute a response category in a multinomial response model. However, the number of cases corresponding to some pair relationships within each age group pair was very small. Hence, it was not feasible to fit multinomial models that cover all the possible pair relationships for a given age group pair. Therefore, in the modeling step, some of the response categories were combined with separate assignments of imputed values within each of the 11 age group pairs. Priority was given to placing the pair relationships "of interest" into separate categories. In some cases, pair relationships that were not of interest were combined with other categories, even if there were sufficient numbers to have a separate category in the multinomial model. Table 11.9 presents the response categories that were used for modeling. The delineation between categories that were combined for modeling was determined in the hot-deck step.

As an example, consider age group pair 5. In this age group pair, there are typically four types of pair relationships that have a sufficient number of respondent pairs to fit an adequate model, including both spouse-spouse domains, sibling-sibling pairs, and all others. Models with fewer response levels are generally easier to fit because there are more observations in each response level. Because only two of those four were pair relationships of interest, the two spouse-spouse domains were used as levels in the response variable. The third level was obtained by combining the sibling-sibling and other relationship pairs. There are typically a small number of parent-child pairs, which also were combined with the other relationship pairs.

[160] Partitioning the observations into two groups is similar to including an indicator variable in the PRD model. Though two models were fitted in practice, in theory this is similar to fitting only one PRD model. Thus, this is still considered the single RP/single PRD type of PMN.

[161] Though the SUDAAN procedure RLOGIST could have been used for all age group pairs except 5, 6, and 9, MULTILOG was used for all models for coding simplicity.

Table 11.9 Modeled Pair Relationships within Age Group Pairs

Age Group Pair Number	Age Group Pair	Number of Levels in Response	Levels of Modeled Response
0	12–14/12–14	2	Sibling-sibling; all others
1	12–14/15–17	2	Sibling-sibling; all others
2	12–14/18–25	2	Sibling-sibling; all others
3	15–17/15–17	2	Sibling-sibling; all others
4	15–17/18–25	2	Sibling-sibling; all others
5	18–20/18–25	3	Both spouse-spouse pair relationships;[1] all others
6	21–25/21–25	3	Both spouse-spouse pair relationships;[1] all others
7	12–14/26+	2	Parent-child; all others
8	15–17/26+	2	Parent-child; all others
9	18–20/26+	2	Parent-child; all others
10	21+/26+	3	Both spouse-spouse pair relationships;[1] all others

[1] The two spouse-spouse pair relationships are spouse-spouse and spouse-spouse-with-children-younger-than-18. The pair relationships labeled "spouse-spouse" include partner-partner pair relationships.

11.5.2.3 Hot-Deck Step

Likeness constraints used in imputation of the pair relationship were generally based on the number of household members in various age groups and on the marital status and genders of the respondents. Logical constraints were limited to the information that was already known about the pair, as denoted by the level of the variable PAIRREL. If, for example, PAIRREL = 14, then no information was available about the identity of the pair relationship and no logical constraint was needed. On the other hand, if PAIRREL = 15, this meant that the pair relationship was either a parent-child pair where the child was aged 12 to 14 or it was some relationship other than spouse-spouse, parent-child, grandparent-grandchild, or sibling-sibling.

11.5.3 Stage Two Imputation: Multiplicity Counts

In many cases where the pair relationships were not defined, multiplicity counts also were not defined. In addition, there were a handful of cases where multiplicity counts were not determined, even when the pair relationship was known. In all of these cases, imputation was required to determine the multiplicity count, and the single RP/single PRD type of PMN was used for imputation.

The multiplicity count variables were divided into six imputation sets: four sets for the multiplicities associated with the four sibling-sibling pair domains; one set for the parent-child, child focus domains; and one set for the parent-child, parent focus domains. The variables in each imputation set are provided in Table 11.10. Because the parent-child counts are hierarchical (i.e., the count for 12-17 must be less than or equal to the count for 12-20), only the counts for the 12-20 age group were modeled. Using the predicted means from these models, a single donor pair for each focus was selected from which the multiplicity counts were determined for the 12-14, 12-17, 15-17, and 12-20 parent-child pairs. No imputation was required for the spouse-

spouse multiplicity counts, because a selected respondent in a spouse-spouse pair was assumed to have had only one spouse.

Table 11.10 Modeled Multiplicities within Domains

Imputation Set	Domain	Model Type	Base Variables	Imputation-Revised Variables
1	Parent-Child, Child Focus	Logistic	MCPCC14, MCPCC57, MCPCC17, MCPCC20	IRMPCC14, IRMPCC57, IRMPCC17, IRMPCC20
2	Parent-Child, Parent Focus	Poisson	MCPCP14, MCPCP57, MCPCP17, MCPCP20	IRMPCP14, IRMPCP57, IRMPCP17, IRMPCP20
3	Sibling-Sibling (12-14/15-17) Older Focus	Poisson	MCS1417	IRMS1417
4	Sibling-Sibling (12-14/15-17) Younger Focus	Poisson	MCS1714	IRMS1714
5	Sibling-Sibling (12-17/18-25) Older Focus	Poisson	MCS1725	IRMS1725
6	Sibling-Sibling (12-17/18-25) Younger Focus	Poisson	MCS2517	IRMS2517

11.5.3.1 Imputation for Parent-Child Multiplicity, Child Focus (Imputation Set 1)

The first imputation set included the four parent-child, child focus multiplicity variables. Modeling was done only for the parent-child (12-20) variable, and multivariate assignment of all four variables was done in the hot-deck step.

11.5.3.1.1 Response Propensity Step

For a respondent pair to be considered complete with regard to the parent-child multiplicities, the multiplicity had to be nonmissing for the domains with children aged 12 to 20. A nonmissing multiplicity for this domain would automatically guarantee nonmissing multiplicities for the subset parent-child domains. Response propensity adjustments were then calculated in order to make the respondent pair weights representative of the entire sample of pairs.

11.5.3.1.2 Prediction Step

For the child-focus parent-child domains, the count being modeled was the number of parents of children aged 12 to 20 who were part of a parent-child pair. In most cases only two responses were possible: one parent or two parents. There were rare instances where the child could have three parents living in the household, with some combination of biological, step, foster, or adoptive parents. For the purposes of modeling, these cases were collapsed with the two-parent households. Similar to the procedure described in Section 11.5.2.2, the data were divided into two groups based on whether the age count variables were nonmissing, and separate models were fitted for each group. The fitted models were binomial logistic regression models using the SUDAAN procedure RLOGIST.

11.5.3.1.3 Hot-Deck Step

Though modeling was only done for the 12-20 age group, the multiplicity counts for the 12-14, 12-17, 15-17, and 12-20 parent-child pairs were assigned from a single donor pair. Likeness constraints were generally based on the pair relationship, the household size, and the number of household members in various age groups. Logical constraints were based on the bounds created during the editing process. These constraints are described in detail in Tables E.112 and E.113 in Appendix E.

11.5.3.2 Imputation for All Other Multiplicities (Imputation Sets 2 through 6)

The imputation process for the other responses (parent-focus parent-child and sibling-sibling multiplicity counts) was similar to that for the child-focus parent-child counts. The main difference is in the PRD step where Poisson regression models were used to model the counts. The counts of the number of children or siblings were underdispersed for a Poisson distribution so that the data had to be scaled using the observed variance.

As with the child-focus parent-child multiplicity counts, the parent-focus parent-child multiplicity counts were only modeled for the 12-20 age group, and counts from all four age groups were assigned from a single donor pair. For the sibling-sibling multiplicity counts, there was only one variable in each imputation set, and each count was modeled separately.

11.5.4 Stage Three Imputation: Household-Level Person Counts

Because of the difficulty in definitively determining household-level counts in many cases, imputation was not uncommon. Household-level person counts were divided into five imputation sets based on the domains listed in Section 11.4: one for each of the two sibling-sibling counts, one for the spouse-spouse counts, one for the spouse-spouse with children counts, and one for the parent-child counts. The first four imputation sets were handled using the single RP/single PRD type of PMN, whereas the imputation of the parent-child household counts was done using the multiple RP/multiple PRD type of PMN. For these counts, separate models were fit for the child-focus and parent-focus counts, and the predicted means from both models were brought together in one hot-deck step.

As with the multiplicities, the parent-child domains were hierarchical, so the imputations could not have been conducted independently if consistency was to be maintained. Hence, models were fitted only to the parent-child domains for the 12-20 age group, and the household-level person counts were assigned for the 12-14, 12-17, and 12-20 parent-child pair domains from a single donor. The household-level person counts for the 15-17 parent-child domains were not determined as they can be easily derived. The spouse-spouse household-level person counts were also hierarchical in that knowledge of whether a spouse-spouse pair was in the household was required before one could say that the pair had children. Therefore, imputations of the spouse-spouse counts were processed first, followed by the imputations of whether the spouse-spouse pairs in the household had children.

Household-level person counts were defined for all respondents, regardless of which pair they belonged to, or even whether they were within a pair at all. For modeling purposes, respondents were partitioned into two groups based on whether they belonged to a pair, and the entire imputation process was conducted separately for each group.

11.5.4.1 Imputation for Sibling-Sibling (12-14/15-17), Older Focus Household Counts (Imputation Set 1)

The first imputation set included one base variable, HCS1417, and the corresponding imputation-revised variable IRHS1417. The imputation process for this variable is detailed in Sections 11.5.4.1.1 through 11.5.4.1.3.

11.5.4.1.1 Response Propensity Step

For a pair or single respondent to be considered complete, the household-level person counts had to be nonmissing for all the variables being imputed. In addition to being separated into pair and nonpair households, respondents also were split by age for both the modeling of the response propensity and the predicted means. For the pairs, households where both pair members were younger than 18 were placed in one group, and the remaining pairs were in the other age group. For the single respondents, one age group consisted of respondents who were younger than 18, and the other consisted of respondents who were 18 or older. For pairs, response propensity adjustments were calculated in order to make the household weights representative of the entire sample of pairs. For single respondents, household weights also were used, and the adjustments were calculated in order to make the respondent household weights representative of the entire sample of households that were not part of a pair.

11.5.4.1.2 Prediction Step

After the response propensity adjustment was complete, the data were split into two groups based on whether the household-level age count variables were nonmissing, as previously described in Section 11.5.2.2. The outcome variable for the household-level sibling-sibling (12-14/15-17) count models was the number of household members aged 15 to 17 with a sibling aged 12 to 14 living with them. These counts could have a value of zero, which distinguished them from the multiplicities from a modeling point of view. Poisson regression was used to fit the models for the household-level person counts corresponding to the sibling-sibling domains. The data were underdispersed for a Poisson distribution so that the data had to be scaled using the observed variance. Modeling was done using the SUDAAN procedure LOGLINK.

11.5.4.1.3 Hot-Deck Step

After the modeling steps were complete, the two age groups were combined for one hot-deck imputation step. Imputation was conducted separately for pair and nonpair households. Likeness constraints used in imputation of household counts were generally based on the number of families in the household, the household size, and the number of household members in various age groups. Logical constraints were based on the bounds created during the editing process. These constraints are described in detail in Tables E.121 and E.122 in Appendix E.

In those instances where an imputed value could not be found after loosening all the likeness constraints, the imputed value was determined by doing a random imputation within bounds derived from the household composition.

11.5.4.2 Imputation for Sibling-Sibling (12-17/18-25) and Both Spouse-Spouse Household Counts (Imputation Sets 2 through 4)

The second, third, and fourth imputation sets were processed similarly to the first. One large difference was that when modeling the spouse-spouse-with-children counts, the data were not separated according to age groups. This applied to both the response propensity adjustments and the calculation of predicted means. Because of the hierarchical relationship between the spouse-spouse and the spouse-spouse-with-children counts, the response propensity adjustment for the spouse-spouse-with-children domain adjusted the weights to be representative of all spouse-spouse pairs rather than the entire sample. Missing counts for the spouse-spouse-with-children domain were not replaced via imputation until it was known definitively, after the hot-deck step of the PMN imputation, whether a household had spouse-spouse pairs.

Polytomous logistic regression was used to model the count of spouse-spouse pairs, with the possible responses being zero, one, and two or more spouse-spouse pairs in the household. Whether or not the spouse-spouse pairs had children younger than 18 was modeled with binomial logistic regression. These models were fitted using the SUDAAN MULTILOG procedure.

In some cases, two family units were in a household. If these resulted in unusual household-level person counts, they were excluded from the modeling step and were considered nonrespondents for the purposes of weight adjustment. No predicted mean was calculated in these cases, and instead of matching donors and recipients using predicted means, the imputed value was determined using random imputation within the preset bounds. One case where this may have occurred was with the spouse-spouse-with-children counts. Having two spouse-spouse pairs with children younger than 18 was an extremely rare category. Therefore, the two response categories that resulted for the spouse-spouse-with-children models were zero or one or more. Households with two family units did not need to be excluded from the spouse-spouse models, because having two spouse-spouse pairs in a household, though not common, was not rare.

11.5.4.3 Imputation for Parent-Child Household Counts (Imputation Set 5)

In contrast to the first four imputation sets, the parent-child household counts were imputed using the multiple RP/multiple PRD type of PMN. This imputation set included the three child-focus counts, HCPCC14, HCPCC17, and HCPCC20, as well as the three parent-focus counts, HCPCP14, HCPCP17, and HCPCP20. The corresponding imputation-revised variables are IRHPCC14, IRHPCC17, IRHPCC20, IRHPCP14, IRHPCP17, and IRHPCP20, respectively. The child-focus and parent-focus counts were modeled separately and joined together in one final hot-deck step. Just as was done for the first four imputation sets, the respondents were grouped into pair and nonpair households, and the entire imputation process was completed separately for each group. Unlike most imputation sets processed using the multiple RP/multiple PRD type of PMN, no provisional hot-deck steps were implemented. Because the child-focus counts were not used as covariates for the parent-focus counts and the parent-focus counts were not used as covariates for the child-focus counts, no provisional hot-deck steps were necessary, and the RP and PRD steps for the child-focus counts were run in parallel with the RP and PRD steps for the parent-focus counts.

11.5.4.3.1 First Response Propensity Step: Child Focus

All respondents were in the domain for all household counts. For a pair or single respondent to be considered complete, the three child-focus counts had to be nonmissing. Similar to the processing of the sibling-sibling counts and the spouse-spouse counts, the data were divided into two age groups based on whether the respondent or pair of respondents was 18 or older.

11.5.4.3.2 First Prediction Step: Child Focus

For the child-focus counts, the modeled response was the number of children aged 12 to 20 in the household with at least one parent living with them. This was modeled using Poisson regression, where the data were scaled using the observed variance to account for underdispersion. The LOGLINK procedure in SUDAAN was used to fit the model. The predicted mean used in the final hot-deck step was the predicted number of children aged 12 to 20 in the household with at least one parent living with them.

11.5.4.3.3 Second Response Propensity Step: Parent Focus

This RP step was identical to the RP step for the child-focus counts, except item respondents were those whose three parent-focus counts were nonmissing.

11.5.4.3.4 Second Prediction Step: Parent Focus

For the parent-focus counts, the modeled response was a three-level variable, based on the number of parents in the household with children aged 12 to 20: zero, one, or two or more. Polytomous logistic regression was used to fit the model as implemented by the MULTILOG procedure in SUDAAN. The predicted means used in the final hot-deck step were the predicted probabilities associated with each of the three levels of the response variable.

11.5.4.3.5 Hot-Deck Step

The predicted means from the models fit for the child-focus and parent-focus counts were brought together in one hot-deck imputation step. Because there are six variables in the imputation set, and any combination of them could be missing, there were 63 possible missingness patterns. These patterns are enumerated in Table E.127 in Appendix E. As with the other four imputation sets, logical constraints were based on the bounds created during the editing process, and likeness constraints were generally based on the number of families in the household, the household size, and the number of household members in various age groups. These constraints are described in detail in Tables E.121 and E.122.

In cases where there were two family units in the household, resulting in unusual counts, the counts were not included in the parent-focus models, and no predicted means were calculated. Even though two-family households were included in the model for the child-focus parent-child counts, the resulting predicted means were not used. This was because the parent-focus and child-focus parent-child counts were in the same imputation set, and the predicted means could not be used in the imputation of the parent-focus parent-child counts when two families were in the household. In these cases, imputation was random between the bounds.

This page intentionally left blank

References

Allison, P. D. (2008). Missing data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology*. Thousand Oaks, CA: Sage Publications, Inc.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (DSM-IV) (4th ed.). Washington, DC: Author.

Ault, K., Aldworth, J., Barnett-Walker, K., Carpenter, L., Copello, E., Frechtel, P., Liu, B., & Martin, P. (2009). Imputation report. In *2007 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.377.007). Research Triangle Park, NC: RTI International. Retrieved from http://media.samhsa.gov/data/

Ault, K., Barnett-Walker, K., Carpenter, L., Copello, E., Cummiskey, C., Frechtel, P., Laufenberg, J., Liu, B., Martin, P., & Moore, A. (2010). Imputation report. In *2008 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.477.007). Research Triangle Park, NC: RTI International. Retrieved from http://media.samhsa.gov/data/

Ault, K., Barnett-Walker, K., Carpenter, L., Cummiskey, C., Frechtel, P., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2011). Imputation report. In *2009 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.577.007). Research Triangle Park, NC: RTI International. Retrieved from http://media.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015a). *Results from the 2014 National Survey on Drug Use and Health: Detailed tables*. Retrieved from http://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015b). *2013 National Survey on Drug Use and Health: Methodological resource book (Section 10, Editing and imputation report)*. Retrieved from http://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015c). *2014 National Survey on Drug Use and Health: Editing and imputation evaluation report* (unpublished internal documentation). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2015d). *2014 National Survey on Drug Use and Health: Methodological resource book (Section 2, Sample design report)*. Retrieved from http://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2015e). *2014 National Survey on Drug Use and Health: Methodological summary and definitions*. Retrieved from http://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016a). *2014 National Survey on Drug Use and Health: Methodological resource book (Section 11, Person-level sampling weight calibration)*. Retrieved from http://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2016b). *2014 National Survey on Drug Use and Health: Methodological resource book (Section 12, Questionnaire dwelling unit-level and person pair-level sampling weight calibration)*. Retrieved from http://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (in press). *National Survey on Drug Use and Health: Evaluation of imputation methods for the National Survey on Drug Use and Health*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

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

Chromy, J. R., & Singh, A. C. (2001). Estimation for person-pair drug-related characteristics in the presence of pair multiplicities and extreme sampling weights. In *Proceedings of the 2001 Joint Statistical Meetings*, *American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Cox, B. G. (1980). The weighted sequential hot deck imputation procedure. In *Proceedings of the 1980 American Statistical Association, Survey Research Methods Section, Houston, TX* (pp. 721-726). Washington, DC: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Cox, B. G., & Cohen, S. B. (1985). *Methodological issues for health care surveys*. New York: Marcel Dekker, Inc.

Cox, D. R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). (Monographs on Statistics and Applied Probability Series). Boca Raton, FL: CRC Press.

Frechtel, P., Archambault, H., Carpenter, L., Cummiskey, C., Edwards, S., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2013). Imputation report. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.006.007). Research Triangle Park, NC: RTI International. Retrieved from http://www.samhsa.gov/data/

Frechtel, P. A., Ewing, G. S., Gulledge, K. E., Edwards, S. L., & Bose, J. (2014). An imputation model database and its relevance to analysis. In *JSM Proceedings,* pp. 3685–3689. Alexandria, VA: American Statistical Association.

Iannacchione, V. (1982). Weighted sequential hot deck imputation macros. In *Proceedings of the Seventh Annual SAS Users Group International Conference* (pp. 759-763). Cary, NC: SAS Corporation.

Kroutil, L. A., Chien, C., Handley, W., & Bradshaw, M. R. (2013). Procedures for editing interviewer-administered data in the 2011 NSDUH computer-assisted interview. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.003). Research Triangle Park, NC: RTI International. Retrieved from http://www.samhsa.gov/data/

Kroutil, L. A., Handley, W., & Bradshaw, M. R. (2013). General principles and procedures for editing drug use data in the 2011 NSDUH computer-assisted interview. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.003). Research Triangle Park, NC: RTI International. Retrieved from http://www.samhsa.gov/data/

Kroutil, L., & Myers, L. (2002). Development of editing rules for CAI substance use data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 85-109). Rockville, MD: Substance Abuse and Mental Health Services Administration.

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

Office of Applied Studies. (2008). *Results from the 2007 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 08-4343, NSDUH Series H-34). Retrieved from http://media.samhsa.gov/data/

Office of Applied Studies. (2009). *Results from the 2008 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 09-4434, NSDUH Series H-36). Retrieved from http://media.samhsa.gov/data/

Office of Applied Studies. (2010a). *Results from the 2009 National Survey on Drug Use and Health: Volume I. Summary of national findings* (HHS Publication No. SMA 10-4586Findings, NSDUH Series H-38A). Retrieved from http://media.samhsa.gov/data/

Office of Applied Studies. (2010b). *Results from the 2009 National Survey on Drug Use and Health: Volume II. Technical appendices and selected prevalence tables* (HHS Publication No. SMA 10-4586Appendices, NSDUH Series H-38B). Retrieved from http://media.samhsa.gov/data/

Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register, 62*(210), 58781-58790.

Office of Management and Budget. (2009). *Update of statistical area definitions and guidance on their uses* (OMB Bulletin No. 10-02). Washington, DC: The White House. Retrieved from https://www.whitehouse.gov/sites/default/files/omb/assets/bulletins/b10-02.pdf

RTI International. (2013). *SUDAAN® language manual, Release 11.0.1*. Research Triangle Park, NC: RTI International.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics, 4*(1), 87-94.

Ruppenkamp, J., Davis, T., Kroutil, L., & Aldworth, J. (2006). Methamphetamine analysis report. In *2005 National Survey on Drug Use and Health: Methodological resource book* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, RTI/0208726.187.038). Research Triangle Park, NC: RTI International. Retrieved from http://media.samhsa.gov/data/

Ruppenkamp, J., Frechtel, P., Aldworth, W. J., Carpenter, L., Clarke, A., Davis, T., Kroutil, L. A, & Martin, P. C. (2007). Methamphetamine analysis report. In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 15, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009). Research Triangle Park, NC: RTI International. Retrieved from http://media.samhsa.gov/data/

SAS Institute Inc. (2014). *SAS/STAT® 13.2 user's guide: High-performance procedures.* Cary, NC: SAS Institute Inc.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (No. 72, Monographs on Statistics and Applied Probability). Boca Raton, FL: Chapman and Hall/CRC.

Shiffman, S., Hickcox, M., Gnys, M., Paty, J. A., & Kassel, J. D. (1995, March). *The Nicotine Dependence Syndrome Scale: Development of a new measure.* Poster presented at the annual meeting of the Society for Research on Nicotine and Tobacco, San Diego, CA.

Shiffman, S., Waters, A. J., & Hickcox, M. (2003). The Nicotine Dependence Syndrome Scale: A multi-dimensional measure of nicotine dependence. Unpublished manuscript.

Singh, A., Grau, E., & Folsom, R. (2004). Imputation and unbiased estimation: Use of centered predictive mean neighborhoods method. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Ontario, Canada* (pp. 4351-4358). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Appendix A: Percentage of Cases Imputed or Logically Assigned, 2014 Survey

Table A.1 Percentage of Cases Imputed for General Cigarette Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Weighted Percentages: Imputed	Logically Assigned	Imputed or Logically Assigned
Cigarette Recency	Lifetime Cigarette Users	35,609	303	1	304	0.40%	0.00%	0.40%
Cigarette 30-Day Frequency	Past Month Cigarette Users	14,105	17	94	111	0.10%	0.72%	0.82%
Cigarette Age at First Use	Lifetime Cigarette Users	35,609	374	12	386	0.84%	0.06%	0.90%
Cigarette Day of First Use	Lifetime Cigarette Users	35,609	35,609	0	35,609	100.00%	0.00%	100.00%
Cigarette Month of First Use	Lifetime Cigarette Users	35,609	33,987	16	34,003	98.12%	0.01%	98.14%
Cigarette Year of First Use	Lifetime Cigarette Users	35,609	33,871	3	33,874	98.00%	0.00%	98.00%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.2 Percentage of Cases Imputed for Daily Cigarette Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Weighted Percentages: Imputed	Logically Assigned	Imputed or Logically Assigned
Cigarette Daily Use	Lifetime Cigarette Users	35,609	27	29	56	0.09%	0.11%	0.20%
Cigarette Age at First Daily Use	Daily Cigarette Users	18,405	151	0	151	0.97%	0.00%	0.97%
Cigarette Day of First Daily Use	Daily Cigarette Users	18,405	18,405	0	18,405	100.00%	0.00%	100.00%
Cigarette Month of First Daily Use	Daily Cigarette Users	18,405	17,825	0	17,825	98.71%	0.00%	98.71%
Cigarette Year of First Daily Use	Daily Cigarette Users	18,405	17,782	0	17,782	98.60%	0.00%	98.60%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.3 Percentage of Cases Imputed for Cigarette Dependence Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Smoke Cigarettes Regularly throughout the Day	Past Month Cigarette Users	14,105	1	0	1	0.00%	0.00%	0.00%
Smoke Same Number of Cigarettes from Day to Day	Past Month Cigarette Users	14,105	7	0	7	0.03%	0.00%	0.03%
Number Cigarettes Smoked per Day Often Changes	Past Month Cigarette Users	14,105	12	0	12	0.08%	0.00%	0.08%
Number Cigarettes Smoked per Day Influenced by Other Things	Past Month Cigarette Users	14,105	5	0	5	0.04%	0.00%	0.04%
Smoke Same Number of Cigarettes on Weekends As on Weekdays	Past Month Cigarette Users	14,105	11	0	11	0.05%	0.00%	0.05%
Smoke Lots of Cigarettes in an Hour, Then No Cigarettes Until Hours Later	Past Month Cigarette Users	14,105	20	0	20	0.10%	0.00%	0.10%
Amount of Smoking Has Increased Since Started Smoking	Past Month Cigarette Users	14,105	2	0	2	0.00%	0.00%	0.00%
Must Smoke Much More Now Before Start To Feel Anything	Past Month Cigarette Users	14,105	26	0	26	0.17%	0.00%	0.17%
Need To Smoke a Lot More To Be Satisfied	Past Month Cigarette Users	14,105	3	0	3	0.03%	0.00%	0.03%
Crave Cigarettes when Haven't Smoked for a Few Hours	Past Month Cigarette Users	14,105	2	0	2	0.00%	0.00%	0.00%
Need To Smoke To Feel Less Irritable	Past Month Cigarette Users	14,105	19	0	19	0.16%	0.00%	0.16%
Feel in Control of Smoking	Past Month Cigarette Users	14,105	7	0	7	0.05%	0.00%	0.05%
Cravings for Cigarettes like Force Can't Control	Past Month Cigarette Users	14,105	6	0	6	0.03%	0.00%	0.03%
Worry about Running Out of Cigarettes	Past Month Cigarette Users	14,105	1	0	1	0.00%	0.00%	0.00%

Table A.3 Percentage of Cases Imputed for Cigarette Dependence Variables (continued)

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Smoking Not Affected by Other Things	Past Month Cigarette Users	14,105	11	0	11	0.10%	0.00%	0.10%
Tend To Avoid Places That Don't Allow Smoking	Past Month Cigarette Users	14,105	20	0	20	0.12%	0.00%	0.12%
No Travel by Airplane Because No Smoking Allowed	Past Month Cigarette Users	14,105	13	0	13	0.19%	0.00%	0.19%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.4 Percentage of Cases Imputed for Cigar Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Cigar Lifetime Use	All Respondents	67,901	11	0	11	0.01%	0.00%	0.01%
Cigar Recency	Lifetime Cigar Users	20,801	267	1	268	0.64%	0.00%	0.64%
Cigar 30-Day Frequency	Past Month Cigar Users	3,522	10	26	36	0.17%	0.64%	0.81%
Cigar Age at First Use	Lifetime Cigar Users	20,801	443	0	443	2.15%	0.00%	2.15%
Cigar Day of First Use	Lifetime Cigar Users	20,801	20,801	0	20,801	100.00%	0.00%	100.00%
Cigar Month of First Use	Lifetime Cigar Users	20,801	19,075	4	19,079	96.06%	0.02%	96.08%
Cigar Year of First Use	Lifetime Cigar Users	20,801	18,959	0	18,959	95.80%	0.00%	95.80%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.5 Percentage of Cases Imputed for Chewing Tobacco Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Chewing Tobacco Lifetime Use	All Respondents	67,901	15	0	15	0.02%	0.00%	0.02%
Chewing Tobacco Recency	Lifetime Chewing Tobacco Users	7,807	118	0	118	0.63%	0.00%	0.63%
Chewing Tobacco 30-Day Frequency	Past Month Chewing Tobacco Users	1,017	8	11	19	0.37%	1.05%	1.42%
Chewing Tobacco Age at First Use	Lifetime Chewing Tobacco Users	7,807	132	0	132	1.27%	0.00%	1.27%
Chewing Tobacco Day of First Use	Lifetime Chewing Tobacco Users	7,807	7,807	0	7,807	100.00%	0.00%	100.00%
Chewing Tobacco Month of First Use	Lifetime Chewing Tobacco Users	7,807	7,360	1	7,361	97.52%	0.00%	97.52%
Chewing Tobacco Year of First Use	Lifetime Chewing Tobacco Users	7,807	7,327	0	7,327	97.35%	0.00%	97.35%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.6 Percentage of Cases Imputed for Snuff Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Snuff Lifetime Use	All Respondents	67,901	45	0	45	0.08%	0.00%	0.08%
Snuff Recency	Lifetime Snuff Users	9,480	168	0	168	0.97%	0.00%	0.97%
Snuff 30-Day Frequency	Past Month Snuff Users	2,294	9	16	25	0.32%	0.50%	0.81%
Snuff Age at First Use	Lifetime Snuff Users	9,480	175	0	175	1.60%	0.00%	1.60%
Snuff Day of First Use	Lifetime Snuff Users	9,480	9,480	0	9,480	100.00%	0.00%	100.00%
Snuff Month of First Use	Lifetime Snuff Users	9,480	8,756	3	8,759	96.38%	0.02%	96.40%
Snuff Year of First Use	Lifetime Snuff Users	9,480	8,702	0	8,702	96.12%	0.00%	96.12%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.7 Percentage of Cases Imputed for Smokeless Tobacco Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Smokeless Tobacco Lifetime Use	All Respondents	67,901	50	0	50	0.09%	0.00%	0.09%
Smokeless Tobacco Recency	Lifetime Smokeless Tobacco Users	11,647	217	0	217	0.96%	0.00%	0.96%
Smokeless Tobacco Age at First Use	Lifetime Smokeless Tobacco Users	11,647	266	0	266	1.79%	0.00%	1.79%
Smokeless Tobacco Day of First Use	Lifetime Smokeless Tobacco Users	11,647	11,647	0	11,647	100.00%	0.00%	100.00%
Smokeless Tobacco Month of First Use	Lifetime Smokeless Tobacco Users	11,647	10,838	3	10,841	96.83%	0.01%	96.84%
Smokeless Tobacco Year of First Use	Lifetime Smokeless Tobacco Users	11,647	10,778	0	10,778	96.61%	0.00%	96.61%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.8 Percentage of Cases Imputed for Pipe Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Pipe Lifetime Use	All Respondents	67,901	20	0	20	0.02%	0.00%	0.02%
Pipe Past Month Use	Lifetime Pipe Users	6,339	5	0	5	0.06%	0.00%	0.06%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.9 Percentage of Cases Imputed for Alcohol Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Alcohol Lifetime Use	All Respondents	67,901	15	0	15	0.03%	0.00%	0.03%
Alcohol Recency	Lifetime Alcohol Users	49,745	673	7	680	0.96%	0.00%	0.96%
Alcohol 12-Month Frequency	Past Year Alcohol Users	42,002	727	1,190	1,917	1.36%	1.89%	3.25%
Alcohol 30-Day Frequency	Past Month Alcohol Users	32,108	605	210	815	1.65%	0.66%	2.31%
Alcohol 5+ Drinks Past Month	Past Month Alcohol Users	32,108	952	65	1,017	2.66%	0.17%	2.82%
Alcohol Age at First Use	Lifetime Alcohol Users	49,745	649	0	649	1.47%	0.00%	1.47%
Alcohol Day of First Use	Lifetime Alcohol Users	49,745	49,745	0	49,745	100.00%	0.00%	100.00%
Alcohol Month of First Use	Lifetime Alcohol Users	49,745	46,029	4	46,033	96.90%	0.00%	96.91%
Alcohol Year of First Use	Lifetime Alcohol Users	49,745	45,811	0	45,811	96.73%	0.00%	96.73%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.10 Percentage of Cases Imputed for Marijuana Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Marijuana Lifetime Use	All Respondents	67,901	38	0	38	0.08%	0.00%	0.08%
Marijuana Recency	Lifetime Marijuana Users	28,636	409	3	412	0.91%	0.00%	0.91%
Marijuana 12-Month Frequency	Past Year Marijuana Users	11,786	381	781	1,162	2.60%	4.92%	7.52%
Marijuana 30-Day Frequency	Past Month Marijuana Users	7,243	182	70	252	2.04%	0.74%	2.78%
Marijuana Age at First Use	Lifetime Marijuana Users	28,636	242	0	242	0.74%	0.00%	0.74%
Marijuana Day of First Use	Lifetime Marijuana Users	28,636	28,636	0	28,636	100.00%	0.00%	100.00%
Marijuana Month of First Use	Lifetime Marijuana Users	28,636	26,717	6	26,723	96.99%	0.01%	97.00%
Marijuana Year of First Use	Lifetime Marijuana Users	28,636	26,586	0	26,586	96.79%	0.00%	96.79%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.11 Percentage of Cases Imputed for Inhalant Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Inhalant Lifetime Use	All Respondents	67,901	119	0	119	0.13%	0.00%	0.13%
Inhalant Recency	Lifetime Inhalant Users	5,365	168	1	169	1.82%	0.00%	1.83%
Inhalant 12-Month Frequency	Past Year Inhalant Users	667	79	72	151	7.11%	7.20%	14.31%
Inhalant 30-Day Frequency	Past Month Inhalant Users	208	30	11	41	9.95%	3.13%	13.08%
Inhalant Age at First Use	Lifetime Inhalant Users	5,365	250	0	250	3.59%	0.00%	3.59%
Inhalant Day of First Use	Lifetime Inhalant Users	5,365	5,365	0	5,365	100.00%	0.00%	100.00%
Inhalant Month of First Use	Lifetime Inhalant Users	5,365	5,006	6	5,012	96.89%	0.03%	96.92%
Inhalant Year of First Use	Lifetime Inhalant Users	5,365	4,972	0	4,972	96.66%	0.00%	96.66%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.12 Percentage of Cases Imputed for Heroin Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Heroin Lifetime Use	All Respondents	67,901	43	0	43	0.04%	0.00%	0.04%
Heroin Recency	Lifetime Heroin Users	1,124	37	0	37	2.29%	0.00%	2.29%
Heroin 12-Month Frequency	Past Year Heroin Users	274	28	28	56	8.14%	10.76%	18.90%
Heroin 30-Day Frequency	Past Month Heroin Users	120	12	2	14	7.54%	1.06%	8.60%
Heroin Age at First Use	Lifetime Heroin Users	1,124	13	0	13	1.54%	0.00%	1.54%
Heroin Day of First Use	Lifetime Heroin Users	1,124	1,124	0	1,124	100.00%	0.00%	100.00%
Heroin Month of First Use	Lifetime Heroin Users	1,124	1,025	1	1,026	93.75%	0.06%	93.81%
Heroin Year of First Use	Lifetime Heroin Users	1,124	1,019	0	1,019	93.40%	0.00%	93.40%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.13 Percentage of Cases Imputed for Hallucinogen Use Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Hallucinogen Lifetime Use	All Respondents	67,901	205	0	205	0.20%	0.00%	0.20%
Hallucinogen Recency	Lifetime Hallucinogen Users	9,361	201	34	235	1.48%	0.25%	1.74%
Hallucinogen 12-Month Frequency	Past Year Hallucinogen Users	1,669	62	75	137	2.72%	3.42%	6.14%
Hallucinogen 30-Day Frequency	Past Month Hallucinogen Users	428	20	5	25	3.71%	0.83%	4.54%
Hallucinogen Age at First Use	Lifetime Hallucinogen Users	9,361	116	108	224	1.27%	1.24%	2.52%
Hallucinogen Day of First Use	Lifetime Hallucinogen Users	9,361	9,361	0	9,361	100.00%	0.00%	100.00%
Hallucinogen Month of First Use	Lifetime Hallucinogen Users	9,361	8,667	47	8,714	96.64%	0.23%	96.87%
Hallucinogen Year of First Use	Lifetime Hallucinogen Users	9,361	8,636	13	8,649	96.45%	0.05%	96.51%
Hallucinogen Use Other than LSD, PCP, or Ecstasy	All Respondents	67,901	184	0	184	0.18%	0.00%	0.18%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.14 Percentage of Cases Imputed for LSD Use Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
LSD Lifetime Use	All Respondents	67,901	58	0	58	0.07%	0.00%	0.07%
LSD Recency	Lifetime LSD Users	5,286	86	0	86	1.13%	0.00%	1.13%
LSD Age at First Use	Lifetime LSD Users	5,286	96	7	103	1.75%	0.07%	1.82%
LSD Day of First Use	Lifetime LSD Users	5,286	5,286	0	5,286	100.00%	0.00%	100.00%
LSD Month of First Use	Lifetime LSD Users	5,286	4,884	0	4,884	96.78%	0.00%	96.78%
LSD Year of First Use	Lifetime LSD Users	5,286	4,864	0	4,864	96.64%	0.00%	96.64%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.15 Percentage of Cases Imputed for PCP Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
PCP Lifetime Use	All Respondents	67,901	54	0	54	0.06%	0.00%	0.06%
PCP Recency	Lifetime PCP Users	1,093	23	0	23	1.60%	0.00%	1.60%
PCP Age at First Use	Lifetime PCP Users	1,093	46	2	48	4.91%	0.27%	5.18%
PCP Day of First Use	Lifetime PCP Users	1,093	1,093	0	1,093	100.00%	0.00%	100.00%
PCP Month of First Use	Lifetime PCP Users	1,093	1,060	0	1,060	99.20%	0.00%	99.20%
PCP Year of First Use	Lifetime PCP Users	1,093	1,058	0	1,058	99.16%	0.00%	99.16%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.16 Percentage of Cases Imputed for Ecstasy Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Ecstasy Lifetime Use	All Respondents	67,901	74	0	74	0.08%	0.00%	0.08%
Ecstasy Recency	Lifetime Ecstasy Users	4,984	64	1	65	0.91%	0.00%	0.91%
Ecstasy Age at First Use	Lifetime Ecstasy Users	4,984	45	1	46	0.98%	0.00%	0.98%
Ecstasy Day of First Use	Lifetime Ecstasy Users	4,984	4,984	0	4,984	100.00%	0.00%	100.00%
Ecstasy Month of First Use	Lifetime Ecstasy Users	4,984	4,543	0	4,543	94.55%	0.00%	94.55%
Ecstasy Year of First Use	Lifetime Ecstasy Users	4,984	4,519	0	4,519	94.27%	0.00%	94.27%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.17 Percentage of Cases Imputed for Cocaine Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Cocaine Lifetime Use	All Respondents	67,901	25	0	25	0.05%	0.00%	0.05%
Cocaine Recency	Lifetime Cocaine Users	8,022	87	33	120	0.76%	0.34%	1.10%
Cocaine 12-Month Frequency	Past Year Cocaine Users	1,377	57	108	165	4.26%	6.32%	10.58%
Cocaine 30-Day Frequency	Past Month Cocaine Users	451	30	18	48	7.16%	4.13%	11.29%
Cocaine Age at First Use	Lifetime Cocaine Users	8,022	58	103	161	0.79%	0.83%	1.62%
Cocaine Day of First Use	Lifetime Cocaine Users	8,022	8,022	0	8,022	100.00%	0.00%	100.00%
Cocaine Month of First Use	Lifetime Cocaine Users	8,022	7,558	0	7,558	97.34%	0.00%	97.34%
Cocaine Year of First Use	Lifetime Cocaine Users	8,022	7,538	0	7,538	97.21%	0.00%	97.21%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.18 Percentage of Cases Imputed for Crack Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Crack Lifetime Use	All Respondents	67,901	26	0	26	0.06%	0.00%	0.06%
Crack Recency	Lifetime Crack Users	1,951	19	0	19	0.98%	0.00%	0.98%
Crack 12-Month Frequency	Past Year Crack Users	184	8	12	20	3.10%	3.60%	6.70%
Crack 30-Day Frequency	Past Month Crack Users	79	1	1	2	0.25%	1.02%	1.26%
Crack Age at First Use	Lifetime Crack Users	1,951	17	0	17	1.05%	0.00%	1.05%
Crack Day of First Use	Lifetime Crack Users	1,951	1,951	0	1,951	100.00%	0.00%	100.00%
Crack Month of First Use	Lifetime Crack Users	1,951	1,881	0	1,881	98.27%	0.00%	98.27%
Crack Year of First Use	Lifetime Crack Users	1,951	1,877	0	1,877	98.19%	0.00%	98.19%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.19 Percentage of Cases Imputed for Tranquilizer Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Tranquilizer Lifetime Use	All Respondents	67,901	146	0	146	0.15%	0.00%	0.15%
Tranquilizer Recency	Lifetime Tranquilizer Users	6,089	122	0	122	1.86%	0.00%	1.86%
Tranquilizer 12-Month Frequency	Past Year Tranquilizer Users	1,633	76	103	179	3.50%	4.92%	8.42%
Tranquilizer Age at First Use	Lifetime Tranquilizer Users	6,089	193	0	193	3.69%	0.00%	3.69%
Tranquilizer Day of First Use	Lifetime Tranquilizer Users	6,089	6,089	0	6,089	100.00%	0.00%	100.00%
Tranquilizer Month of First Use	Lifetime Tranquilizer Users	6,089	5,466	0	5,466	94.02%	0.00%	94.02%
Tranquilizer Year of First Use	Lifetime Tranquilizer Users	6,089	5,431	0	5,431	93.65%	0.00%	93.65%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.20 Percentage of Cases Imputed for Sedative Use Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Sedative Lifetime Use	All Respondents	67,901	163	0	163	0.18%	0.00%	0.18%
Sedative Recency	Lifetime Sedative Users	1,267	38	0	38	1.72%	0.00%	1.72%
Sedative 12-Month Frequency	Past Year Sedative Users	230	22	22	44	6.56%	8.22%	14.78%
Sedative Age at First Use	Lifetime Sedative Users	1,267	71	0	71	4.58%	0.00%	4.58%
Sedative Day of First Use	Lifetime Sedative Users	1,267	1,267	0	1,267	100.00%	0.00%	100.00%
Sedative Month of First Use	Lifetime Sedative Users	1,267	1,181	0	1,181	97.48%	0.00%	97.48%
Sedative Year of First Use	Lifetime Sedative Users	1,267	1,174	0	1,174	97.31%	0.00%	97.31%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.21 Percentage of Cases Imputed for Pain Reliever Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Pain Reliever Lifetime Use	All Respondents	67,901	217	0	217	0.23%	0.00%	0.23%
Pain Reliever Recency	Lifetime Pain Reliever Users	9,817	343	16	359	3.12%	0.08%	3.20%
Pain Reliever 12-Month Frequency	Past Year Pain Reliever Users	3,303	258	181	439	8.23%	4.00%	12.23%
Pain Reliever Age at First Use	Lifetime Pain Reliever Users	9,817	579	14	593	6.64%	0.10%	6.74%
Pain Reliever Day of First Use	Lifetime Pain Reliever Users	9,817	9,817	0	9,817	100.00%	0.00%	100.00%
Pain Reliever Month of First Use	Lifetime Pain Reliever Users	9,817	8,962	18	8,980	95.03%	0.10%	95.13%
Pain Reliever Year of First Use	Lifetime Pain Reliever Users	9,817	8,887	3	8,890	94.54%	0.01%	94.55%
Pain Reliever Use Other than OxyContin	All Respondents	67,901	242	0	242	0.25%	0.00%	0.25%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.22 Percentage of Cases Imputed for OxyContin Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
OxyContin Lifetime Use	All Respondents	67,901	166	0	166	0.17%	0.00%	0.17%
OxyContin Recency	Lifetime OxyContin Users	2,208	86	1	87	3.57%	0.02%	3.58%
OxyContin 12-Month Frequency	Past Year OxyContin Users	467	42	36	78	7.73%	5.68%	13.41%
OxyContin Age at First Use	Lifetime OxyContin Users	2,208	71	2	73	4.07%	0.03%	4.11%
OxyContin Day of First Use	Lifetime OxyContin Users	2,208	2,208	0	2,208	100.00%	0.00%	100.00%
OxyContin Month of First Use	Lifetime OxyContin Users	2,208	2,011	3	2,014	94.79%	0.05%	94.84%
OxyContin Year of First Use	Lifetime OxyContin Users	2,208	1,992	1	1,993	93.82%	0.01%	93.83%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.23 Percentage of Cases Imputed for Stimulant Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Core Plus Noncore Stimulant Lifetime Use	All Respondents	67,901	145	0	145	0.16%	0.00%	0.16%
Core Plus Noncore Stimulant Recency	Lifetime Core Plus Noncore Stimulant Users	5,308	123	17	140	1.96%	0.25%	2.21%
Stimulant Lifetime Use	All Respondents	67,901	145	0	145	0.16%	0.00%	0.16%
Stimulant Recency	Lifetime Stimulant Users	4,810	103	17	120	1.59%	0.27%	1.87%
Stimulant 12-Month Frequency	Past Year Stimulant Users	1,184	63	110	173	5.08%	9.64%	14.72%
Stimulant Age at First Use	Lifetime Stimulant Users	4,810	119	34	153	2.38%	0.80%	3.17%
Stimulant Day of First Use	Lifetime Stimulant Users	4,810	4,810	0	4,810	100.00%	0.00%	100.00%
Stimulant Month of First Use	Lifetime Stimulant Users	4,810	4,420	3	4,423	95.59%	0.06%	95.65%
Stimulant Year of First Use	Lifetime Stimulant Users	4,810	4,406	1	4,407	95.46%	0.00%	95.46%
Stimulant Use Other than Methamphetamine	All Respondents	67,901	139	0	139	0.17%	0.00%	0.17%

Note: The estimates for stimulant lifetime use and stimulant recency include data from the methamphetamine items added in 2005 and 2006, but other estimates in this table do not.
Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.24 Percentage of Cases Imputed for Methamphetamine Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Core Plus Noncore Methamphetamine Lifetime Use	All Respondents	67,901	66	0	66	0.07%	0.00%	0.07%
Core Plus Noncore Methamphetamine Recency	Lifetime Core Plus Noncore Methamphetamine Users	2,824	65	0	65	1.94%	0.00%	1.94%
Methamphetamine Lifetime Use	All Respondents	67,901	66	0	66	0.07%	0.00%	0.07%
Methamphetamine Recency	Lifetime Methamphetamine Users	2,142	45	0	45	1.59%	0.00%	1.59%
Methamphetamine 12-Month Frequency	Past Year Methamphetamine Users	309	24	13	37	7.87%	2.87%	10.73%
Methamphetamine Age at First Use	Lifetime Methamphetamine Users	2,142	59	0	59	2.85%	0.00%	2.85%
Methamphetamine Day of First Use	Lifetime Methamphetamine Users	2,142	2,142	0	2,142	100.00%	0.00%	100.00%
Methamphetamine Month of First Use	Lifetime Methamphetamine Users	2,142	2,055	1	2,056	97.94%	0.02%	97.96%
Methamphetamine Year of First Use	Lifetime Methamphetamine Users	2,142	2,051	1	2,052	97.86%	0.00%	97.87%

Note: The estimates for methamphetamine lifetime use and methamphetamine recency include data from the methamphetamine items added in 2005 and 2006, but other estimates in this table do not.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.25 Percentage of Cases Imputed for Demographic Variables

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
General								
Education Level	All Respondents	67,901	9	0	9	0.02%	0.00%	0.02%
Marital Status	Respondents at Least 15 Years of Age	59,506	13	0	13	0.03%	0.00%	0.03%
Employment Status	Respondents at Least 15 Years of Age	59,506	57	0	57	0.12%	0.00%	0.12%
Employment Status: 18+	Respondents at Least 18 Years of Age	50,894	50	0	50	0.12%	0.00%	0.12%
Immigrant Status								
Born in the United States	All Respondents	67,901	37	6	43	0.07%	0.01%	0.08%
Immigrant Age of Entry in the United States	Respondents Not Born in the United States	8,307	23	0	23	0.41%	0.00%	0.41%
Hispanic or Latino Origin								
Hispanic or Latino Origin	All Respondents	67,901	165	1	166	0.12%	0.00%	0.12%
Hispanic or Latino Origin Group	Hispanic or Latino Respondents	11,743	59	86	145	0.42%	0.75%	1.18%
Single/Multiple Hispanic or Latino Origin Group	Hispanic or Latino Respondents	11,743	65	88	153	0.45%	0.79%	1.24%
Race								
Native Hawaiian	All Respondents	67,901	2,397	0	2,397	3.51%	0.00%	3.51%
Guamanian/Chamorro	All Respondents	67,901	2,397	0	2,397	3.51%	0.00%	3.51%
Samoan	All Respondents	67,901	2,397	0	2,397	3.51%	0.00%	3.51%
Other Pacific Islander	All Respondents	67,901	2,397	1	2,398	3.51%	0.00%	3.51%
Asian	All Respondents	67,901	2,404	357	2,761	3.52%	0.58%	4.10%
Black or African American	All Respondents	67,901	2,432	41	2,473	3.57%	0.05%	3.62%
American Indian/Alaska Native	All Respondents	67,901	2,510	38	2,548	3.60%	0.04%	3.64%
White	All Respondents	67,901	2,520	159	2,679	3.64%	0.23%	3.87%
Detailed Race: 15 Levels	All Respondents	67,901	2,561	565	3,126	3.70%	0.85%	4.54%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.26 Percentage of Cases Imputed for Roster Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Household Roster								
Number of People in Household	All Respondents	67,901	71	0	71	0.13%	0.00%	0.13%
Number of Children in Household Aged < 18	All Respondents	67,901	266	0	266	0.37%	0.00%	0.37%
Number of People in Household Aged ≥ 65	All Respondents	67,901	351	66	417	0.40%	0.10%	0.50%
Family Roster								
Presence of Family Members in Household	All Respondents	67,901	92	0	92	0.16%	0.00%	0.16%
Number of Respondent's Family Members in Household Excluding Foster Relationships	All Respondents	67,901	101	15	116	0.16%	0.03%	0.19%
Number of Respondent's Family Members in Household Including Foster Relationships	All Respondents	67,901	107	9	116	0.16%	0.03%	0.19%
Number of Respondent's Family Members in Household Aged < 18 Excluding Foster Relationships	All Respondents	67,901	195	0	195	0.28%	0.00%	0.28%
Number of Respondent's Family Members in Household Aged < 18 Including Foster Relationships	All Respondents	67,901	207	0	207	0.29%	0.00%	0.29%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.27 Percentage of Cases Imputed for Income Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Amount of Income								
Respondent's Total Income > or < $20,000	All Respondents	67,901	1,056	33	1,089	2.21%	0.04%	2.25%
Respondent's Total Income (Finer Categories)	All Respondents	67,901	2,245	0	2,245	4.77%	0.00%	4.77%
Total Family Income > or < $20,000	All Respondents	67,901	2,867	0	2,867	4.40%	0.00%	4.40%
Total Family Income (Finer Categories)	All Respondents	67,901	6,589	2,343	8,932	10.46%	6.36%	16.82%
Source of Income								
Family Received Income from Job	All Respondents	67,901	250	0	250	0.37%	0.00%	0.37%
Family Received Social Security or Railroad Retirement Payments	All Respondents	67,901	651	0	651	0.83%	0.00%	0.83%
Family Received Public Assistance	All Respondents	67,901	453	0	453	0.55%	0.00%	0.55%
Family Received Supplemental Security Income	All Respondents	67,901	862	0	862	1.07%	0.00%	1.07%
Respondent/Other Family Member Received Food Stamps	All Respondents	67,901	311	0	311	0.41%	0.00%	0.41%
Family Received Welfare/Job Placement/Child Care	All Respondents	67,901	334	0	334	0.41%	0.00%	0.41%
Number of Months on Welfare	Family Receives Public Assistance or Welfare/Job Placement/Child Care	3,916	174	0	174	4.39%	0.00%	4.39%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.28 Percentage of Cases Imputed for Health Insurance Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Respondent Has Health Insurance								
As Defined by the Constituent Variables Method	All Respondents	67,901	546	0	546	0.59%	0.00%	0.59%
Type of Insurance								
Private – As Defined by the Constituent Variables Method	All Respondents	67,901	465	0	465	0.55%	0.00%	0.55%
Medicare	All Respondents	67,901	275	0	275	0.33%	0.00%	0.33%
Military Health Care (CHAMPUS, TRICARE, CHAMPVA, VA)	All Respondents	67,901	262	0	262	0.31%	0.00%	0.31%
Medicaid/CHIP	All Respondents	67,901	572	0	572	0.71%	0.00%	0.71%
Other	Respondent Does Not Have Private Health Insurance, Medicare, Medicaid/CHIP, or CHAMPUS	10,183	200	0	200	1.44%	0.00%	1.44%

CHIP = Children's Health Insurance Program; CHAMPUS = Civilian Health and Medical Program of the Uniformed Services; CHAMPVA = Civilian Health and Medical Program of the Department of Veteran's Affairs; VA = Department of Veteran's Affairs.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Table A.29 Percentage of Cases Imputed for Pair Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Pair Relationships								
Family Pair Relationship Indicator	All Pair Members	36,458	420	0	420	1.24%	0.00%	1.24%
Multiplicities								
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-14	Parent-Child Pair, Child is 12-14	5,626	8	0	8	0.13%	0.00%	0.13%
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-17	Parent-Child Pair, Child is 12-17	10,356	10	0	10	0.10%	0.00%	0.10%
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-20	Parent-Child Pair, Child is 12-20	12,080	22	0	22	0.25%	0.00%	0.25%
Multiplicity Count: Child-Parent, Parent Focus, Child is 15-17	Parent-Child Pair, Child is 15-17	4,730	6	0	6	0.16%	0.00%	0.16%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-14	Parent-Child Pair, Child is 12-14	5,626	108	0	108	1.85%	0.00%	1.85%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-17	Parent-Child Pair, Child is 12-17	10,356	196	0	196	1.68%	0.00%	1.68%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-20	Parent-Child Pair, Child is 12-20	12,080	246	0	246	2.01%	0.00%	2.01%
Multiplicity Count: Child-Parent, Child Focus, Child is 15-17	Parent-Child Pair, Child is 15-17	4,730	88	0	88	1.51%	0.00%	1.51%
Multiplicity Count: Sibling-Sibling (12-14/15-17), 12-14 Focus	Sibling-Sibling Pair: Younger is 12-14, Older is 15-17	2,836	16	0	16	0.67%	0.00%	0.67%
Multiplicity Count: Sibling-Sibling (12-14/15-17), 15-17 Focus	Sibling-Sibling Pair: Younger is 12-14, Older is 15-17	2,836	16	0	16	0.65%	0.00%	0.65%

Table A.29 Percentage of Cases Imputed for Pair Variables (continued)

	Domain	Unweighted Frequencies: Respondents in Domain	Unweighted Frequencies: Imputed	Unweighted Frequencies: Logically Assigned	Unweighted Frequencies: Imputed or Logically Assigned	Weighted Percentages: Imputed	Weighted Percentages: Logically Assigned	Weighted Percentages: Imputed or Logically Assigned
Multiplicity Count: Sibling-Sibling (12-17/18-25), 12-17 Focus	Sibling-Sibling Pair: Younger is 12-17, Older is 18-25	3,342	32	0	32	0.81%	0.00%	0.81%
Multiplicity Count: Sibling-Sibling (12-17/18-25), 18-25 Focus	Sibling-Sibling Pair: Younger is 12-17, Older is 18-25	3,342	34	0	34	0.81%	0.00%	0.81%
Household Person-Level Count								
Household Count: Number of Spouse-Spouse Pairs in Household	All Respondents	67,901	143	0	143	0.25%	0.00%	0.25%
Household Count: Number of Spouse-Spouse Pairs with Children	All Respondents	67,901	75	0	75	0.14%	0.00%	0.14%
Household Count: Child-Parent, Parent Focus, Child is 12-14	All Respondents	67,901	301	1	302	0.35%	0.00%	0.35%
Household Count: Child-Parent, Parent Focus, Child is 12-17	All Respondents	67,901	468	1	469	0.50%	0.00%	0.50%
Household Count: Child-Parent, Parent Focus, Child is 12-20	All Respondents	67,901	555	0	555	0.66%	0.00%	0.66%
Household Count: Child-Parent, Child Focus, Child is 12-14	All Respondents	67,901	94	0	94	0.12%	0.00%	0.12%
Household Count: Child-Parent, Child Focus, Child is 12-17	All Respondents	67,901	151	0	151	0.17%	0.00%	0.17%
Household Count: Child-Parent, Child Focus, Child is 12-20	All Respondents	67,901	212	0	212	0.27%	0.00%	0.27%
Household Count: Sibling-Sibling (12-14/15-17), 15-17 Focus	All Respondents	67,901	87	0	87	0.13%	0.00%	0.13%
Household Count: Sibling-Sibling (12-17/18-25), 18-25 Focus	All Respondents	67,901	115	0	115	0.17%	0.00%	0.17%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Appendix B: Detailed Documentation of Edits for Variables from the Noncore Demographics, Core Substance Use, and Noncore Self-Administered Sections in the 2014 NSDUH

This page intentionally left blank

Appendix B: Detailed Documentation of Edits for Variables from the Noncore Demographics, Core Substance Use, and Noncore Self-Administered Sections in the 2014 NSDUH

B.1 Introduction

This appendix provides detailed documentation of edits that were implemented (where applicable) for variables from the noncore demographics, core substance use, and noncore self-administered sections. The remaining sections of the appendix are organized as follows:

- Section B.2 presents tables describing edits for selected noncore demographics variables corresponding to editing procedures that were discussed in Chapter 5.
- Section B.3 presents tables describing edits for core substance use variables corresponding to editing procedures that were discussed in Chapter 6.
- Section B.4 presents tables describing edits for variables in noncore self-administered modules corresponding to editing procedures that were discussed in Chapter 7.

The issues that are shown in the tables in these sections reflect cumulative response patterns that have been observed in the data since a given module or set of questions was added to the interview. However, not all of these patterns may be encountered in a given survey year. In addition, no editing was done for some types of responses that are presented in the tables. Nevertheless, this information is included to document the types of noteworthy responses for which the data were not edited.

B.2 Detailed Documentation of Edits for Selected Noncore Demographics Variables

This section presents tables describing edits for the noncore education and the employment and workplace variables from the noncore demographics.

- Table B.1 presents edits pertaining to variables in the noncore education section.
- Table B.2 presents edits pertaining to variables in the employment and workplace section.

Table B.1 Edits Pertaining to the Noncore Education Section

Issue	Edits Implemented
The R reported not being currently enrolled in school. In the question about reasons for leaving school without getting a high school diploma (LFSCHWHY), however, the R specified that he or she was still in school (LFSCHWHY = 778).	The R was logically inferred to be currently enrolled in school. A special code of 3 was assigned to the edited school enrollment variable SCHENRL. For the following variables, it was logically inferred that they should have skipped: HSDIPLMA (receipt of a high school diploma), HSGED (receipt of a GED certificate), and LFTSCHAG (age when the R left school). Consequently, any data in these items were overwritten in the edited variables with codes to indicate logical inference that these questions should have been skipped (see Section 2.4.2). Data were not overwritten for LFSCHWHY (reason for leaving school) because that was the variable responsible for inferring that the R was currently enrolled.
The R reported not being currently enrolled in school. In the question about reasons for leaving school without getting a high school diploma (LFSCHWHY), however, the R specified that he or she was being homeschooled (LFSCHWHY = 601).	The R was logically inferred to be currently enrolled in school. A special code of 5 was assigned to the edited school enrollment variable SCHENRL. As above, any data in HSDIPLMA, HSGED, and LFTSCHAG were overwritten. Data were not overwritten for LFSCHWHY because that was the variable responsible for inferring that the R was currently enrolled. If the R was aged 12 to 18, a code of 14 was assigned to the variable SCHTYPE (Homeschooled, LOGICALLY ASSIGNED) to indicate that the R was logically inferred to be homeschooled. No editing was done to SCHTYPE if the R was older than 18. In addition, this edit was not implemented if interviewers specified that the respondent had *completed* homeschooling. In these latter situations, LFSCHWHY was assigned a code other than 601 (see below).
The R reported being currently enrolled in school, and a hard error was triggered between QD11 and QD18. The FI's comments for suppressing the hard error indicated that the R was currently enrolled in technical or vocational school.	The R was logically inferred not to be currently enrolled in school. A special code of 4 was assigned to the edited school enrollment variable SCHENRL. For the following variables, it was logically inferred that they should have skipped: EDUCATND (current grade), SDNTFTPT (full- or part-time status), SCHDSICK (number of days in the past 30 days that the R missed school because the R was sick), and SCHDSKIP (number of days the R skipped school in the past 30 days). Consequently, any data in these items were overwritten in the edited variables with codes to indicate that these items should have been skipped.
The R did not know or refused to report whether he or she was enrolled in school, reported being on a holiday or break from school (QD17A = 1), but reported that he or she did not intend to return to school once the break was over (QD17B = 2).	The R was defined as not being enrolled in school (SCHENRL=2). The variables pertaining to the current grade through the number of days that the R skipped school in the past 30 days (EDUCATND, SDNTFTPT, SCHDSICK, and SCHDSKIP) were assigned legitimate skip codes.
The R was aged 23 or younger and reported not being enrolled in school. However, the interview was conducted in June, July, or August (i.e., when school was not in session). The R also originally reported getting a high school diploma but was inferred not to have received one (i.e., the R completed the 9th grade or lower). In addition, the highest completed grade was within 1 year (in either direction) of the expected completed grade for that R's age.	A code of 52 was assigned to the school enrollment variable SCHENRL only if there was some ambiguity (i.e., a response of "don't know" or "refused") in the follow-up questions QD17B or QD17C, regarding the R being on break from school or intending to return to school when the R's break was over, respectively. This code of 52 was intended to indicate to analysts that there was some uncertainty about the R's current enrollment status due to (a) the interview being conducted in summer months when most schools are not in session, and (b) ambiguity in the R's answers to QD17B or QD17C. Otherwise, no editing was done to SCHENRL (i.e., the R continued to be classified as not enrolled in school) when the R indicated in QD17B that he or she was not on break from or the R was on break but indicated in QD17C that he or she did not intend to return to school once this break was over.

Table B.1 Edits Pertaining to the Noncore Education Section (continued)

Issue	Edits Implemented
The current grade (QD18) was potentially inconsistent with the highest grade that the respondent (R) reported completing (QD11), and (a) the hard error between QD11 and QD18 was not triggered (e.g., the current grade from QD18 was two or more grades higher than the highest grade from QD11); or (b) the hard error was triggered and suppressed, but the field interviewer (FI) did not provide sufficient information to determine what corrections needed to be made.	An algorithm was developed that compared the self-reported current and highest grades with the R's current age (see text). A noncore edited variable for the highest grade completed (EDTEDUC) also was created. Edits were generally implemented as follows: • When both the current grade and the highest completed grade were potentially consistent with the R's age, the edits picked the response from QD18 or QD11 that would yield the most consistent data. The second variable in the pair then was edited for consistency with the response that was picked as being most consistent. • When the current grade was more consistent with the R's current age than was the reported highest grade from the core demographics, then EDTEDUC was logically assigned a code to indicate that the R had completed the lower grade that was adjacent to his or her current grade. • When the highest completed grade was more consistent with the R's current age than was the reported current grade, then the edited current grade (EDUCATND) was logically assigned a code to indicate that the R was in the next highest grade relative to the one he or she had completed, or else EDUCATND was coded as bad data. When neither the current grade (QD18) nor the highest completed grade (QD11) was consistent with the R's age, either EDTEDUC or EDUCATND (or both) were coded as bad data. If the current grade was exactly two grades higher than the last grade but the highest grade was lower than the expected highest grade, then EDTEDUC was coded as bad data. If the current grade was more than two grades higher than the last grade but the current grade was lower than the expected current grade, then EDUCATND was coded as bad data. If the current grade was lower than the highest grade, the one that was closest to the expected grade was chosen, and the other was set to bad data. If both EDTEDUC and EDUCATND were both close to their expected grades, both were set to bad data.
The R reported not being currently enrolled in school, reported receiving a high school diploma, but reported completing the 10th or 11th grade.	No editing was done, and the variable pertaining to receipt of a high school diploma (HSDIPLMA) retained a value of 1 (i.e., "yes"). The rationale was that the R may have gone through school on an accelerated pace or may have otherwise qualified for a high school diploma with fewer than 12 years of education (e.g., if the R went to school in another country).
The R reported not being enrolled in school but having received a high school diploma. However, the R had completed only the 9th grade or lower.	The R was logically inferred in HSDIPLMA not to have received a high school diploma. HSDIPLMA was assigned a code of 4 (No LOGICALLY ASSIGNED).
The R reported not being enrolled and not having received a high school diploma. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had gotten a diploma. That included situations where the R may have received a diploma in another country.	The R was logically inferred to have received a high school diploma, provided that the R had completed the 10th grade or higher. The edited variable HSDIPLMA was assigned a code of 3 (Yes LOGICALLY ASSIGNED). Although Rs would have skipped the question about getting a GED if they had answered "yes" to the question about getting a high school diploma (QD22), no editing was done to the GED variable HSGED (QD23) when Rs were logically inferred to have gotten a high school diploma. In addition, no editing was done to LFSCHWHY and LFTSCHAG in this situation.

Table B.1 Edits Pertaining to the Noncore Education Section (continued)

Issue	Edits Implemented
The R reported not being enrolled in school, not having received a high school diploma, and not having received a GED certificate. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had received a GED (LFSCHWHY = 606).	The R was logically inferred to have gotten a GED certificate. For this edit to be implemented, however, the R had to have indicated explicitly that he or she had actually received a GED, not that he or she was working on a GED.
The R reported not being enrolled in school, not having received a high school diploma, and having received a GED certificate. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had completed homeschooling but had not received a GED (LFSCHWHY = 621).	The R's response of "yes" in HSGED was considered to be questionable based on the response in LFSCHWHY. HSGED was set to a value of 11 (i.e., "bumped" by 10). For this edit to be implemented, however, the R had to have indicated explicitly that he or she had actually received no GED after having completed homeschooling.
The R reported leaving school at an age greater than his or her current age.	The edited variable corresponding to the age at leaving school was assigned a bad data code.
The R reported leaving school at age 3 or younger, or the R reported leaving school at an age that was considered too young for the highest grade that he or she reported completing (e.g., completed the 11th grade but reported leaving school at age 13 or younger).	The edited variable corresponding to the age at leaving school was assigned a bad data code.
The number of days in the past 30 that the R missed school because the R was sick and the number of days that the R skipped school added up to more than 30.	No editing was done in this situation.
The R reported being currently enrolled in school but reported skipping school all 30 days in the past 30 days.	A code of 11 was assigned to the school enrollment variable SCHENRL. This code was intended to indicate to analysts that there was some uncertainty about the R's current enrollment status.
The R reported being currently enrolled in school but reported in question QD20 that he or she missed school because of sickness for more than 30 days. This pattern was observed in earlier years because a code of 90 was used to mean "school not in session," and the computer-assisted interviewing (CAI) program code did not allow for discontinuities in the allowable range.	Values of 31 days were set to 30 days. Values greater than 31 days but fewer than 90 (i.e., school not in session) were replaced with bad data codes. This logic was in place, but values greater than 31 and fewer than 90 did not occur in the data.

Table B.2 Edits Pertaining to the Employment and Workplace Section

Issue	Edits Implemented
The R did not know or refused to report in question QD26 whether he or she worked in the past year. However, the R also reported in question QD33 (edited variable WRKEDYR) that he or she did not have a job in the past 12 months.	The R was logically inferred not to have worked in the past week (WRKEDWK = 4) and not to have had a job in the past week (WRKHAVJB = 4), where 4 = No LOGICALLY ASSIGNED. Subsequent employment and workplace variables that could be assigned legitimate skip codes were edited as though QD26 and QD27 had been answered as "no."
The R reported working in the past week. However, the R subsequently reported missing work for all 30 of the past 30 days because he or she was sick or did not want to be at work (or both). Because the past week was included in the 30 days prior to the interview, it would be inconsistent for an R to report working in the past week but missing work for every day in the past month.	The following edits were implemented in this situation: • If the R reported that he or she missed work for all 30 days in the past month because he or she was sick, the edited variable (WORKDAYS) was assigned a bad data code. • If the R reported missing work for all 30 days in the past month because he or she did not want to be there, the edited variable (WORKBLAH) was assigned a bad data code.
The respondent (R) reported working in the past week in question QD26. However, the R subsequently reported being without a job at some point in the past 12 months and reported being without a job during all 52 weeks in the past 12 months. Because all 52 weeks of the 12-month period prior to the interview would include the week prior to the interview, it would be inconsistent for an R to report working in the past week but not working for all 52 weeks in the past year.	The edited variable pertaining to the number of weeks without a job in the past 12 months (WRKUNWKS) was assigned a bad data code.
The R answered question QD26 (worked in the past week) as "no" but answered question QD27 (having a job in the past week) as "don't know." Edit logic prior to 2003 left the variable pertaining to the number of hours worked in the past week WRKHRSWK as blank (i.e., a legitimate skip code was not assigned to WRKHRSWK). The prior logic for assigning legitimate skip codes to WRKHRSWK was part of the logic for assigning legitimate skips when both QD26 and QD27 were answered as "no." However, only the response to question QD26 truly applies to WRKHRSWK.	Since 2003, a legitimate skip code has been assigned to WRKHRSWK when QD26 is answered as "no" (QD26 = 2), independent of how QD27 was answered.
The reported year when the R last worked for pay was fewer than 5 years from the R's birth year (including situations where the year the R reported last working for pay was earlier than the year the R was born).	The edited variables WRKLSTMN and WRKLSTYR were assigned bad data codes.

Table B.2 Edits Pertaining to the Employment and Workplace Section (continued)

Issue	Edits Implemented
The R was not asked whether he or she was self-employed in the past 12 months because the R had already given an answer indicating that he or she had been self-employed. This could occur in one of the ways listed below. • The R reported not working in the past week because he or she was self-employed and did not have any business (QD30 = 5). • The R reported in question INOC06 that the category that best described the business in which he or she worked was one in which the R was self-employed (INOC06 answered as 7 or 8).	The edited variable pertaining to self-employment in the past 12 months (WRKSLFEM) was assigned a code to indicate that "yes" could be logically inferred. This was done instead of assigning a legitimate skip code. This edit did not apply if INOC06 indicated that Rs worked without pay in a family business or farm.
The R did not report being self-employed at any time in the past 12 months but reported having a job. However, the industry and occupation (I&O) question pertaining to the R's last job (INOC08) indicated that the R was self-employed in an incorporated or unincorporated business (edited variable WRKBZCYR, corresponding to INOC08, had a value of 7 or 8).	The edited variable WRKSLFEM was logically inferred to have been answered as "yes," provided that the following conditions held: • The R reported working in the past year (WRKEDYR = 1), such that reported self-employment in INOC08 would pertain to self-employment in the past year. • The year and month that the R reported last working for pay (WRKLSTYR and WRKLSTMN) also were consistent with the R reporting that he or she worked in the past year. The following data in WRKLSTYR and WRKLSTMN were considered to be consistent (or at least not contradictory) with indications that the R worked in the past year (WRKEDYR = 1): • The R reported last working for pay in the current interview year. • The R reported last working for pay in the previous year, and the month that the R reported last working for pay was within 12 months of the interview date, or was the same month as the interview date. • The R reported last working for pay in the previous year, but the month that the R reported last working for pay had a missing value. In this situation, WRKEDYR = 1, and an indication of self-employment in INOC08 was still allowed to infer in WRKSLFEM that the R had been self-employed in the past 12 months. WRKSLFEM was not logically inferred to be "yes" if the R reported working in the past year (WRKEDYR = 1), WRKBZCYR = 7 or 8, but any of the following occurred: • The R reported last working for pay in the previous year, and the month that the R reported last working for pay was more than 12 months beyond the interview date. • The R had missing data for the year when he or she last worked for pay (e.g., if WRKLSTYR was refused). • A problem had been identified with the interview date that was stored by the computer-assisted interviewing (CAI) system while the interview was in progress.

B.3 Detailed Documentation of Edits for Core Substance Use Variables

This section presents tables that detail the edits that were applied to variables in the core substance use modules. Because data on use of methamphetamine and other stimulants from the noncore special drugs module were used to create core-plus-noncore (CPN) recency variables for these substances as described in Section 6.2.6, Section B.3 also includes details about the editing of variables that were relevant to creating these CPN recency variables.

- Tables B.3 and B.4 present edits pertaining to variables for lifetime substance use, corresponding to the discussion of procedures in Section 6.2.1.
- Tables B.5 to B.7 present edits pertaining to variables for the frequency of use of different substances in the past 12 months or past 30 days, corresponding to the discussion of procedures in Section 6.2.4. These tables also document procedures for editing the frequency-of-use variables for "parent/child" drug combinations (e.g., pain relievers and OxyContin®) when the only substance that the respondent used was the child drug (Section 6.2.3).
- Tables B.8 to B.10 present edits pertaining to variables for incidence (i.e., initiation of use), corresponding to the discussion of editing procedures in Section 6.2.5. These tables also document procedures for editing the incidence variables for "parent/child" drug combinations when the only substance that the respondent used was the child drug (Section 6.2.3), as well as a few edits that are applied during imputation processing (Section 6.2.8).
- Table B.11 presents edits for variables on methamphetamine prevalence, stimulant prevalence, and related follow-up variables in the noncore special drugs, corresponding to the discussion of procedures in Section 6.2.6 for creating the CPN recency variables for these substances.
- Finally in this section, Table B.12 presents miscellaneous edits pertaining to the tobacco variables.

Table B.3 Edits Pertaining to Lifetime Variables for Hallucinogens through Sedatives Based on "OTHER, Specify" Data

Issue	Edits Implemented
Respondents (Rs) could specify something that corresponded to a drug in the list. For example, if an R specified use of "marijuana laced with PCP" as "some other hallucinogen," this response would indicate PCP use—even if such use had not previously been reported.	If the corresponding drug in the list was not answered as "yes," then "yes" was logically inferred, and the "specify" response was retained to indicate to analysts the source of the logical inference.
Rs could use street names or slang terms when specifying their use of some other drug besides the ones they previously had been asked about. Use of these slang terms to infer use of a drug that Rs had already been asked about (e.g., LSD, PCP) could be problematic, however, because the meaning of a particular slang term could vary by region, could apply to more than one drug, or could change over time.	Use of street or slang terms to infer lifetime use of a specific drug was generally restricted to situations where that term was supplied to respondents as a synonym for that drug. For example, "angel dust" was listed in question LS01b as an alternative term for PCP. Thus, situations where Rs specified angel dust were used to infer lifetime use of PCP, if question LS01b had not already been answered as "yes." However, other potential slang terms for PCP (e.g., "boat") that were not listed in question LS01b were not used in editing. Additional situations where slang terms were classified with particular drugs included "shrooms" (coded as psilocybin/mushrooms), "X" and "X-TC" (coded as Ecstasy), "Roofies" (coded as Rohypnol®), and "rock" (coded as crack cocaine). In addition, question ST01 in the stimulants module listed "speed" as an alternative term for methamphetamine. However, the decision was made that indications of speed in and of themselves should not be used to infer lifetime methamphetamine use because speed may often be used to refer to other stimulants or to stimulants as a group, not just methamphetamine.
Rs could specify use of some other drug within a particular category of drugs, but the drug being mentioned applies to another drug category covered in the survey (e.g., specifying Valium®, a tranquilizer, when asked about nonmedical use of pain relievers).	No editing across modules was done when these types of responses occurred (Section 2.4.4). However, the "OTHER, Specify" data within a module indicate to analysts when Rs have specified use of a drug that fits another category. This approach assumed that some Rs may specify the use of other drugs according to their *functional* properties. For example, Rs may specify certain tranquilizers in the sedatives section because tranquilizers can cause drowsiness. Similarly, the definition of hallucinogens (i.e., drugs that "often cause people to see or experience things that are not real") could apply to other drugs besides hallucinogens that alter one's perception of reality.

Table B.4 Edits Pertaining to Reports of Over-the-Counter Drugs in "OTHER, Specify" Data for Pain Relievers through Sedatives

Issue	Edits Implemented
Rs may indicate lifetime use of one or more prescription-type psychotherapeutics (e.g., Darvon®, Darvocet-N®, or Tylenol® with codeine) but specify only over-the-counter (OTC) medications as the other drug that they used (e.g., specifying only Tylenol® as the other prescription pain reliever that they used nonmedically).	Because Rs were instructed in the psychotherapeutics section not to include use of OTC medications, edits in this situation inferred that Rs logically had not used some other drug in that category. However, Rs who reported lifetime nonmedical use of a prescription-type medication in response to a previous question were still at least lifetime abusers of that category of drugs. This edit also was done in situations where Rs answered at least one prescription-type gate question as "don't know" (DK) or "refused" (REF). This edit was *not* done if Rs reported use of "some other" psychotherapeutic drug and (a) specified use of a prescription-type medication and an OTC drug, or (b) had some response of DK or REF along with the indication of OTC use. In this latter situation, the assumption was that a response of DK or REF meant that the R was still potentially a nonmedical user of a prescription-type drug.
The *only* indication of lifetime nonmedical use was an OTC drug that was specified. (In this situation, unlike the one described above, Rs have denied ever using other prescription-type medications that were covered in the particular category.)	Because Rs were instructed in the psychotherapeutics sections not to include use of OTC medications, edits in this situation inferred that the R had never used prescription or street psychotherapeutics. This edit was implemented if only OTC drugs were specified, or the data included only OTC drugs and codes for "bad data." This edit was not done in situations where one or more other gate questions for a type of psychotherapeutic medication were answered as DK or REF because that was not conclusive evidence that the R had *never* used. This edit also was not done in situations where the "OTHER, Specify" data included responses of DK or REF in addition to reports of OTC drugs (i.e., but no reports of nonmedical use of prescription-type drug). As for the previous issue, responses of DK or REF were interpreted to mean that the R was still potentially a nonmedical user of a prescription-type drug.

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use

Issue	Edits Implemented
Number of days that the respondent (R) used in the past 12 months is greater than 365 days (e.g., if the R reported using 31 days per month on average, then 31 x 12 = 372 days).	Nonmissing values greater than 365 were trimmed back to a maximum of 365 days. A flag was set to indicate that this trimming was done.
Number of days that the R reported using in the past 30 days implies less use in the past 12 months than what the R reported in the 12-month frequency (e.g., if the R reported using on only 1 day in the past 30 days and did not use on 29 days, the maximum number of days that the R could have used in the past 12 months would be 336 days); this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	The 12-month frequency was trimmed to be consistent with the number of days that the R reported using in the past 30 days. The relevant flag variable indicated that this trimming was done.
Rs who answered the 30-day frequency questions as "don't know" (DK) or "refused" (REF) had the opportunity to give their best estimate of the number of days they used in the past 30 days. These best estimate variables were categorical as opposed to being continuous (e.g., 1 = 1 or 2 days, 2 = 3 to 5 days). For alcohol through inhalants, the maximum possible value of the "estimated" 30-day frequency could imply less use in the past 12 months than what the R reported in the 12-month frequency. For example, if the R reported using a drug on "10 to 19 days" in the past 30 days, the R would not have used the drug on at least 11 days in that period. It would therefore be inconsistent for the R to have reported using the drug on more than 354 days in the past 12 months.	The 12-month frequency was trimmed to be consistent with the minimum number of days that the R estimated using in the past 30 days. For example, if the R estimated using a drug on "10 to 19" days in the past 30 days, it was assumed that the R could have used on 345 days in the past 12 months, which would be consistent with use on 10 days in the past 30 days (and nonuse on 20 days). This value would not be considerably different from that derived by assuming use on the maximum number of estimated days in the past 30 days (e.g., 354 possible days of use in the past 12 months, if the R reported using on 10 to 19 days in the past 30 days).
For Rs who estimated their 30-day frequency, the final 12-month frequency was consistent with some, but not all, values in the range for the estimated 30-day frequency.	No editing was done to the 12-month frequency. For example, if an R reported using a drug on "6 to 9 days" in the past 30 days and reported using on 8 days in the past 12 months, the 12-month frequency was within the range of 6 to 8 days. Thus, a 12-month frequency of 8 days would be consistent with the estimated 30-day frequency, as long as the R used on 6 to 8 days in the past 30 days (and did not use on 9 days in that period). In this situation, minimum and maximum possible values were created for use by the statistical imputation team in assigning a final 30-day frequency. This procedure narrowed the allowable range of the 30-day frequency for consistency with the 12-month frequency, with the final 30-day frequency being picked from within that narrowed range. In the above example, the allowable range for assigning the final 30-day frequency was 6 to 8 days rather than 6 to 9 days.

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
R initiated use at some point in the past 12 months. Relative to when the R was interviewed, the R therefore could not have used over the entire 12-month period.	The value of the 12-month frequency was reduced (i.e., prorated) according to the maximum possible allowable number of days that the R could have used in the past 12 months, relative to when the R was interviewed. If the R answered the 12-month frequency in terms of a total number of days used in that period and the answer was greater than the maximum number of days that the R could have used, the 12-month frequency was reduced to this upper limit. The maximum possible allowable number of days was determined from the interview date and the month- and year-of-first use or the date of the R's last birthday, whichever yielded the smallest number of days that the R could have used. If the R reported first using in the survey year but did not report the month when he or she first used, it was assumed that the R potentially started use in January of the survey year.
The R initiated use in the past 12 months but definitely initiated use more than 30 days prior to the interview date. The R also reported use on only 1 day in the past 12 months. However, the recency had a value of 1 (Used in the past 30 days) or 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED).	The 12-month frequency of 1 day was set to bad data. If the R last used in the past 30 days and the R initiated use more than 30 days ago but within the past 12 months, the R logically would have used on at least 1 day in the past 30 days and on at least 1 other day in the period that was more than 30 days ago but within the past 12 months. Similarly, a recency of 8 indicated that there was uncertainty about whether the R last used in the past 30 days or more than 30 days ago but within the past 12 months. Consequently, use on only 1 day in the past 12 months and initiation more than 30 days ago within the past 12 months would have forced the recency to be more than 30 days ago. In addition, when the recency had indicated use in the past 30 days and the above edit had changed a 12-month frequency from 1 day to bad data, whatever values that existed in the 30-day variables were retained. In particular, if the 30-day frequency had a missing value, that missing value was retained. (Other edits described below would have logically inferred that the R used on 1 day in the past 30 days if the R also had indicated use on exactly 1 day in the past 12 months.)
Information was needed for establishing constraints on donors for the imputation of the 12-month frequency-of-use variables. For example, if a 12-month frequency variable had been changed from 1 day to bad data, the imputation team needed to constrain the imputation of the 12-month frequency to a value greater than 1 day.	Indicator flags were created to aid the statistical imputation team in establishing specific constraints for the imputation of the 12-month frequency-of-use variables. The indicator flags had the following values: 1 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN 1 DAY. 2 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN RECIPIENT'S 30-DAY FREQUENCY OF USE. (Did not apply to the psychotherapeutics, which do not have 30-day frequency variables.) 3 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN RECIPIENT'S 30-DAY FREQUENCY OF BINGE ALCOHOL USE. (Applied only to alcohol).

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
R initiated use at some point in the past 30 days. Relative to when the R was interviewed, however, the 30-day frequency (or the minimum value of the "estimated" 30-day frequency) exceeded the maximum possible number of days that the R could have used the drug in that period.	If the R gave an exact number of days for the 30-day frequency, the value was reduced to agree with the maximum possible number of days that the R could have used the drug. If the R estimated the 30-day frequency as a range of a possible number of days, the range for the estimated 30-day frequency was reduced to agree with the maximum possible number of days that the R could have used. For example, if the R estimated using a drug on 20 to 29 days in the past 30 days but could have used the drug at most on 15 days in that period, the estimated 30-day frequency was revised to the range indicating use on 10 to 19 days.
The 30-day frequency was greater than 12-month frequency; this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	The value for the 30-day frequency was retained, and the 12-month frequency was set to a missing value (i.e., bad data). (Any of the corresponding 12-month frequency source variables that the respondent answered were set to bad data as well.) This edit conserved data reported by the R that suggest more frequent use in the past 30 days. Although both the 12-month and 30-day frequencies could have been set to missing (i.e., bad data) and subsequently imputed, a drawback of this approach is that a donor could be chosen whose 30-day and 12-month frequency data were consistent but whose data were considerably different from what the recipient R originally reported, especially if the recipient R's original answer to the 30-day frequency was at or close to 30.
For cigarettes, chewing tobacco, snuff, and cigars, Rs who answered the 30-day frequency question as DK or REF were asked to give their best estimate of the number of days they used these tobacco products in the past 30 days. These best estimate variables were categorical as opposed to being continuous (e.g., 1 = 1 or 2 days, 2 = 3 to 5 days). Because there were no 12-month frequency questions for tobacco products, inconsistencies would not exist between these "estimated" 30-day frequency variables and other data for these tobacco products.	No editing was done to the corresponding 30-day frequency variables (i.e., leaving the original answer of DK or REF). A final value was assigned by the statistical imputation team.
For alcohol through inhalants, if Rs gave their best estimate of the number of days they used in the past 30 days, it was possible for the bottom end of the range for an "estimated" 30-day frequency to exceed the value for the 12-month frequency (i.e., after any opportunities that the R had for inconsistency resolution).	The value indicating the range for the "estimated" 30-day frequency was retained, and the 12-month frequency was set to a missing value (i.e., bad data). For example, if the R reported using a drug on 5 days in the past 12 months but estimated use of that drug on 6 to 9 days in the past 30 days, there would be no way for these two answers to be consistent. Setting the 12-month frequency to bad data in this situation is consistent with the edit when Rs report an exact number of days for the final 30-day frequency that is greater than the final 12-month frequency.

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
The bottom end of the range for the final "estimated" number of days that the R used a drug in the past 30 days is *exactly* equal to the final 12-month frequency; this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	The final, edited 30-day frequency was set to the value for the 12-month frequency. The code was retained for the "estimated" 30-day frequency to indicate to analysts where the value came from for the 30-day frequency. For example, if the R reported using a drug on exactly 6 days in the past 12 months and estimated using that drug on 6 to 9 days in the past 30 days, the 30-day frequency was set to exactly 6 days. The two answers in this example are not necessarily inconsistent, as long as the R used on exactly 6 days in the past month. Thus, the response of 6 days in the past 12 months could be thought of as "trapping" the possible number of days that the R could have used in the past 30 days.
The 30-day frequency has a final response of DK or REF (i.e., the R did not give a best estimate for the 30-day frequency), but the 12-month frequency indicates use on exactly 1 day.	The final, edited 30-day frequency was logically inferred to be 1 day. If the R last used in the past 30 days and used on only 1 day in the past 12 months, that 1 day of use had to have occurred in the 30 days prior to the interview.
The value for the final 12-month frequency is somewhere within the range for the final "estimated" number of days that the R used a drug in the past 30 days; this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	No further editing was done because the data were not necessarily inconsistent. For example, if the R reported using a drug on 8 days in the past 12 months and estimated use of that drug on 6 to 9 days in the past 30 days, use of the drug on 6, 7, or 8 days in the past 30 days would still be consistent with use on 8 days in the past 12 months. In this situation, minimum and maximum possible values were created for use by the statistical imputation team in assigning a final 30-day frequency. This procedure narrowed the allowable range of the 30-day frequency for consistency with the 12-month frequency, with the final 30-day frequency being picked from within that narrowed range. In the above example where the R reported using on 8 days in the past 12 months and 6 to 9 days in the past 30 days, the allowable range for the 30-day frequency would be 6 to 8 days, not 6 to 9, in order for the 30-day frequency to be consistent with the 12-month frequency.

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
Overall 12-month frequency does not agree with the reported preference for giving the answer (e.g., if the R originally indicated that he or she would give an answer in terms of a total number of days per year, answered the days per year question as DK/REF, and then gave an answer in average days per month). This is a function of the computer-assisted interviewing (CAI) logic, which keeps routing Rs through the series until they give a nonmissing answer or answer all 12-month frequency questions as DK/REF; this applied to the alcohol through psychotherapeutics sections. Similarly, the R could change his or her preferred way of reporting the 12-month frequency in response to a consistency check with the 30-day frequency (e.g., starting out as reporting in terms of the total number of days in the past 12 months but changing to average number of days per month in response to a consistency check); this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	The preferred way of reporting the 12-month frequency that best agreed with the final answer to the 12-month frequency was assigned to the "best way" variable (e.g., ALBSTWAY for alcohol), pertaining to the R's preferred way of answering the 12-month frequency questions. Thus, if an R started out as preferring to give an answer in total days per year but the final answer came from an average number of days per month, the preferred method of reporting would be consistent with the reporting in days per month. Any other 12-month frequency questions that the R answered that did not correspond with the final result were overwritten with LEGITIMATE SKIP Logically assigned codes. In the above example, the initial answer in terms of the total number of days used in the past 12 months would be overwritten, and the final answer of average number of days per month would be retained.

Table B.6 Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine, Pain Relievers, and Stimulants

Issue	Edits Implemented
The 12-month frequency of methamphetamine use was greater than the corresponding frequency of any stimulant use. The same issue applied in the pain relievers module when the 12-month frequency of OxyContin® use was greater than the corresponding frequency of any pain reliever use.	The higher value from the methamphetamine frequency was assigned to the stimulant frequency. This also involved moving over the source variables from methamphetamine that were associated with the overall methamphetamine frequency. In addition, the methamphetamine variable associated with the preferred method of reporting the 12-month frequency (STBSTWAY) was bumped by a value of 20 to indicate that data had been moved over from the methamphetamine frequency variables. The same edits described above for methamphetamine and any stimulants were applied to the 12-month frequency variables for any pain relievers when the 12-month frequency of OxyContin® use was greater than the corresponding frequency of any pain reliever use.
The 30-day or 12-month frequency of crack use is greater than the corresponding frequency of any cocaine use.	The higher value from the crack frequency was assigned to the cocaine frequency. For the 12-month frequency, this also involved moving over the source variables from crack that were associated with the overall crack frequency. In addition, the cocaine variable associated with the preferred method of reporting the 12-month frequency (CCBSTWAY) was bumped by a value of 20 to indicate that data had been moved over from the crack cocaine variables.
Respondents (Rs) indicated that their month of first use for any cocaine, any pain relievers, or any stimulants was in the same month they were interviewed, and they used cocaine, pain relievers, or stimulants on only 1 day in the past 12 months. However, the Rs also were users of crack cocaine, OxyContin®, or methamphetamine, respectively, and they had missing data for the corresponding 12-month frequency variables pertaining to crack cocaine, OxyContin®, or methamphetamine.	Logically, these Rs also would have used crack cocaine, OxyContin®, or methamphetamine on 1 day in the past 12 months. Therefore, the corresponding 12-month frequency variables for crack cocaine, methamphetamine, or OxyContin® were set to 1 day in this situation.
Rs indicated that their month of first use for any cocaine, any pain relievers, or any stimulants was in the same month they were interviewed, and they used cocaine, pain relievers, or stimulants on only 1 day in the past 12 months. However, the Rs also were users of crack cocaine, OxyContin®, or methamphetamine, respectively, and they had missing data for the corresponding 12-month frequency variables pertaining to crack cocaine, OxyContin®, or methamphetamine.	Logically, these Rs also would have used crack cocaine, OxyContin®, or methamphetamine on 1 day in the past 12 months. Therefore, the corresponding 12-month frequency variables for crack cocaine, methamphetamine, or OxyContin® were set to 1 day in this situation.

Table B.6 Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine, Pain Relievers, and Stimulants (continued)

Issue	Edits Implemented
In the stimulants module, the R was a lifetime user only of methamphetamine and reported last using methamphetamine in the past 30 days or more than 30 days ago but within the past 12 months. The same issue applied in the pain relievers module when the R was a lifetime user only of OxyContin® and reported last using OxyContin® at some point in the past 12 months.	Values from the methamphetamine 12-month frequency variables were assigned to the corresponding stimulant 12-month frequency variables. The variable STBSTWAY (preferred way of answering the stimulant 12-month frequency questions, corresponding to question ST10) was assigned a code in the 20s to indicate that data had been assigned from the methamphetamine 12-month frequency variables. Similarly, when the R was a user only of OxyContin®, values from the OxyContin® 12-month frequency variables were assigned to the corresponding pain reliever 12-month frequency variables, and the variable PRBSTWAY (preferred way of answering the pain reliever 12-month frequency questions) was assigned a code in the 20s.
The R was skipped out of the 12-month or 30-day frequency questions for cocaine because of his or her original answer to the cocaine recency, but nonmissing values existed in the corresponding frequency variables for crack. For example, if an R reported last using any cocaine more than 30 days ago but within the past 12 months, but reported last using crack within the past 30 days, 30-day frequency data could exist for crack but not for any cocaine.	No editing was done to the 12-month or 30-day cocaine frequency that had been skipped. In the imputation stage, the cocaine 12-month or 30-day frequency was subsequently edited or imputed for consistency with the corresponding crack frequency data.
The R estimated the frequency of use of both cocaine and crack in the past 30 days, with the estimated frequency for crack being greater than the estimated frequency for any cocaine.	The "estimated" value for the crack 30-day frequency (CK30EST) was assigned as the estimate for the number of days that the R used any cocaine in that period. The code assigned to the "estimated" cocaine 30-day frequency variable CC30EST was bumped by a value of 10 to indicate that data had been moved over from CK30EST.
The "estimated" 30-day frequency of crack use is greater than the total number of days that the R reported using any cocaine in that same period.	The "estimated" value for the crack 30-day frequency (CK30EST) was assigned as the estimate for the number of days that the R used any cocaine in that period. The code assigned to the "estimated" cocaine 30-day frequency variable CC30EST was bumped by a value of 10 to indicate that data had been moved over from CK30EST. The edited cocaine 30-day frequency variable COCUS30A was assigned a bad data code to wipe out the R's original answer. For example, if the R reported using any cocaine on exactly 5 days in the past 30 days but reported using crack on 6 to 9 days in that period (i.e., a code of 3 in question CK06DKRE), the R was logically inferred to have used any cocaine on 6 to 9 days in the past 30 days. A code of 13 was assigned to CC30EST, corresponding to the code of 3 in CK30EST. In this example, it would be impossible for the R to have used crack on at least 6 days in the past 30 days but to have used any cocaine on only 5 days in that same period.

Table B.6 Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine, Pain Relievers, and Stimulants (continued)

Issue	Edits Implemented
The value for the cocaine 30-day frequency is somewhere within the range given by the "estimated" crack 30-day frequency.	Examination of preliminary data from 2000 indicated that all cocaine use in the past 30 days could not be automatically inferred to be crack use. Therefore, if the value for the cocaine 30-day frequency was the minimum possible number of days that the R could have used crack in that period, the crack 30-day frequency CRKUS30A was set to be equal to the cocaine 30-day frequency COCUS30A. For example, if the R reported using cocaine on exactly 6 days in the past 30 days and estimated using crack on 6 to 9 days in that period, it could be logically inferred that the R used crack on exactly 6 days. If the value for the cocaine 30-day frequency was greater than the minimum value for the estimated crack frequency, no further editing was done. However, information was provided to the statistical imputation team regarding the narrower range in which to assign a final value for the crack 30-day frequency. For example, if the R reported using cocaine on 7 days and crack on 6 to 9 days, the allowable range for the final crack 30-day frequency would be 6 or 7 days.

Table B.7 Edits Pertaining to the 30-Day Alcohol Frequency Variables Related to Binge Alcohol Use

Issue	Edits Implemented
Frequency of consumption of five or more drinks in the past 30 days (DR5DAY, from AL08) is greater than the overall frequency of consumption of any alcohol in the past 30 days (ALCDAYS, from AL06), despite the presence of a consistency check.	If the respondent (R) reported having five or more drinks per occasion on all 30 days, then the edits inferred that the R drank on all 30 days. If the R reported having five or more drinks per occasion on fewer than 30 days, then the value from the five-drink frequency DR5DAY was retained and the overall 30-day frequency ALCDAYS was set to missing (i.e., bad data). This approach conserved data reported by the R that could indicate heavy alcohol use (i.e., consumption of five or more drinks on 5 or more days in the past 30 days).
Frequency of consumption of five or more drinks in the past 30 days (DR5DAY) is greater than the maximum possible value for the estimated frequency of consumption of alcohol in the past 30 days (AL30EST, from AL06DKRE).	If the R reported having five or more drinks per occasion on all 30 days, then the edits inferred that the R drank on all 30 days (i.e., the variable ALCDAYS was assigned a value of 30). This edit also was done if the R reported having five or more drinks per occasion on all 30 days and the R continued to answer the follow-up question AL06DKRE as DK or REF. If the R reported having five or more drinks per occasion on fewer than 30 days, then the value from DR5DAY was retained, and the estimated 30-day frequency variable AL30EST was assigned a bad data value. For example, if an R reported having five or more drinks per occasion on 10 days in the past 30 days but estimated drinking alcohol on 6 to 9 days in that period, it would be impossible for these two answers to be consistent.
R drank on 1 day in the past month and the usual number of drinks was five or more on this 1 day, but frequency of consumption of five or more drinks per occasion is 0 (or is answered as DK/REF).	A value of 1 day was assigned to DR5DAY. If the R reported having five or more drinks on the 1 day when he or she drank, then the R logically had to have had five or more drinks on exactly 1 day in the past month.
R drank on 1 day in the past month and the usual number of drinks was fewer than five, but the frequency of consumption of five or more drinks per occasion is answered as DK/REF.	A code of 80 was assigned to DR5DAY, where 80 = NO OCCASION OF 5+ DRINKS IN PAST 30 DAYS Logically assigned. If the R reported having fewer than five drinks on the 1 day when he or she drank, then the R logically had no occasions of consuming five or more drinks in the past month.
Frequency of consumption of five or more drinks in the past 30 days (DR5DAY) falls within the range of the estimated frequency of consumption of alcohol in the past 30 days (AL30EST).	No further editing was done because the data were not necessarily inconsistent. For example, if the R reported drinking any alcohol on 6 to 9 days in the past 30 days and reported having five or more drinks per occasion on 7 days in that period, these data would be consistent as long as the R drank any alcohol on 7, 8, or 9 days in the past 30 days. In this situation, minimum and maximum possible values were created for use by the statistical imputation team in assigning a final 30-day frequency for alcohol. This procedure narrowed the allowable range of the alcohol 30-day frequency variable ALCDAYS for consistency with the five-drink frequency variable DR5DAY. The final 30-day frequency for alcohol was picked from within that narrowed range. In the above example where the R reported having five or more drinks per occasion on 7 days in the past 30 days and 6 to 9 days in that period, the allowable range for the 30-day frequency would be 7 to 9 days, not 6 to 9, in order for ALCDAYS to be consistent with DR5DAY.

Table B.7 Edits Pertaining to the 30-Day Alcohol Frequency Variables Related to Binge Alcohol Use (continued)

Issue	Edits Implemented
The R drank on more than 1 day in the past month and the usual number of drinks per day was five or more, but the reported frequency of five or more drinks per occasion is 0.	*Both* the usual number of drinks (i.e., NODR30A, from AL07) and DR5DAY (from AL08) were set to bad data. This approach was consistent with cognitive testing results that suggested that Rs sometimes incorrectly answer AL07 by indicating the total number of drinks they had *over the entire 30-day period* instead of the usual number of drinks they had per day. Consequently, these edits did not automatically assume that these Rs were binge alcohol users. If DR5DAY had a value of DK/REF and the usual number of drinks reported was five or more, then DR5DAY retained the corresponding DK/REF value.
The number of days that the R reported having five or more drinks equals the total number of days that the R reported drinking any alcohol (DR5DAY = ALCDAYS), but the reported usual number of drinks is fewer than five (or is answered as DK/REF).	A code of 975 was assigned to NODR30A (i.e., the edited version of AL07) to indicate usual consumption of at least five drinks.

Table B.8 Edits Pertaining to Incidence Variables Based on Consistency Checks

Issue	Edits Implemented
The respondent (R) indicated in the final verification check (e.g., CGCC22 for cigarettes) that the age at first use (AFU) based on the new month and year of first use (MFU and YFU) was correct (e.g., CGCC22 = 4). The computer-assisted interviewing (CAI) program updated the AFU (such as AGE1STCG for cigarettes) with the value of the age calculated from the MFU and YFU (referred to subsequently as the MYR1ST age, such as MYR1STCG, for cigarettes). However, the new value for the AFU indicated that the R was more than 1 year younger than his or her current age at the time the R first used the drug (e.g., the R was 16, reported first use of the drug at age 15, but then confirmed an MFU and YFU that meant the R was 14 when the R first used the drug). Had the R initially reported this AFU, the R would not have been routed to the MFU and YFU questions.	The updated value was retained for the AFU (e.g., first use at age 14 for a 16-year-old R in this example). Based on this updated AFU, it was logically inferred that the R should have skipped the MFU and YFU items. A code of 9989 was assigned to the YFU variable (e.g., CIGYFU for cigarettes), and a code of 89 was assigned to the MFU variable (e.g., CIGMFU). Consistent with procedures in prior years, therefore, the values for the MFU and YFU were statistically imputed.
The final verification check (e.g., CGCC22) was skipped because the R entered revised data for the MFU and YFU that made them consistent with the AFU.	No editing was done because the R was considered to have resolved the inconsistency.
The final verification check (e.g., CGCC22) was skipped because the R entered a new MFU that was the same as the R's birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, the MFU and YFU were set to bad data if they could never be consistent with the AFU. Suppose, for example, that a hypothetical R was born in June 1996, was interviewed in March 2014 (age 17 at the time of the interview), reported first use of a drug at age 17, and initially reported first use in May 2013. First use in May 2013 would have meant that the R was 16 when he or she first used the drug because the R's 17th birthday was not until June 2013. If the R changed the month and year to June 2013, that could be consistent with first use at age 17, if the use occurred after the R's birthday. However, if the R changed the month and year to June 2012, it would never be possible for the R to have first used at age 17 and also to have first used in June 2012. In this latter situation, the MFU and YFU would be set to bad data.
The R entered a new MFU or YFU that differed from what the R previously reported. The MYR1ST age based on the revised MFU and YFU still mismatched the AFU, but the R indicated in the final verification check that this MYR1ST age was correct.	No editing was done in this situation. The CAI program automatically updated the AFU to be consistent with the updated values reported for the MFU and YFU.

Table B.8 Edits Pertaining to Incidence Variables Based on Consistency Checks (continued)

Issue	Edits Implemented
The consistency check was triggered between the AFU and the MYR1ST age calculated from the MFU and YFU. However, the R answered the first consistency check (e.g., CGCC19 for cigarettes) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The AFU was retained, but the MFU and YFU were set to bad data. Consequently, the MFU and YFU needed to be imputed, but the AFU did not.
The R entered values in the consistency checks for the MFU and YFU that again yielded a nonmissing MYR1ST age based on the MFU and YFU. However, the R failed to resolve the inconsistency between the AFU and the MYR1ST age. The R also reported either in the first verification check that the MYR1ST age was not correct (e.g., CGCC19 = 6 for cigarettes) or reported in the second verification check that the AFU was correct (e.g., CGCC20 = 2 for cigarettes).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: • The default edit was to set the MFU and YFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the MFU and YFU were assigned the code that corresponded to the answer in the final verification check.
The R entered values in the consistency checks for the MFU and YFU that again yielded a nonmissing MYR1ST age based on the MFU and YFU. However, the R failed to resolve the inconsistency between the AFU and the MYR1ST age. The R also reported in the second verification check that neither the AFU nor original MYR1ST age was correct (e.g., CGCC20 = 3 for cigarettes).	The following edits were implemented for the AFU, MFU, and YFU: • The default edit was to set the AFU, MFU, and YFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the AFU, MFU, and YFU were assigned the code that corresponded to the answer in the final verification check.
The R triggered the initial consistency check between the AFU, MFU, and YFU and reported that the MYR1ST age calculated from the MFU and YFU was correct (e.g., CGCC19 = 4). However, the R answered the second consistency check (e.g., CGCC20 for cigarettes) as "don't know" or "refused." Consequently, the R did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for "don't know" or "refused" were assigned to the AFU, MFU, and YFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R's correct age when he or she first used a drug, or whether the MFU and YFU indicated the R's correct age at initiation. Therefore, the AFU, MFU, and YFU all were set to missing values.

Table B.9 Edits Pertaining to Incidence Variables Other than for Parent/Child Relationships

Issue	Edits Implemented
The age at first use (AFU) was greater than the respondent's (R's) reported age; this applied to all drugs except pipe tobacco.	The final age was accepted as the standard, and the inconsistent AFU was set to bad data. Any month and year data associated with the AFU also were set to bad data.
The AFU had a value of "don't know" (DK) or "refused" (REF), including situations where this assignment has been made from the consistency check data.	If the MFU and YFU questions were skipped because the R answered the AFU as DK or REF to begin with, the DK or REF value from the AFU was propagated onto the skipped MFU and YFU variables. This edit was designed to indicate the reason that the MFU and YFU variables had been skipped. In addition, because the R may have first used the drug within 1 year of his or her current age, the MFU and YFU questions may have been relevant to the R. If the R had answered the MFU and YFU questions but the final AFU had a value of DK or REF (i.e., due to a consistency check response), the MFU and YFU data were overwritten with the corresponding DK or REF value from the edited AFU. Retaining the MFU and YFU data in this situation would imply that the R used within 1 year of his or her current age.
The MFU and/or YFU questions were answered because the initial answer to an AFU was within 1 year of the R's current age. However, the final, edited AFU is more than 1 year younger than the R's current age based on the consistency checks between the AFU, YFU, and MFU that were described in Table B.8.	The original answers to the MFU and YFU were overwritten with logically assigned legitimate skip codes (Section 2.4.2). If the R originally had answered the relevant AFU as being more than 1 year younger than his or her current age, the CAI program would have skipped the R past the questions about the MFU and YFU.
The MFU has been skipped because the R answered the YFU as DK or REF.	The DK or REF value from the YFU was propagated onto the skipped MFU. That is, if the R did not know in what year he or she first used the drug, it was assumed that the R would not know the month either. Similarly, a refusal to answer the YFU was interpreted to be a blanket refusal to answer the month as well as the year.
The R had missing data for the MFU. However, one of the following occurred: • The YFU was the current calendar year, and the R was interviewed in January. • The R first used at his or her current age, the R first used in the current calendar year, and the R's most recent birthday occurred in the interview month. • The R first used at his or her current age, the R first used in the prior calendar year, and the R's most recent birthday occurred in December.	The MFU could be logically inferred, as indicated below. • The MFU was logically inferred to be January of the interview year. That would be the only month in which the R could have initiated use in the current year. • The MFU was logically inferred to be the interview month. If the R first used in the current calendar year and attained his or her current age in the interview month, the R logically could not have initiated use in any month other than the interview month. • The MFU was logically inferred to be December. If the R first used in the prior calendar year and attained his or her current age in December of that year, the R logically could not have initiated use in any month other than December of the prior calendar year.

Table B.10 Edits Pertaining to Parent/Child Incidence Variables

Issue	Edits Implemented
The age at first use (AFU) for a child drug (e.g., LSDAGE for LSD) indicated first use at age 1 or 2, but the AFU for the parent drug (e.g., HALLAGE for any hallucinogen) had a missing value. This applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack, hallucinogens/LSD/PCP/ Ecstasy, pain relievers/ OxyContin®, and stimulants/methamphetamine.	Ages of 1 or 2 in the AFU variables for child drugs (e.g., LSDAGE) were set to bad data. The imputation procedures for AFU variables did not allow Rs with AFUs of 1 or 2 to be donors for imputation. Logically, however, an AFU of 1 or 2 for a child drug would preclude the AFU for the parent drug from being imputed to any value other than 1 or 2. For example, if LSDAGE was 1 or 2 and the AFU for any hallucinogen (HALLAGE) had a missing value, the only potential imputation donors for HALLAGE would be those Rs with HALLAGE of 1 or 2. By definition, however, no one could be a donor for imputation. By setting LSDAGE to bad data, the AFUs for both LSD and any hallucinogen would be imputed.
The AFU for a child drug (e.g., LSDAGE) indicated first use at age 3, but the AFU for the parent drug (e.g., HALLAGE) had a missing value. This applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack, hallucinogens/LSD/PCP/ Ecstasy, pain relievers/OxyContin®, and stimulants/methamphetamine.	No editing was done when this pattern occurred. Logically, however, the AFU for the parent drug had to have a value of 3 because respondents with AFUs of 1 or 2 (e.g., HALLAGE of 1 or 2) were not eligible to be donors in the imputations. The final imputed AFU (e.g., IRHALAGE for any hallucinogen) was set to a value of 3.
The AFU for a child drug (e.g., crack cocaine) was younger than AFU for the parent (e.g., any cocaine), This applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack, hallucinogens/LSD/PCP/ Ecstasy, pain relievers/OxyContin®, and stimulants/ methamphetamine.	This edit required the AFUs to be defined (i.e., not missing) for both the parent and child drugs. The younger AFU from the child drug was assigned to the parent AFU. For example, if the AFU of crack cocaine was younger than the AFU of any cocaine, then the edits assigned the value from the crack AFU to the cocaine AFU.
Respondents (Rs) in the stimulants module were lifetime users only of methamphetamine. This included situations in which the only reported use of a stimulant was "some other stimulant," but only methamphetamine was specified. The computer-assisted interviewing (CAI) program skipped these respondents out of questions related to general stimulant use.	Methamphetamine data were moved over to the corresponding general stimulant variables.
Rs in the pain relievers module were lifetime users only of OxyContin®. This included situations in which the only reported use of a pain reliever was "some other pain reliever," but only OxyContin® was specified. The CAI program skipped these respondents out of general questions related to use of any pain relievers.	OxyContin® data were moved over to the corresponding general variables for any pain relievers.

Table B.10 Edits Pertaining to Parent/Child Incidence Variables (continued)

Issue	Edits Implemented
Rs in the hallucinogens module were lifetime users only of LSD, only of PCP, or only of Ecstasy. This included situations in which the only reported use of a hallucinogen was "some other hallucinogen," but only LSD, only PCP, or only Ecstasy was specified. The CAI logic skipped these respondents out of questions related to the age, month, and year that they first used the specific hallucinogen.	Data on initiation of hallucinogen use were moved over to the corresponding specific hallucinogen variables that had been skipped. For example, if the R had used only LSD, the hallucinogen age (from HALLAGE) was assigned to the LSD AFU (LSDAGE). Similarly, if the R had provided data on the month and year that he or she had first used any hallucinogen, these data were assigned to the LSD month and year variables LSDMFU and LSDYFU. Similar edits were done when Rs had been skipped out of questions related to specific hallucinogens but were logically inferred to be lifetime users only of LSD, only of PCP, or only of Ecstasy, based on the "OTHER, Specify" responses.
Rs in the hallucinogens module were lifetime users of two or more of the following: LSD, PCP, or Ecstasy. No other hallucinogen use was indicated. However, the R answered the AFU question for any hallucinogen (edited variable HALLAGE) as "don't know" or "refused." In addition, the specific hallucinogens that the R had used had legitimate values for his or her AFU variables (i.e., LSDAGE, PCPAGE, or ECSAGE, for LSD, PCP, or Ecstasy, respectively).	Logically, the AFU for any hallucinogen (HALLAGE) had to be the minimum of the ages indicated in LSDAGE, PCPAGE, or ECSAGE. Therefore, HALLAGE was assigned the minimum value from LSDAGE, PCPAGE, or ECSAGE. If this edit was done and month and year data were available for a specific hallucinogen, the month and year variables for any hallucinogen (HALMFU and HALYFU) were edited accordingly. For example, if a missing value in HALLAGE was replaced with the value from LSDAGE, and the month of first use (MFU) and year of first use (YFU) for LSD (LSDMFU and LSDYFU) had valid values, the values from LSDMFU and LSDYFU were assigned to HALMFU and HALYFU, respectively. This edit was done *only* to replace missing data in HALLAGE. No editing was done if HALLAGE had a value that was lower than the AFU values reported in LSDAGE, PCPAGE, or ECSAGE.
The Rs in the cocaine/crack, hallucinogens, pain relievers, or stimulants sections indicated first use of any drug in that category (e.g., any hallucinogen) at their current age but had missing data for the age when they first used specific drugs in that category (e.g., LSD).	The Rs were logically inferred to have first used the specific drug at their current age.
The AFU the parent drug was at the R's current age and the AFU for the child drug was at 1 year below the R's current age. Consequently, the child AFU had been assigned to the parent AFU.	The associated MFU and YFU from the child drug (e.g., crack) were assigned to the parent' drug's MFU and YFU data (e.g., for any cocaine). However, this edit was not done in the hallucinogens module when the sole hallucinogen use involved use of LSD, PCP, or Ecstasy. Consequently, some inconsistencies could remain in the edited hallucinogens data.
The Rs in the cocaine/crack, hallucinogens, pain relievers, or stimulants sections indicated first use for the parent drug category (e.g., any hallucinogen) in the current survey year but had missing data for the year when they first used a child drug (e.g., LSD).	The Rs were logically inferred to have first used the child drug in the current year. If the Rs also reported first use of drugs in the parent category (e.g., any hallucinogen) in the same month that they were interviewed and they had missing MFU data for child drugs within that category, the Rs also were logically inferred to have first used the child drug in the same month that they were interviewed. Otherwise, if the Rs originally had reported a month when they first used a child drug but that month had been assigned a code of bad data due to the "flag and impute" edits, the MFU (e.g., LSDMFU) also was reset to the originally reported value if it matched the MFU for any drug in the parent category (e.g., HALMFU).

Table B.10 Edits Pertaining to Parent/Child Incidence Variables (continued)

Issue	Edits Implemented
The Rs in the cocaine/crack, hallucinogens, pain relievers, or stimulants sections indicated first using any drug in that category (e.g., any hallucinogen) in the year immediately prior to the current one (i.e., 2012) but that point of initiation was within 12 months of the interview date. These Rs also had missing data for the year and month when they first used specific drugs in that category (e.g., LSD).	If the Rs originally had reported year and month data for when they first used a specific drug but these data had been assigned codes for bad data due to the "flag and impute" edits, this information was restored *only* if the original answers matched the MFU and YFU for any drug in that category. For example, if LSDMFU and LSDYFU had been set to bad data but the original answers for the LSD month and year matched the month and year in HALMFU and HALYFU, LSDMFU was equated to HALMFU, and LSDYFU was equated to HALYFU.
The MFU and/or YFU questions for the parent drug were answered because the initial answer to the parent AFU was within 1 year of the R's current age. However, the final, edited AFU for the parent drug was more than 1 year younger than the R's current age (e.g., if the unedited AFU for cocaine was within 1 year of the R's current age, but the AFU of crack was more than 1 year younger than the R's current age).	The original answers to the parent MFU and YFU were overwritten with logically assigned legitimate skip codes (see Section 2.4.2). If the R originally had answered the parent AFU as being more than 1 year younger than his or her current age, the CAI program would have skipped the R past the questions about the parent MFU and YFU.
The AFUs for the parent and child drugs were the same, but the MFU or YFU was earlier for the child drug (e.g., crack).	The earlier month and year data from the child drug (e.g., crack) were assigned to the parent drug's month and year data (e.g., cocaine). However, this edit was not done in the hallucinogens module when the sole hallucinogen use involved use of LSD, PCP, or Ecstasy but the YFU and MFU data for the parent and child drugs were inconsistent. Consequently, some inconsistencies could remain in the edited hallucinogens data.
For cigarettes, the AFU for any cigarette use (CIGTRY) equaled the age at initiation of daily cigarette use (CIGAGE). However, the R reported initiating daily cigarette use in a year or month that was earlier than what the R reported for first use of any cigarette.	The month and year data for daily cigarette use (CIGDLMFU and CIGDLYFU, respectively) were assigned to the month and year variables for any cigarette use (CIGMFU and CIGYFU, respectively).
For hallucinogens, the respondent was known never to have used any drugs other than the child drugs, the AFU for the parent drug was nonmissing, only one child AFU was missing, and the minimum of the nonmissing child AFUs was greater than the parent AFU.	The missing child AFU variable was assigned to the parent AFU value.
For hallucinogens, the respondent was known never to have used any drugs other than the child drugs, the parent AFU was nonmissing, only one child AFU was missing, the minimum of the nonmissing child AFUs was equal to the parent AFU, and the earliest of the nonmissing child MFUs and/or YFUs was later than the parent MFU and/or YFU.	The missing child AFU variable was assigned to the parent AFU value.

Table B.10 Edits Pertaining to Parent/Child Incidence Variables (continued)

Issue	Edits Implemented
The parent AFU was equal to the respondent's current age, and some or all of the AFUs for the child drugs were missing.	All missing child AFU variables were assigned to the respondent's current age.
The parent AFU was equal to 1 less than the respondent's current age, the child recency[1] was lifetime but not past year (or, for cigarettes, past 3 years but not past year), and the child AFU value was missing.	The child AFU variable was assigned to 1 less than the respondent's current age.
The AFU for cigarettes was equal to 3 less than the respondent's current age, the cigarette recency was lifetime but not past 3 years, and the AFU for daily cigarettes was missing.	The AFU for daily cigarettes was assigned to 3 less than the respondent's current age.

[1] Because no recency question was associated with daily cigarettes, the overall cigarette recency was used instead.

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module

Issue	Edits Implemented
The respondent (R) reported methamphetamine use in the core stimulants module (i.e., question ST01 or STREF1 was answered as "yes").	The edited recency variables MTHAREC (corresponding to question SD17B) and MTANDLRC (SD18A) were assigned legitimate skip codes. The edited variables MTHAMP (corresponding to question SD17A), MTHANEDL (SD18A), MTHEVCK (from SD17A1 and SD17ALT), MTHNORSN (SD17A2), and MTHNOSP (SD17A2sp) also were assigned legitimate skip codes (see Section 2.4.2).
The R had been routed to the methamphetamine follow-up questions in the special drugs module because the R had not reported methamphetamine use in ST01 or STREF1. However, the R had specified use of methamphetamine as "some other stimulant" in the core stimulants module.	If the lifetime methamphetamine question SD17A had not been answered as "yes," and no other reports of methamphetamine use existed in the special drugs module, then MTHAMP was assigned the code for bad data (see Section 2.4.3).
The R reported in question SD17A that he or she had never used methamphetamine, but the R also specified use of methamphetamine with a needle somewhere in the "OTHER, Specify" items SD05A through SD05E, pertaining to use of other drugs with a needle.	The R was logically inferred to be a methamphetamine user. The edited variables MTHAMP and MTHANEDL were assigned a code of 3 (Yes, LOGICALLY ASSIGNED). Because questions SD17B and SD18B pertaining to the most recent use of methamphetamine and most recent use of methamphetamine with a needle, respectively, had been skipped, the edited variables MTHAREC and MTANDLRC were assigned a code of 9 (Used at some point in the lifetime, LOGICALLY ASSIGNED).
After reporting methamphetamine use in SD17A, the R reported in question SD17A1 or SD17ALT that his or her earlier report from the core stimulants module was correct that he or she had never used methamphetamine. However, the R also specified using methamphetamine with a needle.	The "OTHER, Specify" data overruled the R's denial of methamphetamine use from SD17A1 or SD17ALT. MTHAMP retained a code of 1 (i.e., "yes"), and the data were retained from SD17B for MTHAREC. The variable MTHEVCK was assigned the code for bad data. If SD18A did not indicate use of methamphetamine with a needle, MTHANEDL was assigned a code of 3 and MTANDLRC was assigned a code of 9, as described previously.
After reporting methamphetamine use in SD17A, the R reported in question SD17A1 or SD17ALT that his or her earlier report from the core stimulants module was correct that he or she had never used methamphetamine. The R did not specify any use of methamphetamine with a needle.	The R was logically inferred *not* to be a methamphetamine user. MTHAMP was assigned a code of 4 (No, LOGICALLY ASSIGNED), and a code of 81 (Never used methamphetamine, LOGICALLY ASSIGNED) was assigned to MTHAREC, MTHANEDL, and MTANDLRC. The answer that confirmed that the R had never used methamphetamine was retained in MTHEVCK to document the source of the logical inference of nonuse of methamphetamine. MTHNORSN and MTHNOSP were assigned a code of 99 (LEGITIMATE SKIP).
The R did not know or refused to report in SD17A1 or SD17ALT which answer was correct: the earlier report of nonuse from the core stimulants module or the report of methamphetamine use in the special drugs module.	MTHAMP through MTANDLRC were set to bad data (i.e., including MTHAREC). Data from the core stimulants module indicating lifetime nonuse of methamphetamine were used to determine the core-plus-noncore (CPN) methamphetamine recency.

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module (continued)

Issue	Edits Implemented
The R did not indicate methamphetamine use in the core stimulants module and also reported in SD17A that he or she never used methamphetamine. No other reports of methamphetamine use existed elsewhere in the special drugs module.	MTHAMP retained a code of 2 (i.e., "no"). A code of 91 (NEVER USED METHAMPHETAMINE) was assigned to MTHAREC, MTHANEDL, and MTANDLRC. A code of 99 was assigned to MTHEVCK, MTHNORSN, and MTHNOSP.
Follow-up questions were skipped because the R refused to indicate in SD17A whether he or she ever used methamphetamine.	A code of 97 (i.e., refused) was assigned to the remaining edited methamphetamine variables, and they were treated as having missing data. Thus, a refusal to report methamphetamine use in SD17A was propagated onto MTHAREC, MTHANEDL, and MTANDLRC.
The R reported in SD17A1 or SD17ALT that his or her report of methamphetamine use in question SD17A was correct. The R reported in SD17A2 that there was "some other reason" why he or she had not previously reported methamphetamine use in the core stimulants module. However, the written response in SD17A2SP indicated either that the R did not recognize methamphetamine as a prescription drug, or that the R made a mistake in answering the previous core question about methamphetamine.	MTHNORSN (corresponding to question SD17A2) was edited as follows: • Assigned a code of 11 (Did not consider methamphetamine to be a prescription drug LOGICALLY ASSIGNED) if the R gave an answer in question SD17A2SP indicating that he or she did not recognize methamphetamine in the context of questions about prescription drugs. MTHNOSP (corresponding to SD17A2SP) retained a code of 1 ("Did not think of methamphetamine as a prescription drug") to indicate the source of the logical inference. • Assigned a code of 12 (Made a mistake in answering the previous question LOGICALLY ASSIGNED) if the R gave an answer in SD17A2SP indicating that he or she made a mistake in answering the methamphetamine question in the core stimulants module. MTHNOSP retained a code of 2 (i.e., Made a mistake). Values of 11 and 12 in MTHNORSN were used in editing MTHREC06 (see Section 6.2.6.2).
MTHAMP indicated that the R was a lifetime user of methamphetamine, and there was nothing in the follow-up questions to indicate that this answer was incorrect. However, the R did not know or refused to report when he or she last used methamphetamine.	The methamphetamine recency variable (MTHAREC) was generally assigned a code of 9 (i.e., used at some point in the lifetime). Exceptions were as follows: • If the R reported last using methamphetamine with a needle in the past 30 days, then MTHAREC was assigned a code of 11 (Used in the past 30 days, LOGICALLY ASSIGNED). • Otherwise, if the R reported last using methamphetamine with a needle more than 30 days ago but within the past 12 months, then MTHAREC was assigned a code of 8 (Used at some point in the past 12 months, LOGICALLY ASSIGNED). A code of 8 was assigned because the R could have used methamphetamine in the past 30 days.

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module (continued)

Issue	Edits Implemented
The R indicated more recent use of methamphetamine with a needle in MTANDLRC than the R indicated for most recent use of any methamphetamine in MTHAREC.	MTHAREC was edited to reflect the indication of more recent use from MTADNLRC. • If MTANDLRC indicated that the R last used methamphetamine with a needle in the past 30 days, but MTHAREC did not indicate use in the past 30 days, MTHAREC was assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED). • If MTANDLRC indicated that the R last used methamphetamine with a needle more than 30 days ago but within the last 12 months, and MTHAREC indicated that the R last used methamphetamine more than 12 months ago, MTHAREC was assigned a code of 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED). • If MTANDLRC indicated that the R last used methamphetamine with a needle more than 30 days ago but within the last 12 months, and MTHAREC did not indicate use in the past year, MTHAREC was assigned a code of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED.
MTHAMP (corresponding to question SD17A) indicated that the R was a lifetime user of methamphetamine. There were no indications elsewhere in the special drugs module of the R reporting use of methamphetamine as "some other drug" that he or she used with a needle. The corresponding needle recency variable MTANDLRC was skipped because of the following:	
• The R never used methamphetamine with a needle (from MTHANEDL).	A code of 93 was assigned to the corresponding methamphetamine needle recency variable MTANDLRC to indicate that the R used methamphetamine but never with a needle.
• The R refused to indicate or did not know in MTHANEDL whether he or she had ever used methamphetamine with a needle.	A code of 97 was assigned to MTANDLRC (i.e., the refusal was propagated). Data from MTANDLRC were not relevant to creating the CPN methamphetamine recency.
The R was a lifetime user of methamphetamine or other stimulants in the core stimulants module and reported using methamphetamine or other stimulants with a needle at least once in his or her lifetime. However, the R did not know or refused to indicate when he or she last used the drug with a needle.	The edits depended on the most recent use reported in the corresponding core recency variable: • In general, the edited needle recency variable (e.g., MTNDLREC for methamphetamine, based on question SD10B) was assigned a code of 9 (i.e., used at some point in the lifetime). • However, if the core recency (e.g., METHREC) indicated that the R last used the drug more than 12 months ago, then the edited needle recency variable pertaining to that drug was assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED).

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module (continued)

Issue	Edits Implemented
The R was a lifetime methamphetamine user in the core stimulants module. The R reported in question SD10A that he or she had never used a needle to inject methamphetamine, but the R also specified use of methamphetamine with a needle somewhere in the "OTHER, Specify" items SD05A through SD05E, pertaining to use of other drugs with a needle.	The lifetime methamphetamine needle use variable MTHNEEDL (corresponding to SD10A) was assigned a code of 3 (Yes LOGICALLY ASSIGNED). The methamphetamine needle recency variable MTNDLREC generally was assigned a code of 9 (i.e., used at some point in the lifetime). However, if the R reported last use of any methamphetamine more than 12 months ago (i.e., METHREC = 3), then MTNDLREC was assigned a code of 13 (Used a needle to inject methamphetamine more than 12 months ago, LOGICALLY ASSIGNED).
The R was a lifetime methamphetamine user in the core stimulants module. Questions SD10C and SD11 pertaining to use of other stimulants with a needle had been skipped because methamphetamine was the only stimulant that the R had reported ever using.	If the lifetime methamphetamine variable METHDES was coded as 1 (i.e., "yes") and all other lifetime stimulant variables had values of 2 (i.e., "no"), the edited needle use variables for lifetime use of other stimulants with a needle (OSTNEEDL, corresponding to SD10C) and most recent use of other stimulants with a needle (OSTNLREC, corresponding to SD11) were assigned a code of 99 (legitimate skip). This edit was not done if SD10C and SD11 were skipped when METHDES had the only affirmative response, but at least one of the other lifetime stimulant variables had a value of "don't know" or "refused."
The R was a lifetime methamphetamine user in the core stimulants module. In addition, questions SD10C and SD11 pertaining to use of other stimulants with a needle had been answered because the R reported use of "some other stimulant" in the stimulants module. Based on "OTHER, Specify" data in the stimulants module, however, methamphetamine was the only stimulant that the R had ever used.	Data were retained in the needle recency variable for other stimulants (OSTNLREC) if the R reported more recent use of "other" stimulants with a needle than what the R reported for the methamphetamine needle recency (MTNDLREC); that is, no editing was done. Otherwise, nonblank values in OSTNLREC were replaced with a code of 89 (Legitimate skip LOGICALLY ASSIGNED) (see Section 2.4.2).
The R was a lifetime nonmedical user of stimulants. The R reported in question SD10C that he or she had never used a needle to inject stimulants/other stimulants. However, the R also specified use of stimulants other than methamphetamine with a needle somewhere in the "OTHER, Specify" items SD05A through SD05E, pertaining to use of other drugs with a needle.	The stimulant needle recency variable OSTNLREC generally was assigned a code of 9 (i.e., used at some point in the lifetime). However, if the R reported last nonmedical use of any stimulant more than 12 months ago (i.e., STIMREC = 3), then OSTNLREC was assigned a code of 13 (Used a needle to inject stimulants more than 12 months ago, LOGICALLY ASSIGNED).
The R was asked questions about use of methamphetamine and other stimulants with a needle because the R reported lifetime use of methamphetamine and "some other stimulant" in the stimulants module (and no other stimulant). However, only methamphetamine was specified as the "other" stimulant.	The R was treated as being a lifetime user only of methamphetamine. Therefore, any data in the other stimulant needle use variables OSTNEEDL and OSTNLREC were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned). This edit indicated that the R logically should have skipped the questions pertaining to OSTNEEDL and OSTNLREC.

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module (continued)

Issue	Edits Implemented
The R had been logically inferred to be a nonuser of prescription-type stimulants because the only drugs that the R reported using in the stimulants module were over-the-counter (OTC) drugs. In addition, the R did not specify use of methamphetamine or other stimulants with a needle as "some other drug" that the R used with a needle.	Any data in the methamphetamine and other stimulant needle variables MTHNEEDL, MTNDLREC, OSTNEEDL, and OSTNLREC were overwritten with a code of 81 (i.e., NEVER USED METHAMPHETAMINE/STIMULANTS Logically assigned) for consistency with the inference that the R was a lifetime nonuser of prescription-type stimulants. This edit did not apply to the methamphetamine variables starting with question SD17A that respondents were asked if they reported in the core stimulants module that they never used methamphetamine.
The R reported in the core drug modules that he or she never used methamphetamine (METHREC = 81 or 91) or that he or she had never used stimulants (STIMREC = 81 or 91). The R also had not reported use of methamphetamine in question SD17A. However, the R specified use of methamphetamine or other stimulants as "some other drug" that he or she had ever injected.	No editing was done to the core drug data, including the core recency variables. However, the edits that were described previously for setting the relevant needle recency variable to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) were relevant to creating the CPN recency variables for stimulants and methamphetamine.
The R was logically inferred to be a lifetime user of methamphetamine with a needle in the MTHNEEDL variable that existed prior to 2005 (i.e., MTHNEEDL = 3) based on the R's "OTHER Specify" data in the variables OTDGNDLA through OTDGNDLE.	If the lifetime other stimulant needle use variable OSTNEEDL had missing data regarding use of a needle to inject other stimulants, whatever unedited data existed in questions SD10C and SD11 were reassigned to the corresponding edited variables OSTNEEDL and OSTNLREC, respectively.
The R was logically inferred to be a lifetime user of methamphetamine (METHDES = 3) based on "OTHER, Specify" data in the stimulants module, and the methamphetamine needle variable MTHNEEDL had missing data. Further, the variable MTHANEDL did not indicate use of methamphetamine with a needle.	If OSTNEEDL = 2 and MTHANEDL did not equal 1 (i.e., "yes"), the R was logically inferred not to have used a needle to inject methamphetamine (i.e., MTHNEEDL = 4, where 4 = No LOGICALLY ASSIGNED). If the methamphetamine needle recency variable MTNDLREC had been skipped, it was assigned a code of 93 (USED METHAMPHETAMINE BUT NEVER WITH A NEEDLE). When the R had not indicated lifetime methamphetamine use in questions ST01 or STREF1 in the core stimulants module, question SD10C (corresponding to OSTNEEDL) asked whether the R had used a needle to inject *any* stimulant. Therefore, it could be logically inferred that the R had never used a needle to inject methamphetamine. Consequently, MTNLDREC would not indicate more recent use of methamphetamine with a needle than was indicated by METHREC.

Table B.12   Miscellaneous Core Edits Pertaining to Tobacco Variables

Issue	Edits Implemented
The R reported smoking cigarettes on all 30 days in the past 30 days, but the R reported initiating daily cigarette use within 30 days of the interview date. The latter could occur in one of the following ways: (a) the reported month and year for when the R started smoking cigarettes daily was within 30 days of the interview date; or (b) the R reported starting to smoke cigarettes every day at his or her current age (i.e., CIGAGE = AGE), but the R had been at this age for fewer than 30 days.	The following edits were implemented: • If only the month and year indicated that the R had been smoking daily for fewer than 30 days (i.e., but the R had been at the age given in CIGAGE for at least 30 days), the month and year (CIGDLMFU and CIGDLYFU) were set to bad data. • If the R had been at the age reported in CIGAGE for fewer than 30 days and reported starting to smoke daily in the same month that he or she was interviewed, CIGAGE, CIGDLMFU, and CIGDLYFU all were set to bad data. • If the R had been at the age reported in CIGAGE for fewer than 30 days and CIGDLMFU and CIGDLYFU had codes of DK or REF, CIGAGE, CIGDLMFU, and CIGDLYFU all were set to bad data.
The R reported smoking cigarettes in the past 30 days. The R did not report smoking on all 30 days in the past 30 days but reported initiating daily cigarette use within 30 days of the interview date.	No editing was done when this pattern occurred, even though question CG15 asked respondents if they ever smoked cigarettes every day for at least 30 days. Consequently, data were preserved that indicated that the R recently initiated daily cigarette use, even if the R could not have smoked daily for at least 30 days.
The CAI program had skipped lifetime cigarette questions pertaining to periods where the R had smoked cigarettes daily for at least 30 days (CG15) or whether the R had smoked 100 or more cigarettes (CG16A). Answers to prior questions already indicated that the R had engaged in these behaviors. For example, if the R had reported smoking every day in the past 30 days, the R had to have had a point in his or her life when he or she smoked cigarettes every day for at least a month. Similarly, it could be inferred from answers to questions about the number of days that Rs smoked in the past 30 days and the number of cigarettes they usually smoked per day that they had smoked at least 100 cigarettes in their lifetime.	It was logically inferred in CIGDLYMO (the edited version of CG15) or in CIG100LF (the edited version of CG16A) that the R had engaged in the behavior of interest (i.e., ever smoked cigarettes daily for at least 30 days or smoked 100 or more cigarettes in the R's lifetime). A code of 5 was assigned. This code had the following meaning: 5 = Yes LOGICALLY ASSIGNED (from skip pattern) This code was used instead of a legitimate skip code to signify that the CAI logic had skipped the R out of the question because the R had already provided data to indicate that the question should have been answered affirmatively.
The R was routed to the lifetime 100-cigarette question (CG16A) because the R estimated the number of days that he or she smoked cigarettes in the past 30 days. The R did not answer CG16A as "yes," but the minimum number of days that the R could have smoked in the past 30 days and the minimum usual number of cigarettes smoked per day suggest that the R smoked 100 or more cigarettes in his or her lifetime.	It was logically inferred in CIG100LF (i.e., the edited version of CG16A) that the R had smoked 100 or more cigarettes in his or her lifetime. For example, if an R reported smoking 10 to 19 days and reported smoking 16 to 25 cigarettes per day, then the R would have smoked at least 100 cigarettes in the past 30 days, even if he or she smoked on only 10 days and smoked only 16 cigarettes per day.

B.4 Detailed Documentation of Edits for Variables in Noncore Self-Administered Modules

This section presents tables that detail the edits that were applied to variables in the noncore self-administered modules. Because the content of noncore self-administered modules differs considerably from module to module, many of the edits are module-specific.

- Tables B.13 and B.14 present edits pertaining to variables in the special drugs module (except for issues that were presented in Table B.12 for editing of stimulant and methamphetamine variables from this module), corresponding to the discussion of procedures in Section 7.4.1.
- Tables B.15 and B.16 present edits pertaining to variables in the blunts module, corresponding to the discussion of procedures in Section 7.4.3.
- Table B.17 presents edits pertaining to variables in the special topics module, corresponding to the discussion of procedures in Section 7.4.5.
- Tables B.18 to B.21 present edits pertaining to variables in the prior substance use module, corresponding to the discussion of procedures in Section 7.4.7.
- Tables B.22 to B.24 present edits pertaining to variables in the substance treatment module, corresponding to the discussion of procedures in Section 7.4.8.
- Table B.25 presents edits pertaining to editing of variables in the adult mental health service utilization module, corresponding to the discussion of procedures in Section 7.4.10.
- Finally in this section, Tables B.26 and B.27 present edits pertaining to variables in the consumption of alcohol module, corresponding to the discussion of procedures in Section 7.4.17.

Table B.13 Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Module

Response Pattern	Edit
Variables were skipped because the respondent (R) never used the drug of interest, and there were no other indications elsewhere in the special drugs module that the respondent ever used this drug.	A code of 91 was assigned to the edited variables. For example, if the R never used heroin, the edited variables HERSMOKE, HRSMKREC, HERSNIFF, HRSNFREC, HERNEEDL, and HRNDLREC were assigned a code of 91.
Variables were skipped because the R refused to indicate in the corresponding core module whether he or she ever used the drug of interest, and there were no other indications elsewhere in the special drugs module that the respondent ever used this drug.	A code of 97 (i.e., refused) was assigned to the edited variables. Thus, for example, a refusal from the heroin recency-of-use variable in the core was propagated onto the heroin variables in the special drugs module.
Variables were skipped because the R did not know in the core module whether he or she ever used the drug of interest, and there were no other indications elsewhere in the special drugs module that the R ever used this drug.	The skipped special drugs variables pertaining to this drug retained a value of 98 (i.e., blank).
For the methamphetamine variables that existed prior to 2005, other stimulants, and cocaine, there were no indications elsewhere in the special drugs module of the R reporting use of this drug as "some other drug" that he or she used with a needle. The R was a lifetime user of the drug of interest, but the corresponding needle recency variable was skipped because of the following:	
• The R never used the drug with a needle.	A code of 93 was assigned to the corresponding needle recency variable (e.g., CONDLREC) to indicate that the R used the drug but never with a needle.
• The R refused to indicate whether he or she had ever used the drug with a needle.	A code of 97 was assigned to the edited needle recency variable (i.e., the refusal was propagated).
• The R did not know whether he or she had ever used that drug with a needle.	The edited needle recency variable retained a code of 98 (i.e., blank).

Table B.13 Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Module (continued)

Response Pattern	Edit
The R was a lifetime user of heroin, but relevant recency variables for smoking heroin (HRSMKREC), sniffing heroin (HRSNFREC), or using it with a needle (HRNDLREC) were skipped because of the following:	
• The R never used heroin via the route of interest.	A code of 93 was assigned to the relevant heroin recency variable(s) (e.g., HRSMKREC for smoking heroin) to indicate that the R used heroin but not in that particular way.
• The R refused to indicate whether he or she had ever used heroin via the route of interest.	A code of 97 was assigned to the relevant heroin recency variable(s) (i.e., the refusal was propagated).
• The R did not know whether he or she had ever used heroin via the route of interest.	The edited heroin recency variable(s) retained a code of 98 (i.e., blank).
(In the case of heroin use with a needle, there were no other indications elsewhere in the special drugs module of heroin use with a needle.)	The special situation in which Rs reported lifetime use of heroin in the core but reported that they never smoked, sniffed, or used it with a needle is discussed in the text.
General needle use variables were skipped because the R reported never using heroin, methamphetamine, other stimulants, cocaine, or any other drug (question SD05; edited variable OTDGNEDL) with a needle. For methamphetamine, this included reports in the methamphetamine question SD18a (added in 2005) and in the methamphetamine question that existed prior to 2005 that the R had never used methamphetamine with a needle.	A code of 99 (i.e., legitimate skip) was assigned to all of the general needle use variables that had been skipped (GNNDREUS, GNNDLSH1, GNNDCLEN, GNNDLSH2, and GNNDGET).
General needle use variables were skipped because question SD05 was answered as "no" (OTDGNEDL = 2); there were no affirmative reports of heroin, methamphetamine, other stimulants, or cocaine with a needle, but one or more of the lifetime needle use variables for these drugs was answered as "don't know" or "refused."	When there was no affirmative report of use of heroin, methamphetamine, other stimulants, or cocaine with a needle, question SD05 was worded as follows: "Have you ever, even once, used a needle to inject *any drug* that was not prescribed for you ..." (wording not italicized in the interview). Therefore, a code of 99 (i.e., legitimate skip) was assigned to all of the general needle use variables that had been skipped (GNNDREUS, GNNDLSH1, GNNDCLEN, GNNDLSH2, and GNNDGET) because it could be inferred that the response of "no" in question SD05 pertained to use of any drug with a needle. However, no editing was done to any responses of "don't know" or "refused" in the lifetime needle use variables pertaining to heroin (HERNEEDL), methamphetamine (MTHNEEDL), other stimulants (OSTNEEDL), or cocaine (COCNEEDL).

Table B.13 Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Module (continued)

Response Pattern	Edit
General needle use variables were skipped because question SD05 was refused (OTDGNEDL = 97), and there were no affirmative reports of heroin, methamphetamine, other stimulants, or cocaine with a needle.	The refusal from OTDGNEDL was propagated to the general needle use variables GNNDREUS, GNNDLSH1, GNNDCLEN, GNNDLSH2, and GNNDGET.

Table B.14 Edits in the Special Drugs Module (Other than Skip Patterns) Pertaining to Heroin Use and Needle Use

Issue	Edits Implemented
The respondent (R) was a lifetime user of heroin and reported smoking, sniffing, or using heroin with a needle at least once in his or her lifetime, but did not know or refused to indicate when he or she last smoked, sniffed, or injected heroin.	The edits depended on the most recent use of heroin reported in the corresponding core heroin recency variable: • In general, the edited heroin recency variables in the special drugs module (HRSMKREC, HRSNFREC, HRNDLREC) were assigned a code of 9 (i.e., used at some point in the lifetime). • However, if the core heroin recency indicated that the R last used heroin more than 12 months ago and there was no other indication in the special drugs module that the R had used heroin in the past 12 months (see below), then the edited variables pertaining to smoking, sniffing, or injection of heroin were assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED). This edit did not apply if the R answered "did not know" or refused to report when he or she last used heroin in a particular way (e.g., smoking it) but reported last using it a different way in the past 12 months (e.g., with a needle).
The R was a lifetime user of cocaine and reported using it with a needle at least once in his or her lifetime. However, the R did not know or refused to indicate when he or she last used cocaine with a needle.	The edits depended on the most recent use reported in the corresponding core recency variable COCREC: • In general, the edited needle recency variable CONDLREC for cocaine was assigned a code of 9 (i.e., used at some point in the lifetime). • However, if COCREC indicated that the R last used cocaine more than 12 months ago, then CONDLREC was assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED).
The R reported in the core drug modules that he or she never used heroin (HERREC = 91) or had never used cocaine (COCREC = 91). However, the R specified use of one or more of these drugs as "some other drug" that he or she had ever injected.	No editing was done to the core drug data. However, the R was logically inferred in the special drugs data to be a lifetime user of that drug with a needle, even though the core drug data indicated that the R never used that drug. The corresponding needle recency variable was set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). For example, if the R reported in the heroin module that he or she never used heroin but specified injection of heroin as "some other drug," the lifetime heroin needle use variable HERNEEDL was set to 3 (Yes LOGICALLY ASSIGNED), and the heroin needle recency HRNDLREC was set to 9. Similar edits were done for the needle use variables pertaining to methamphetamine, other stimulants, and cocaine.

Table B.14 Edits in the Special Drugs Module (Other than Skip Patterns) Pertaining to Heroin Use and Needle Use (continued)

Issue	Edits Implemented
The R specified lifetime use of more than five drugs with a needle.	Priority was given to retaining as many unique mentions as possible for other drugs that the R used with a needle. Thus, multiple mentions of the same drug would be counted only once. Priority also would be given to retaining mentions of drugs that were covered in the special drugs module that the R had not previously reported using with a needle (e.g., if the question corresponding to MTHNEEDL had been answered as "no" but methamphetamine had been specified as "some other drug" that the R used with a needle). Conversely, retention of "OTHER, Specify" codes corresponding to drugs that the R had already reported using with a needle were given lower priority. If there were still more than five mentions of unique drugs after the above steps, priority was given to retaining the most serious drugs according to the Drug Enforcement Administration (DEA) drug schedule (e.g., first priority to retaining mention of Schedule I drugs that have no approved medical use in the United States, second priority to retaining Schedule II drugs). Finally, after the drugs had been ranked according to their severity based on the DEA drug schedule, if mention of more than five drugs still remained, the codes were retained in the order they appeared in the data.
The R reported using a needle to inject a drug for nonmedical reasons (SD05 = 1), but the R previously reported never using marijuana, cocaine, heroin, hallucinogens, inhalants, prescription pain relievers, prescription tranquilizers, prescription stimulants, or prescription sedatives.	No editing was done if the R • specified needle use involving a drug that could be abused or had psychoactive properties (e.g., steroids, one or more categories of drugs covered in the core modules that were not covered elsewhere in special drugs, such as prescription pain relievers), or • reported one or more "risky" needle use behaviors (reusing a needle, needle sharing, or cleaning a needle with bleach). The R was inferred not to be a lifetime nonmedical needle user (OTDGNEDL = 4) if the R • specified use of a drug that was typically not abused and did not have psychoactive properties (e.g., if injection of antibiotics was specified), and • reported never reusing a needle, sharing a needle (before or after someone had used it), or cleaning a needle with bleach (i.e., "risky" needle use behaviors). When OTDGNEDL = 4, any data in the general needle use variables GNNDREUS, GNNDLSH1, GNNDCLEN, GNNDLSH2, and GNNDGET were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned).
The general needle use items corresponding to questions SD12 through SD16 had been answered. However, the only indications of lifetime needle use had been set to bad data (e.g., if MTHANEDL had been set to bad data; see Table B.11).	The edited general needle use variables GNNDREUS through GNNDGET were set to bad data.

Table B.14 Edits in the Special Drugs Module (Other than Skip Patterns) Pertaining to Heroin Use and Needle Use (continued)

Issue	Edits Implemented
The R reported getting his or her last needle "some other way" and specified a meaningful response for how he or she last got the needle.	The final, edited variable pertaining to how the R got his or her last needle (GNNDGET) was a composite of the response categories that were offered to the R (i.e., bought the needle from a pharmacy, got the needle from a needle exchange, bought the needle on the street, got the needle in a shooting gallery, got the needle some other way). This was done because the computer-assisted interviewing (CAI) logic did not allow Rs to specify an "other" way that they got the needle if they reported getting the needle in one of the first four ways. If Rs reported getting the needle "some other way" and specified a meaningful way they got the needle, then the "OTHER, Specify" response was assigned to GNNDGET.
The R reported getting his or her last needle "some other way" and did not know what that other way was, refused to specify what that other way was, or gave a response that was coded as bad data (e.g., a nonsensical response).	The final, edited variable pertaining to how the R got his or her last needle (GNNDGET) retained a nonspecific code of "some other way." Stated another way, the response of "some other way" was given precedence over the missing value in the "OTHER, Specify" response. The edit was done in this manner to provide a nonmissing value for analysts to use.
The R answered "don't know" or "refused" at the outset, when asked how he or she got the last needle that he or she used.	The response of "don't know" or "refused" was retained in the final, edited variable (GNNDGET).

Table B.15 Edits in the Blunts Module Pertaining to Consistency Checks for Incidence

Issue	Edits Implemented
The respondent (R) indicated in the final verification check (BLCC06) that the age at first use (AFU) based on the new month and year of first use (MFU and YFU) was correct (i.e., BLCC06 = 4). The computer-assisted interviewing (CAI) program updated the AFU (AGE1STBL) with the value of the age calculated from the MFU and YFU (i.e., MYR1STBL). However, the new value for the AFU indicated that the R was more than 1 year younger than his or her current age at the time the R first used blunts (e.g., the R was 16, reported first use of blunts at age 15, but then confirmed an MFU and YFU that meant the R was 14 when the R first used blunts). Had the R initially reported this AFU, the R would not have been routed to the MFU and YFU questions.	The updated value was retained in the edited AFU variable BLNTAGE (e.g., first use at age 14 for a 16-year-old R in this example). Based on this updated AFU, it was logically inferred that the R should have skipped the MFU and YFU items. A code of 9989 was assigned to the YFU variable BLNTYFU, and a code of 89 was assigned to the MFU variable BLNTMFU.
The final verification check (BLCC06) was skipped because the R entered revised data for the MFU and YFU that made BLNTMFU and BLNTYFU consistent with the AFU in BLNTAGE.	No editing was done because the R was considered to have resolved the inconsistency.
The final verification check (BLCC06) was skipped because the R entered a new MFU that was the same as the R's birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, BLNTMFU and BLNTYFU were set to bad data if they could never be consistent with the AFU. Suppose, for example, that a hypothetical R was born in June 1996, was interviewed in March 2014 (age 17 at the time of the interview), reported first use of blunts at age 17, and initially reported first use in May 2013. First use in May 2013 would have meant that the R was 16 when he or she first used blunts because the R's 17th birthday was not until June 2013. If the R changed the month and year to June 2013, that could be consistent with first use at age 17, if the use occurred after the R's birthday. However, if the R changed the month and year to June 2012, it would never be possible for the R to have first used at age 17 and also to have first used in June 2012. In this latter situation, BLNTMFU and BLNTYFU would be set to bad data.
The R entered a new MFU or YFU that differed from what the R previously reported. The age based on the revised MFU and YFU (updated in MYR1STBL) still mismatched the AFU, but the R indicated in the final verification check that the new value from MYR1STBL was correct.	No editing was done in this situation. The CAI program automatically updated AGE1STBL to be consistent with the updated values reported for the MFU and YFU.

Table B.15 Edits in the Blunts Module Pertaining to Consistency Checks for Incidence (continued)

Issue	Edits Implemented
The consistency check was triggered between AGE1STBL and MYR1STBL. However, the R answered the first consistency check (BLCC03, regarding whether the value in MYR1STBL was correct) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The AFU from AGE1STBL was retained in BLNTAGE, but BLNTMFU and BLNTYFU were set to bad data.
The consistency check was triggered between AGE1STBL and MYR1STBL. However, the R entered the same values for the YFU and MFU that triggered the inconsistency with the AFU in the first place.	No editing was done to BLNTAGE, but BLNTYFU and BLNTMFU were set to bad data.
The R entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for MYR1STBL. However, the R failed to resolve the inconsistency between AGE1STBL and the updated value in MYR1STBL. The R also reported either in the first verification check that MYR1STBL was not correct (i.e., BLCC03 = 6) or reported in the second verification check that the AFU from AGE1STBL was correct (i.e., BLCC04 = 2).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: • The default edit was to set BLNTMFU and BLNTYFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," BLNTMFU and BLNTYFU were assigned the code that corresponded to the answer in the final verification check.
The R entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for MYR1STBL. However, the R failed to resolve the inconsistency between AGE1STBL and the updated value in MYR1STBL. The R also reported in the second verification check that neither answer was correct for what was originally captured in AGE1STBL and MYR1STBL (i.e., BLCC04 = 3).	The following edits were implemented for BLNTAGE, BLNTMFU, and BLNTYFU: • The default edit was to set BLNTAGE, BLNTMFU, and BLNTYFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," BLNTAGE, BLNTMFU, and BLNTYFU were assigned the code that corresponded to the answer in the final verification check.
The R triggered the initial consistency check between the AFU, MFU, and YFU and reported that the age from MYR1STBL was correct (i.e., BLCC03 = 4). However, the R answered the second consistency check (BLCC04) as "don't know" or "refused." Consequently, the R did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for "don't know" or "refused" were assigned to BLNTAGE, BLNTMFU, and BLNTYFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R's correct age when he or she first used blunts, or whether the MFU and YFU indicated the R's correct age at initiation. Therefore, BLNTAGE, BLNTMFU, and BLNTYFU all were set to missing values.

Table B.16 Edits in the Blunts Module Other than Those Pertaining to Consistency Checks for Incidence

Issue	Edits Implemented
The respondent (R) reported lifetime use of blunts but did not know or refused to report in BL02 when he or she last used them.	The blunt recency variable BLNTREC (corresponding to question BL02) was set to 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). The 30-day frequency BLNT30DY was left as blank.
The R reported using blunts (BLNTEVER = 1, corresponding to question BL01), but the R previously reported never using marijuana (MJEVER = 2). The consistency check BLNT30C1 (corresponding to question BL03) was triggered because the R reported using blunts in the past 30 days in question BL02 and the R never used marijuana. BLNT30C1 was not triggered for less recent reports of blunt use.	• If BLNT30C1 = 1 (i.e., R has verified that he or she has never used marijuana): – BLNTEVER was set to 4 (No LOGICALLY ASSIGNED). – BLNTREC was set to 81 (NEVER USED BLUNTS Logically assigned). – BLNT30DY was to 91 (if blank) or 81 (if a value needed to be overwritten). – Incidence variables (BLNTAGE, BLNTYFU, BLNTMFU) were assigned codes of 981, 9981, and 81, respectively. – Other assignment of legitimate skip codes applies, as described in the main text. • If BLNT30C1 = 2 (i.e., R has verified past month use of blunts), no editing was done to BLNTEVER or BLNTREC. Also, these noncore data in blunts were not used to edit the core MJEVER variable (see Section 2.4.1).
The R reported using blunts, but the R previously reported never using marijuana. BLNT30C1 was not triggered because the R reported use of blunts that was less recent than the past 30 days.	• BLNTEVER was set to a value of 11 (i.e., "bump" by 10) to signify that the R reported lifetime use of cigars with marijuana in them but reported never using marijuana in the core, and the R was not asked to verify which answer was correct. • No editing was done to the recency values in BLNTREC.
The R reported using blunts in the past 30 days, and the R is a marijuana user. However, the R reported last using marijuana more than 30 days ago but within the past 12 months or more than 12 months ago. The consistency check BLNT30C2 (corresponding to question BL04) was triggered for these cases.	• If BLNT30C2 = 1 (i.e., R has verified that the core marijuana recency is correct): – BLNTREC was set to 19 (More than 30 days ago LOGICALLY ASSIGNED). – BLNT30DY was set to 93 if blank and to 83 (DID NOT USE BLUNTS IN THE PAST 30 DAYS Logically assigned) if the 30-day frequency data needed to be overwritten. – Note, however, that no editing was done to BLNTREC when BLNT30C2 = 1, but the core marijuana recency MJREC had been set to a value of 8 or 9 (Used at some point in the past 12 months or Used at some point in the lifetime, respectively), based on edits in the marijuana module (see Section 6.2.2.3 in Chapter 6). The rationale for doing no editing to BLNTREC was that values of 8 or 9 in MJREC indicated potential use of marijuana in the past month. • If BLNT30C2 = 2 (i.e., R has verified past month use of blunts), no editing was done to BLNTEVER or BLNTREC. Also, these noncore data in blunts were not used to edit the core marijuana recency MJREC (see Section 2.4.1).

Table B.16 Edits in the Blunts Module Other than Those Pertaining to Consistency Checks for Incidence (continued)

Issue	Edits Implemented
The R reported last using blunts more than 30 days ago but within the past 12 months (BLNTREC = 2), and the R reported last using marijuana more than 12 months ago (MJREC = 3).	BLNT30C2 (BL04) was not triggered in this situation because the R did not report use of blunts in the past 30 days. • BLNTREC was set to a value of 12 (i.e., "bump" by 10), to signify that the R reported last using blunts more than 30 days ago but within the past 12 months but reported last using marijuana more than 12 months ago, and the R was not asked to verify which answer was correct. • The editing procedures continued to set the 30-day frequency BLNT30DY to 93 because the R was not a past month user of blunts.
The R reported last using blunts more than 30 days ago but within the past 12 months (BLNTREC = 2), but the marijuana recency MJREC was assigned codes of 8 or 9 (Used at some point in the past 12 months or lifetime, respectively).	No editing was done to the blunts variables when MJREC had values of 8 or 9 because BLNTREC = 2 was not necessarily inconsistent with when the R last used marijuana.
Question BL07 (edited variable BLNTNOMJ) was skipped. The R reported in the tobacco module that he or she never smoked a cigar and reported in the blunts module that he or she never smoked a blunt.	BLNTNOMJ was set to a value of 91 (NEVER USED CIGARS/BLUNTS). This included situations in which BLNTEVER had been set to 4 (No LOGICALLY ASSIGNED) and BLNTREC to 81 (NEVER USED BLUNTS Logically assigned) because the R confirmed in BLNT30C2 that he or she never used marijuana.
BLNTNOMJ was skipped, and the R reported lifetime use of blunts but definitely did not smoke blunts in the past 30 days. The R also reported in the tobacco module that he or she did not smoke a cigar in the past 30 days (including situations in which the R reported never smoking a cigar).	BLNTNOMJ was set to value of 93 (DID NOT USE CIGARS/BLUNTS IN THE PAST 30 DAYS). This edit also was implemented if the R reported use of blunts less recently than the past 30 days but reported in the core tobacco module that he or she never smoked a cigar. Even though the cigar and blunts data were inconsistent, BLNTNOMJ still was not applicable.
BLNTNOMJ was skipped, and the R reported lifetime use of blunts but definitely did not smoke blunts in the past 30 days. Unlike the situation above, the R reported in the tobacco module that he or she smoked cigars in the past 30 days.	BLNTNOMJ was set to a value of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern). For instance, if the R has reported last smoking cigars with marijuana in them more than 30 days ago, it could logically be inferred that all past month cigar use had to involve use of cigars that did not have marijuana in them.
BLNTNOMJ was skipped, even though the R reported use of blunts in the past 30 days (BLNTREC = 1). However, the R had reported lifetime use of cigars in the tobacco module but reported not smoking cigars in the past 30 days.	Rather than assign a code of 99 (LEGITIMATE SKIP), BLNTNOMJ was set to a value of 14, where 14 = USED BLUNTS PAST 30 DAYS/DIDN'T USE CIGARS PAST 30 DAYS.
BLNTNOMJ was skipped, even though the R reported use of blunts in the past 30 days (BLNTREC = 1). However, the R had reported in the tobacco module that he or she had never smoked part or all of a cigar.	Rather than assign a code of 99, BLNTNOMJ was set to a value of 24, where 24 = USED BLUNTS PAST 30 DAYS/NEVER USED CIGARS.

Table B.16 Edits in the Blunts Module Other than Those Pertaining to Consistency Checks for Incidence (continued)

Issue	Edits Implemented
BLNTNOMJ was skipped because the cigar recency variable CIGARREC or the blunts recency variable BLNTREC had missing values or "indefinite" values (i.e., used at some point in the past 12 months, or used at some point in the lifetime).	BLNTNOMJ was left as 98 (blank).
Question BL07 (corresponding to BLNTNOMJ) was answered but was not answered as "yes," and the R had not smoked blunts in the past 30 days.	BLNTNOMJ was set to a value of 3, where 3 = Yes LOGICALLY ASSIGNED. This edit affected cases where BL07 had been answered (but not as "yes"), and the Rs confirmed in BLNT30C2 that they did not use marijuana in the past 30 days.
The age at first use (AFU) was inconsistent with the R's current age despite the R being prompted to change the AFU.	The final age was accepted as the standard, and BLNTAGE was set to bad data.
BLNTAGE had a value of "don't know" (DK) or "refused" (REF), including situations where this assignment has been made from the consistency check data.	If the questions pertaining to BLNTYFU and BLNTMFU were skipped because the R answered the AFU as DK or REF, the DK or REF value from BLNTAGE was propagated onto BLNTYFU and BLNTMFU. This edit was designed to indicate the reason that BLNTYFU and BLNTMFU had been skipped. In addition, because the R may have first used blunts within 1 year of his or her current age, the month- and year-of-first-use (MFU and YFU) questions may have been relevant to the R. If the R had answered the MFU and YFU questions but BLNTAGE had a final code for DK or REF (i.e., due to a consistency check response), data in BLNTYFU and BLNTMFU were overwritten with the corresponding DK or REF value from BLNTAGE. Retaining the MFU and YFU data in this situation would imply that the R first smoked blunts within 1 year of his or her current age.
The MFU has been skipped because the R answered the YFU as DK or REF.	The DK or REF value from BLNTYFU was propagated onto BLNTMFU. That is, if the R did not know in what year he or she first used blunts, it was assumed that the R would not know the month either. Similarly, a refusal to answer the YFU was interpreted to be a blanket refusal to answer the month as well as the year.
The R had missing data for the MFU. However, one of the following occurred: • The YFU was the current calendar year, and the R was interviewed in January. • The R first used at his or her current age, the R first used in the current calendar year, and the R's most recent birthday occurred in the interview month. • The R first used at his or her current age, the R first used in the prior calendar year, and the R's most recent birthday occurred in December.	The MFU could be logically inferred, as indicated below. • The MFU was logically inferred to be January of the interview year. That would be the only month in which the R could have initiated use of blunts in the current year. • The MFU was logically inferred to be the interview month. If the R first used in the current calendar year and attained his or her current age in the interview month, the R logically could not have initiated use of blunts in any month other than the interview month. • The MFU was logically inferred to be December. If the R first used blunts in the prior calendar year and attained his or her current age in December of that year, the R logically could not have initiated use in any month other than December of the prior calendar year.

Table B.17 Edits Pertaining to the Special Topics Module

Issue	Edits Implemented
The respondent (R) reported never having been arrested or answered the lifetime arrest question as "don't know" or "refused" but reported being on probation or parole in the past 12 months.	The R was logically inferred to have been arrested at least once in his or her lifetime (i.e., BOOKED = 3). The rationale for this edit was that someone could not be on probation or parole without first having been arrested for a crime. The skipped variables pertaining to arrests in the past 12 months retained a value of blank.
The R reported being arrested in the past 12 months, did not report being arrested for a specific crime in that period, but reported being arrested for this crime as "some other offense."	The R was logically inferred to have been arrested for that crime. No further editing was done to the affirmative answer where the R reported being arrested for "some other offense" (BKOTH). Similarly, no further editing was done to the "OTHER, Specify" variable (BKOTHOFF) that indicated the crime for which the R was arrested (see Section 2.4).
The R reported being arrested at least once in the past 12 months and answered all specific past year arrest questions as "no." However, the R specified an offense in the series of follow-up questions SP03R, SP03RSP, SP03S, and SP03SSP.	The variable for "some other offense" (BKOTH) was logically inferred to be "yes." A code of 3 (i.e., Yes LOGICALLY ASSIGNED) was assigned to BKOTH. The offense that was specified in SP03RSP or SP03SSP was assigned to the edited "OTHER, Specify" variable BKOTHOFF.
The R reported being arrested at least once in the past 12 months but answered all specific past year arrest questions as "no" and reported nothing in the "OTHER, Specify" variable BKOTHOFF to support the indication of being arrested.	The response was retained to indicate that the R had been arrested in the past 12 months. A code of 5 (i.e., Offense unknown LOGICALLY ASSIGNED) was assigned to the "some other offense" variable (BKOTH).
The R reported being arrested for every offense in the past 12 months that was asked about in the module. (For youths aged 12 to 17, that included reports of being arrested for possession of tobacco; this question was skipped for adults.)	The edits differed, depending on what Rs specified for their "other" offense: • If a valid "other" offense was not specified, the entire series of past year offense variables was assigned a bad data code. • If the R gave a valid response for some other offense for which he or she was arrested in the past 12 months, the data were retained to indicate that the R was arrested for this other offense. However, the variables pertaining to arrests for all other offenses were set to bad data. • For adults, the variable pertaining to arrests for possession of tobacco (BKPOSTOB) continued to be assigned a legitimate skip code.
The R reported being arrested only one time in the past 12 months, did not report being arrested for some other offense (BKOTH = 2), but reported being arrested for every other offense in that same period.	Not including BKOTH or its associated "OTHER, Specify" variable (BKOTHOFF), the variables pertaining to arrests for specific offenses in the past 12 months were assigned a bad data code. For adults, the BKPOSTOB variable continued to be assigned a legitimate skip code.
The R reported being arrested 80 or more times in the past 12 months.	The variable pertaining to the number of arrests in the past 12 months (NOBOOKYR) was set to bad data.
The R had alternating "yes/no" or "no/yes" patterns to all questions about arrests for specific offenses in the past 12 months (e.g., SP03a = 1, SP03b = 2, SP03c = 1).	All variables pertaining to arrests for specific offenses in the past 12 months were set to bad data.

Table B.17 Edits Pertaining to the Special Topics Module (continued)

Issue	Edits Implemented
The R was asked questions about driving under the influence of alcohol or illegal drugs solely because the R originally reported past year use of one or more psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, or sedatives). However, the R was logically inferred to be a lifetime nonuser of these psychotherapeutics because the only reported lifetime use involved over-the-counter (OTC) drugs.	Any data in the substance use and driving variables (DRVALDR, DRVAONLY, and DRVDONLY) were replaced with a code of 81 (i.e., NEVER USED ALCOHOL OR DRUGS Logically assigned).
The R was asked questions about driving under the influence of alcohol, but the alcohol recency variable ALCREC had been set to bad data.	The edited variables pertaining to driving under the influence of alcohol and illegal drugs in combination (DRVALDR) and driving under the influence of alcohol (DRVAONLY) were set to bad data.
The R was routed into questions about driving under the influence of alcohol and illegal drugs in combination and about driving under the influence of illegal drugs, but (a) the only drug that the R definitely used in the past 12 months was alcohol (i.e., after all editing had been done to the core recency-of-use variables for alcohol and other drugs), and (b) it could not be determined that the R was not a past year user of all of the other drugs.	The edited variables pertaining to driving under the influence of alcohol and illegal drugs in combination (DRVALDR) and driving under the influence of illegal drugs (DRVDONLY) were set to bad data.
The R had not used alcohol in the past 12 months and was routed into the question about driving under the influence of illegal drugs solely because of psychotherapeutic use that turned out to be limited to OTC use. In addition, one or more other drug recency-of-use variables were ambiguous with respect to past year use, so it could not be determined whether the R did or did not use other illegal drugs.	The edited variable (DRVDONLY) was set to bad data.
All core drug recency variables that had triggered respondents being asked questions about driving under the influence of drugs in the past 12 months had been set to bad data.	The edited variables pertaining to driving under the influence of alcohol and illegal drugs in combination (DRVALDR) and driving under the influence of illegal drugs (DRVDONLY) were set to bad data.

Table B.18 Edits in the Prior Substance Use Module Pertaining to Consistency Checks for Age, Year, and Month of Last Use Variables

Issue	Edits Implemented
The respondent (R) indicated in the final verification check (e.g., LUCG08 for cigarettes) that the age of last use (ALU) based on the new month and year of last use (MLU and YLU) was correct (e.g., LUCG08 = 4). The computer-assisted interviewing (CAI) program updated the ALU (such as AGELSTCG for cigarettes) with the value of the age calculated from the MLU and YLU (referred to subsequently as the MYRLST age, such as MYRLSTCG, for cigarettes). However, the new value for the ALU indicated that the R was more than 1 year younger than his or her current age at the time the R last used the drug (e.g., the R was 16, reported last use of the drug at age 15, but then confirmed an MLU and YLU that meant the R was 14 when the R last used the drug). Had the R initially reported this ALU, the R would not have been routed to the MLU and YLU questions.	The updated value was retained for the ALU (e.g., last use at age 14 for a 16-year-old R in this example). Based on this updated ALU, it was logically inferred that the R should have skipped the MLU and YLU items. A code of 9989 was assigned to the YLU variable (e.g., CIGYLU for cigarettes), and a code of 89 was assigned to the MLU variable (e.g., CIGMLU).
The final verification check (e.g., LUCG08 for cigarettes) was skipped because the R entered revised data for the MLU and YLU that made them consistent with the ALU.	No editing was done because the R was considered to have resolved the inconsistency.
The final verification check (e.g., LUCG08) was skipped because the R entered a new MLU that was the same as the R's birth month.	The new MLU could be consistent with the ALU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the ALU, MLU, and YLU, as long as the revised MLU and YLU were potentially consistent with the ALU. However, the MLU and YLU were set to bad data if they could never be consistent with the ALU. Suppose, for example, that a hypothetical R was born in June 1996, was interviewed in March 2014 (age 17 at the time of the interview), reported last use of cigarettes at age 17, and initially reported last use in May 2013. Last use in May 2013 would have meant that the R was 16 when he or she last smoked cigarettes because the R's 17th birthday was not until June 2013. If the R changed the month and year to June 2013, that could be consistent with last use at age 17, if the use occurred after the R's birthday. However, if the R changed the month and year to June 2012, it would never be possible for the R to have last used at age 17 and also to have last used in June 2012. In this latter situation, CIGMLU and CIGYLU would be set to bad data.
The consistency check was triggered between the ALU and the MYRLST age. However, the R answered the first consistency check (e.g., LUCG05 for cigarettes, regarding whether the value in the MYRLST age was correct) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The value for the ALU was retained, but the MLU and YLU were set to bad data.

Table B.18 Edits in the Prior Substance Use Module Pertaining to Consistency Checks for Age, Year, and Month of Last Use Variables (continued)

Issue	Edits Implemented
The final verification check (e.g., LUCG08 for cigarettes) was skipped because the R entered the same values for the MLU and YLU that triggered the inconsistency with the ALU in the first place.	The value for the ALU was retained, but the MLU and YLU were set to bad data.
The R entered a new MLU or YLU that differed from what the R previously reported. The MYRLST age based on the revised MLU and YLU still mismatched the ALU, but the R indicated in the final verification check that this MYRLST age was correct.	No editing was done in this situation. The CAI program automatically updated the ALU to be consistent with the updated values reported for the MLU and YLU.
The consistency check was triggered between the ALU and the MYRLST age calculated from the MLU and YLU. However, the R answered the first consistency check (e.g., LUCG05 for cigarettes) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The ALU was retained, but the MLU and YLU were set to bad data.
The R entered values in the consistency checks for the MLU and YLU that again yielded a nonmissing MYRLST age based on the MLU and YLU. However, the R failed to resolve the inconsistency between the ALU and the MYRLST age. The R also reported either in the first verification check that the MYRLST age was not correct (e.g., LUCG05 = 6 for cigarettes) or reported in the second verification check that the ALU was correct (e.g., LUCG06 = 2 for cigarettes).	No editing was done to the ALU. The following edits were implemented for the MLU and YLU: • The default edit was to set the MLU and the YLU to bad data. As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the MLU and YLU were assigned the code that corresponded to the answer in the final verification check.
The R entered values in the consistency checks for the MLU and YLU that again yielded a nonmissing MYRLST age based on the MLU and YLU. However, the R failed to resolve the inconsistency between the ALU and the MYRLST age. The R also reported in the second verification check that neither the ALU nor original MYRLST age was correct (e.g., LUCG06 = 3 for cigarettes).	The following edits were implemented for the ALU, MLU, and YLU: • The default edit was to set the ALU, MLU, and YLU to bad data. As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the ALU, MLU, and YLU were assigned the code that corresponded to the answer in the final verification check.
The R triggered the initial consistency check between the ALU, MLU, and YLU and reported that the MYRLST age calculated from the MLU and YLU was correct (e.g., LUCG05 = 4). However, the R answered the second consistency check (e.g., LUCG06 for cigarettes) as "don't know" or "refused." Consequently, the R did not have an opportunity to correct the inconsistency between the ALU, MLU, and YLU.	The relevant codes for "don't know" or "refused" were assigned to the ALU, MLU, and YLU. The rationale for this edit is that conclusive information did not exist regarding whether the ALU indicated the R's correct age when he or she last used a drug or whether the MLU and YLU indicated the R's correct ALU. Therefore, the ALU, MLU, and YLU all were set to missing values.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks)

Issue	Edits Implemented
The age of last use (ALU) for a given drug or behavior (MRJAGLST for marijuana, CIGDLLST for daily cigarette smoking) was greater than the respondent's (R's) current age.	The ALU was set to bad data.
The ALU was lower than the age at first use (AFU) from the core (e.g., MJAGE for marijuana). If the ALU for drugs, cigarettes, or daily cigarette use was lower than the corresponding AFU, a consistency check was triggered. The Rs were not allowed to go back and change the AFU from the core; the only way they could resolve the inconsistency was by making the ALU consistent with the AFU. Thus, the Rs may indicate that the ALU was correct as reported, or they could enter a new ALU that still was inconsistent with the AFU.	The ALU was set to bad data, even if the R reported in the consistency check that "yes," what the R had previously reported for the ALU was correct. Thus, as a *partial* exception to the principle of not editing across modules (Section 2.4.1), core AFU data to edit the noncore ALU, but not vice versa.
The AFU from the core (e.g., MJAGE for marijuana) had a missing value, but the ALU (e.g., MRJAGLST) was defined. That included situations in which the AFU was set to bad data as part of the "flag and impute" edits (see Section 6.2.2.4).	No editing was done to the ALU, as long as it was consistent with the R's age (see above). In addition, no editing was done to the ALU if the imputed AFU (e.g., IRMJAGE for marijuana) was given a value greater than the ALU. (To preserve consistency of the imputed AFU data with imputed data from prior years, imputation of the core AFU variables in 2014 did not take into account data from the corresponding ALU variables.)
The edited core recency (e.g., CIGREC for cigarettes) indicated that the last use was more than 12 months ago, but the ALU indicated last use at the R's current age. The answer to the ALU would suggest use in the past 12 months. (For cigarettes, either CIGAGLST or CIGDLLST could be inconsistent with the recency.)	No editing was done to the ALU. However, a standard codebook footnote was included for these variables to alert analysts to the fact that values in this noncore module could be inconsistent with values from core modules.
The core recency was set to an "indefinite" value of 8 or 9 (used at some point in the past 12 months or used at some point in the lifetime, respectively). These indefinite recency values could be imputed to past month use. If the unedited recency in the corresponding core module indicated use more than 30 days ago, the R would be routed to the prior substance use module and could provide data for the ALU. The latter would suggest that the R is not a past month user, even though the final imputed recency might indicate that the R is a past month user.	No editing was done to the ALU. The standard codebook footnote described above also applied to this issue.
For Rs who last used drugs more than 30 days ago but within the past 12 months, the number of days that the R could have used a drug (e.g., marijuana) in the past 12 months based on MRJAGLST was less than the number of days in the edited or imputed 12-month frequency for marijuana (MJYRTOT or IRMJFY, respectively).	No editing was done to MRJAGLST. The standard codebook footnote described above also applied to this issue.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The ALU variable had a missing value, but the corresponding AFU indicated that the R first used (or started smoking cigarettes daily) at his or her current age.	• The ALU was set to equal the AFU. Logically, last use cannot be any later than the R's current age. • As part of this same edit, if the year of first use (YFU) indicated that the R first used in the year of the interview, then the R was logically inferred to have last used in the current year.
The year of last use (YLU) had a missing value. In addition, the R reported first using a drug (or first smoking cigarettes daily) at an age that was 1 year younger than his or her current age. The R also reported first using the drug (or first smoking cigarettes daily) in the current year.	No editing was done to the ALU, but the YLU was set to equal the YFU. Logically, if the first use was in the current year, then the last use had to be in the current year as well.
The month of last use (MLU) had a missing value, but the corresponding month of first use (MFU, or first daily cigarette use) was in the calendar month prior to in the interview month.	The MLU was set to equal the MFU. Logically, if the R was routed to the prior substance use module because he or she had not used a given drug in the past 30 days, the R also had to have last used that drug in the month prior to the interview month.
The R was interviewed in January and reported first use of a drug (or first daily cigarette use) in December of the previous year.	If the YLU had a missing value or had a value that was inconsistent with the YFU, then the YLU was set to equal the YFU. Concurrently, the MLU was set to equal the MFU under similar constraints.
The ALU had a missing value, but the calculated ALU variable (MYR; e.g., MYRLSTMJ for marijuana) from the YLU, MLU, and birth date was consistent with the AFU and the R's current age. In addition, none of the above edits had been applied.	The ALU was set equal to the calculated ALU (MYR).
The R gave a valid value for the YLU and reported last use in the same month when he or she had a birthday, or else the R answered the MLU question as "don't know" or "refused."	Two possible values for the ALU were calculated, based on the YLU that the R reported and the R's birth year. If neither of these two possible values matched the ALU that the R reported, then the YLU and MLU were set to bad data.
The R reported an ALU (and also may have reported values for the YLU and MLU) for when he or she last used any hallucinogen. However, the ALU, YLU, and MLU variables were skipped for LSD, PCP, or Ecstasy because the R had used only LSD, PCP, or Ecstasy, respectively.	Values from the ALU, YLU, and MLU for any hallucinogen (HALAGLST, HALYLU, and HALMLU, respectively) were transferred to the edited variables that had been skipped for LSD, PCP, or Ecstasy. For example, if the R was a lifetime user only of Ecstasy, data from HALAGLST, HALYLU, and HALMLU were transferred to the corresponding variables ECSAGLST, ECSYLU, and ECSMLU for Ecstasy.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The R reported an ALU (and also may have reported values for the YLU and MLU) for when he or she last used OxyContin® in the core pain relievers module or methamphetamine in the core stimulants module, respectively. However, the ALU, YLU, and MLU variables were skipped for any pain relievers or any stimulants because the R was a lifetime user of only these drugs in the respective modules.	Values from the ALU, YLU, and MLU for OxyContin® (OXYAGLST, OXYCYLU, and OXYCMLU) or methamphetamine (MTHAGLST, METHYLU, and METHMLU) were transferred to the corresponding edited variables that had been skipped for any pain relievers or any stimulants. For example, if the only pain reliever that the R had ever used nonmedically was OxyContin®, data from OXYAGLST, OXYCYLU, and OXYCMLU were transferred to the corresponding variables ANLAGLST, ANALYLU, and ANALMLU for any pain relievers. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module (see the text for further explanation).
The R reported last using any cocaine and crack cocaine in the same year. However, the R also reported last using any cocaine in a month that was earlier than the month when the R reported *first* using crack cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The MLU for any cocaine was set to bad data. Similar edits were implemented for the MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, this edit for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module (see the text for further explanation).
The ALU for crack cocaine was later than the corresponding ALU for any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The ALU for cocaine was set to equal the ALU for crack cocaine. Similar edits were implemented for the ALUs for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, this edit for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The ALU for any cocaine had a missing value. However, the R reported last using crack cocaine at his or her current age. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	• The ALU for any cocaine was set to equal the ALU for crack cocaine. Logically, if the R last used crack at his or her current age, then that also had to be the age at which the R last used any cocaine. • As part of this same edit, if the crack YLU variable indicated that the R last used crack in the current year, then the cocaine YLU was set to the current year. Similar edits were implemented for the ALUs for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The YLU for crack cocaine was later than the YLU for any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The YLU for any cocaine was set to equal the YLU for crack cocaine. Values in the MLU for crack cocaine also were carried over to the MLU for any cocaine. Similar edits were implemented for the YLU and MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The YLU for cocaine had a missing value. However, the R also indicated that he or she last used crack cocaine at an age that was 1 year younger than his or her current age. The R also reported last using crack cocaine in the current year. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	No editing was done to the ALU for any cocaine, but the YLU for any cocaine was set to the current year. Logically, if the last use of crack cocaine was in the current year, then the last use of any cocaine had to be in the current year as well. Similar edits were implemented for the YLU and MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The R last used any cocaine and crack cocaine at the same age and in the same year. However, the R reported last using crack cocaine in a month that was later than what the R reported for last use of any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The MLU for any cocaine was set to equal the MLU for crack cocaine. Similar edits were implemented for the MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The cocaine MLU had a missing value, but the R reported last using crack cocaine in the calendar month prior to the interview month. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The MLU for any cocaine was set to the calendar month prior to the interview month. Logically, if the R was asked the questions about any cocaine and crack cocaine in the prior substance use module because he or she had not used these drugs in the past 30 days, the R also had to have last used any cocaine in the month prior to the interview month. Similar edits were implemented for the MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The ALU for any cocaine had a missing value, but the calculated value in the MYR variable for crack cocaine was consistent with both the crack ALU and the R's current age. In addition, none of the above edits was applied. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The ALU for any cocaine was set to equal the MYR value for crack cocaine. Similar edits were implemented for the ALUs for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The R reported last using any hallucinogen and LSD, PCP, or Ecstasy in the same year. However, the R also reported last using any hallucinogens in a month that was earlier than the month when the R reported *first* using LSD, PCP, or Ecstasy.	The MLU for any hallucinogens was set to bad data.
At least one of the ALUs for LSD, PCP, and Ecstasy was greater than the ALU for any hallucinogen.	• The hallucinogen ALU was set to the latest ALU from the ALUs for LSD, PCP, or Ecstasy. • As part of this same edit, if a given ALU for LSD, PCP, or Ecstasy was moved over to the ALU for any hallucinogens and the corresponding YLU for LSD, PCP, or Ecstasy was later than the YLU for any hallucinogen, then the later YLU was moved over to the YLU for any hallucinogen. If two or all three of these drugs had the same latest ALU value relative to the hallucinogen ALU, then the latest YLU from these drugs was moved over to the YLU for any hallucinogen. • If the YLU for LSD, PCP, or Ecstasy was moved over to the YLU for any hallucinogen and the hallucinogen MLU had a nonmissing value that was earlier than the corresponding MLU for LSD, PCP, or Ecstasy, then the MLU for any hallucinogen was set to the latest MLU from LSD, PCP, or Ecstasy.
The ALU for any hallucinogens had a missing value, but one or more ALUs for LSD, PCP, or Ecstasy indicated that the R last used at his or her current age.	• The hallucinogen ALU was set to be equal to the R's current age. • As part of this same edit, if a given YLU for LSD, PCP, or Ecstasy indicated that the R last used in the current year, then the hallucinogen YLU was set to the current year.
The ALU for any hallucinogens had a missing value, and one or more ALUs for LSD, PCP, or Ecstasy indicated that the R last used at a year younger than his or her current age. At least one YLU for LSD, PCP, or Ecstasy indicated that the R last used any of these drugs in the current year.	The hallucinogen YLU was set to the current year. No editing was done to the hallucinogen ALU.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The MLU for any hallucinogens had a missing value, but the R reported last using LSD, PCP, or Ecstasy in the calendar month prior to the interview month.	The MLU for any hallucinogens was set to the calendar month prior to the interview month. Logically, if the R was asked the questions about any hallucinogens and LSD, PCP, or Ecstasy because he or she had not used these drugs in the past 30 days, the R also had to have last used any hallucinogens in the month prior to the interview month.

Table B.20 Edits in the Prior Substance Use Module Pertaining to the Source of Psychotherapeutics Variables

Issue	Edits Implemented
Respondents (Rs) last used a particular psychotherapeutic drug in the past 30 days. They reported getting the drug "in some other way" in the past 30 days and specified a source that corresponded to one of the available response options for obtaining the drug in the past 30 days.	If Rs did not report getting the drug from that particular source in the past 30 days, the corresponding variable was assigned a code of 3. For example, if Rs did not report that they got pain relievers from a friend or relative for free in the past 30 days but they reported this in the "OTHER, Specify" variable ANLOTHSP (corresponding to question LU28SP), then the 30-day variable for getting pain relievers from a friend or relative for free (ANLFRFRE) was assigned a code of 3. In this example, a code of 3 meant, "Response entered LOGICALLY ASSIGNED (from ANLOTHSP)." Similar edits were implemented if Rs reported that they obtained the drug from a friend or relative for free in the past 30 days, the friend or relative obtained the drug "in some other way," and the source of the friend or relative's drug corresponded to one of the available response options for how the friend or relative obtained medication that the R used nonmedically in the past 30 days.
Rs last used a particular drug in the past 30 days. They reported getting the drug in a specific way the last time they used it nonmedically that they did not report for the past 30 days.	Logically, if Rs last used a drug in the past 30 days, the way that they got the drug the last time they used it nonmedically also applied to the past 30 days. The corresponding past month variable was assigned a code of 5. For example, if Rs did not report that they got pain relievers from a friend or relative for free in the past 30 days but they reported in question LU28 that they got the last pain relievers they misused in this way, then ANLFRFRE was assigned a code of 5. In this example, a code of 5 meant, "Response entered LOGICALLY ASSIGNED (from ANLGTLAS)." Similar edits were implemented if Rs reported that they obtained the drug from a friend or relative for free in the past 30 days, and the Rs also reported that the friend or relative obtained the drug that the Rs last used in a way that the Rs did not report for the 30-day items on how the friend or relative obtained the drug.
Rs were logically inferred to have gotten medication from friends or relatives for free in the past 30 days (e.g., ANLFRFRE = 3 or 5). Consequently, Rs were not asked how the friends or relatives obtained the medication.	The edited variables pertaining to how the friends or relatives obtained the medication retained a code of 98 (i.e., blank). If the Rs had reported directly that they obtained a psychotherapeutic drug from a friend or relative for free (instead of the response being logically inferred), they would have been asked the subsequent questions about how the friend or relative obtained the drug.

Table B.20 Edits in the Prior Substance Use Module Pertaining to the Source of Psychotherapeutics Variables (continued)

Issue	Edits Implemented
Rs reported that they got the last drug they used nonmedically in "some other way," but they specified getting it in one of the ways that they had been asked about previously.	The code from the "OTHER, Specify" variable plus a value of 10 was added to the variable for how Rs obtained the drug the last time they used it nonmedically (i.e., "bump" by 10). For example, if Rs reported getting pain relievers in some other way when they last used them nonmedically (LU28 = 10), but they specified getting them from a friend or relative for free (i.e., ANLGTOSP = 5), then ANLGTLAS was coded as 15. Similar edits were implemented if Rs reported that friends or relatives obtained a drug that the Rs last used in "some other way" and then the Rs specified one of the ways they had asked about previously. For example, if ANLFFOSP was coded as 5 (i.e., the friend or relative got the pain reliever from another friend or relative for free), then ANLFFLAS (how the friend or relative obtained the pain reliever that the R last used nonmedically) was coded as 15.
Rs were logically inferred to have gotten the last medication they used nonmedically from friends or relatives for free (e.g., ANLGTLAS = 15 for pain relievers; see above). Consequently, Rs were not asked how the friends or relatives obtained the medication that the Rs last used.	The corresponding edited variable pertaining to how the friends or relatives obtained the medication that the Rs last used retained a code of 98 (i.e., blank).
Rs were logically inferred to have gotten a drug from one of the sources they were asked about for the last time they used it nonmedically. The Rs also used the drug in the past 30 days and did not report getting the drug from that source in the past 30 days.	The corresponding 30-day variable was assigned a code of 5. For example, if ANLGTLAS = 15 because getting pain relievers from a friend or relative for free was specified as "some other way" that Rs got the pain relievers the last time they used them nonmedically, then ANLFRFRE was coded as 5, if it was not already coded as 1 or 3 (e.g., if Rs had not previously specified getting pain relievers from a friend or relative for free as "some other way" they got them in the past 30 days). Similar edits were in place for the variables pertaining to how friends or relatives obtained a drug that Rs used nonmedically in the past 30 days. For example, if Rs used prescription pain relievers nonmedically in the past 30 days, got the pain relievers they used the last time from friends or relatives for free, and the friends or relatives were logically inferred to have gotten the pain relievers in a particular way, then the friends or relatives were logically inferred to have gotten the pain relievers in this way for pain relievers that the Rs used nonmedically in the past 30 days.
Rs reported that they or friends/relatives got a drug in the past 30 days or the last time in "some other way." However, the Rs explicitly specified that the drug they or friends/relatives obtained was an over-the-counter (OTC) drug.	No editing was done when Rs specified that they had obtained OTCs.

Table B.20 Edits in the Prior Substance Use Module Pertaining to the Source of Psychotherapeutics Variables (continued)

Issue	Edits Implemented
Rs last used a particular drug in the past 30 days. They reported getting the drug in "some other way" the last time they used it nonmedically, and they did not report getting the drug in some other way in the past 30 days.	The variable for obtaining the drug in some other way in the past 30 days was assigned a code of 5. The "OTHER, Specify" response for how Rs got the drug the last time they used it nonmedically also was transferred over to the "OTHER, Specify" variable for the past 30 days (provided that Rs had not specified that they obtained OTCs). For example, if Rs reported that they got pain relievers in some other way the last time they used them nonmedically and they specified getting them from a friend, relative, or at home, but they did not specify whether they got them for free, bought them, or took them without asking (ANLGTOSP = 11), then the 30-day variable ANLOTHWY (for "some other way") was assigned a code of 5, and the corresponding "OTHER, Specify" variable ANLOTHSP was coded as 11 as well. Similar edits were implemented if Rs reported that they got a drug from a friend or relative in the past 30 days. In particular, the "other" response for how friends or relatives obtained the drug that the Rs last used nonmedically was transferred to the corresponding "OTHER, Specify" variable for how friends or relatives obtained the drug that Rs used nonmedically in the past 30 days (e.g., ANLFFOSP for pain relievers), provided that the R did not specify that the friend or relative had obtained OTCs.

Table B.21 Edits in the Prior Substance Use Module Pertaining to the Sequence of Initiation for Alcohol, Cigarettes, and Marijuana

Issue	Edits Implemented
All age-at-first-use (AFU) variables for these drugs had a code of 997, indicating the respondent (R) refused to answer the AFU questions for these drugs, or the R refused to answer the lifetime use question.	A code of 97 (REFUSED) was assigned to each sequence-of-use variable USEALCG, USEMJCG, USEALMJ, USEACM, or USENEXT that was blank. Otherwise, if the R was routed to a given question for some reason, a bad data code was assigned to the edited variable.
One or both AFUs for the drugs of interest had missing values, but the corresponding sequence-of-use variable had a valid value. For example, the AFU variables AGE1STAL and AGE1STCG that were stored by the computer-assisted interviewing (CAI) program had equal values while the interview was in progress, such that the R was asked USEALCG. In editing of the core data, however, the cigarette AFU variable CIGTRY could have been set to bad data because of an inconsistency with the cigarette recency CIGREC. Similarly, the alcohol AFU variable ALCTRY could have been set to bad data because of an inconsistency with the alcohol variable ALCREC.	A code was assigned that was equal to the original value of the sequence variable plus 10 (i.e., "bump" by 10). This edit preserved data indicating that the R was routed to a particular item and how the R answered, but it would alert analysts that one or both of the AFUs that routed the R to the item was questionable. For example, if question LU22, corresponding to USEALCG, was coded as 1, indicating that the R started using alcohol before using cigarettes, but the AFU for alcohol or the AFU for cigarettes had missing values (e.g., bad data), then USEALCG was assigned a code of 11.
The R initiated use of a given pair of drugs (e.g., alcohol and cigarettes) within 1 year of his or her current age. However, the sequence variables were not consistent with the *edited* year-of-first-use (YFU) and month-of-first-use (MFU) variables from the core modules. For example, the R reported in question LU22 that he or she first used alcohol before using cigarettes. However, the edited YFU and MFU data from the tobacco and alcohol modules indicated that the R first used cigarettes in an earlier year or month than when the R first used alcohol.	Where edited YFU and MFU data existed for a given pair of drugs, the edits logically inferred a sequence of use that was consistent with the core YFU and MFU data. For example, if the R answered question LU22 as 1 (corresponding to USEALCG), that would indicate that the R used alcohol before using cigarettes. However, if the cigarette YFU and MFU data indicated use of cigarettes before alcohol, USEALCG was assigned a code of 4, where 4 = Cigarettes LOGICALLY ASSIGNED. Similarly, if the R answered LU22 as 2, indicating that the R used cigarettes before using alcohol, but the alcohol YFU and MFU data indicated use of alcohol before cigarettes, then a code of 3 was assigned to USEALCG, where 3 = Alcohol LOGICALLY ASSIGNED. These edits also were performed when USEALCG has missing values (e.g., "don't know" [DK] or "refused" [REF]), but YFU and MFU data from the core could be used to infer a nonmissing value in USEALCG. These principles also applied to edits of USEMJCG and USALMJ. The rationale for these edits was that the Rs themselves provided data in the core modules that indicated the sequence with which they used these drugs.

Table B.21 Edits in the Prior Substance Use Module Pertaining to the Sequence of Initiation for Alcohol, Cigarettes, and Marijuana (continued)

Issue	Edits Implemented
Values of 1 or 2 in question LU26 (corresponding to USENEXT) had different meanings, depending on how the R answered question LU25 (corresponding to USEACM).	USENEXT was recoded to parallel the levels in USEACM (1 = Alcohol; 2 = Cigarettes; 3 = Marijuana). • USEACM = 1 indicated that the R first used alcohol before using cigarettes or marijuana. When USEACM = 1, a value of 1 in LU26 indicated that the R used cigarettes next. A value of 2 in LU26 indicated that the R used marijuana next. Therefore, when USEACM = 1, USENEXT was coded as 2 when LU26 = 1 and was coded as 3 when LU26 = 2. • USEACM = 2 indicated that the R first used cigarettes before using alcohol or marijuana. When USEACM = 2, a value of 1 in LU26 indicated that the R used alcohol next. A value of 2 in LU26 indicated that the R used marijuana next. Therefore, when USEACM = 2, USENEXT was coded as 1 when LU26 = 1 and was coded as 3 when LU26 = 2. • USEACM = 3 indicated that the R first used marijuana before using cigarettes or alcohol. When USEACM = 3, a value of 1 in LU26 indicated that the R used alcohol next. A value of 2 in LU26 indicated that the R used cigarettes next. Therefore, when USEACM = 3, USENEXT was coded as 1 when LU26 = 1 and was coded as 2 when LU26 = 2.
Data existed in USEACM (LU25), but one or more AFUs had been set to bad data (e.g., because of an inconsistency in the core modules between the AFU and the recency).	Values other than those for DK, RE, or bad data were bumped by 10 in USEACM and USENEXT. As was the case for USEALCG, USEMJCG, and USEALMJ, this edit preserved data in USEACM and USENEXT while making analysts aware that there is a potential issue with the data.
Question LU25 (corresponding to USEACM) was answered as 1, indicating that the R first used alcohol. However, this response was inconsistent with one or more YFU or MFU variables from the core modules for cigarettes, alcohol, or marijuana.	The core YFU and MFU data for cigarettes and marijuana were checked for indications that the R used either of these drugs first. • If it could be determined unambiguously that the R used cigarettes before using marijuana or alcohol, the edits assigned a code of 5 to USEACM, where 5 = Cigarettes LOGICALLY ASSIGNED. For example, if the respective YFU variables for cigarettes, alcohol, and marijuana, CIGYFU, ALCYFU, and MJYFU, all were defined and CIGYFU was earlier than the other two (but, by definition, the R first used all three substances at the same age), it could logically be inferred that cigarette use occurred first. • Similarly, if it could be determined unambiguously that the R first used marijuana before using cigarettes or alcohol, the edits assigned a code of 6 to USEACM, where 6 = Marijuana LOGICALLY ASSIGNED. Otherwise, if there was some indication that the R used either marijuana or cigarettes before using alcohol but incidence data were not fully defined for all three substances, USEACM was set to bad data.

Table B.21 Edits in the Prior Substance Use Module Pertaining to the Sequence of Initiation for Alcohol, Cigarettes, and Marijuana (continued)

Issue	Edits Implemented
Question LU25 (corresponding to USEACM) was answered as 2, indicating that the R first used cigarettes. However, this response was inconsistent with one or more YFU or MFU variables from the core modules for cigarettes, alcohol, or marijuana.	The core YFU and MFU data for alcohol and marijuana were checked for indications that the R used either of these drugs first. • If it could be determined unambiguously that the R used alcohol before using marijuana or cigarettes, the edits assigned a code of 4 to USEACM, where 4 = Alcohol LOGICALLY ASSIGNED. • Similarly, if it could be determined unambiguously that the R first used marijuana before using cigarettes or alcohol, the edits assigned a code of 6 to USEACM, where 6 = Marijuana LOGICALLY ASSIGNED. Otherwise, if there was some indication that the R used either alcohol or marijuana before using cigarettes but incidence data were not fully defined for all three substances, USEACM was set to bad data.
Question LU25 (corresponding to USEACM) was answered as 3, indicating that the R first used marijuana. However, this response was inconsistent with one or more YFU or MFU variables from the core modules for cigarettes, alcohol, or marijuana.	The core YFU and MFU data for cigarettes and alcohol were checked for indications that the R used either of these drugs first. • If it could be determined unambiguously that the R used alcohol before using marijuana or cigarettes, the edits assigned a code of 4 to USEACM, where 4 = Alcohol LOGICALLY ASSIGNED. • Similarly, if it could be determined unambiguously that the R first used cigarettes before using alcohol or marijuana, the edits assigned a code of 5 to USEACM, where 5 = Cigarettes LOGICALLY ASSIGNED. Otherwise, if there was some indication that the R used either alcohol or cigarettes before using marijuana but incidence data were not fully defined for all three substances, USEACM was set to bad data.
USEACM had been assigned codes of 4, 5, 6, or 85 (bad data).	To edit USENEXT, a sequence of logic was followed similar to that outlined above for USEACM. For example, if USEACM had been set to a value of 4 (used alcohol first) and the core YFU and MFU data for cigarettes and marijuana indicated that the R used cigarettes next, set USENEXT was set to a value of 5 (Cigarettes LOGICALLY ASSIGNED), if USENEXT did not already indicate that the R used cigarettes next. If USEACM had been set to bad data and USENEXT was not blank, USENEXT also was set to bad data.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004

Issue	Edits Implemented
The respondent's (R's) only report(s) of drug use in the core drug modules that routed the R into question TX01 about lifetime substance treatment had been set to bad data as part of the core drug editing.	Nonblank values in the edited variables pertaining to receipt of substance treatment were replaced with bad data codes.
Responses to the questions on the receipt of treatment in the past 12 months and the last time that the R received treatment were inconsistent (e.g., if the R reported that he or she did not receive treatment in the past 12 months but subsequently reported last receiving treatment during that period).	The edits favored responses that indicated more recent receipt of treatment: • If an R responded affirmatively that he or she had received treatment in the past 12 months but reported last receiving treatment "more than 12 months ago," the edits logically inferred that the R last received treatment at some point in the past 12 months (i.e., TXLASREC = 8). • If an R reported that he or she did not receive treatment in the past 12 months but reported last receiving treatment in the past 12 months, the edits logically inferred that the R had received treatment in that period (i.e., TXYREVER = 3).
The question on the receipt of treatment in the past 12 months had missing data (e.g., a response of "don't know" or "refused"), but the question on the last time that the R received treatment did not. Alternatively, the question on the last time that the R received treatment had missing data, but the question on receipt of treatment in the past 12 months did not.	Where possible, data were used to replace the missing value with a nonmissing value. Suppose, for example, that the R did not know or refused to report whether he or she had received treatment in the past 12 months. • If the R reported last receiving treatment in this period, the ambiguous response was replaced with a value to indicate that the R had received treatment in this period (i.e., TXYREVER = 3). • If the R reported last receiving treatment more than 12 months ago, it was logically inferred that the question about receipt of any treatment in the past 12 months should have been answered as "no" (i.e., TXYREVER = 4).
The question about the most recent receipt of treatment had missing data.	Data from the past year treatment variable (TXYREVER) or the lifetime treatment variable (TXEVER) were used to replace missing values in the edited treatment recency variable TXLASREC. • If the R had received treatment in the past 12 months (TXYREVER = 1), TXLASREC was assigned a code of 8 to indicate treatment at some point in the past 12 months. • If the R had not received treatment in the past 12 months (TXYREVER = 2), TXLASREC was assigned a code of 13 to indicate that the last treatment episode was more than 12 months ago. • If the question about treatment in the past 12 months (TX02) was answered as "don't know" or "refused" and the R did not report currently being in treatment in question TX07, TXLASREC was assigned a code of 9 to indicate treatment at some point in the R's lifetime. (The R had to have answered the lifetime treatment question TX01 as "yes" in order to have been routed to TX02.)

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R reported currently being in treatment in question TX07, so the question about the most recent time that the R had been in treatment was skipped.	The edited variable corresponding to question TX07 (TXRCVNOW) continued to be coded as 1 (i.e., "yes"). Instead of a legitimate skip code being assigned, the edited treatment recency variable (TXLASREC) was assigned a code of 7, where 7 = Still in treatment LOGICALLY ASSIGNED. A code of 21 (still in treatment) also was assigned to the treatment outcome variable TXLTYOUT.
The R reported currently being in treatment in question TX07 but did not know or refused to report in question TX02 whether he or she had received treatment in the past 12 months.	The R was logically inferred to have received treatment in the past 12 months (TXYREVER = 3).
The R was routed to question TX07 but did not know or refused to report whether he or she was still in treatment.	Data from the treatment outcome variable (TXLTYOUT) and treatment recency variable (TXLASREC) were used to replace missing values in the edited variable TXRCVNOW, corresponding to TX07. • If the R had received treatment in the past 30 days (TXLASREC = 1) and reported still being in treatment (TXLTYOUT = 1 or 21), TXRCVNOW was assigned a code of 3 to indicate that the R logically was still in treatment. • Otherwise, if the R definitely had not received treatment in the past 30 days, TXRCVNOW was assigned a code of 4 to indicate that the R logically was no longer in treatment.
The R reported that he or she was not currently in treatment (TXRCVNOW = 2 or 4), but the R reported still being in treatment when asked about the outcome of the last treatment episode.	The treatment outcome variable (TXLTYOUT) was assigned a bad data code.
The R specified receiving treatment for an over-the-counter (OTC) psychotherapeutic medication (e.g., aspirin).	This information on OTC drugs was not used to infer treatment for any of the psychotherapeutic drugs because the questions about receipt of treatment for psychotherapeutic drugs referred specifically to treatment for prescription-type medications (i.e., not to OTCs).
The R did not report receiving treatment for a particular drug during his or her last (or current) treatment episode, but treatment for this drug was specified as treatment for "some other drug." In the case of the psychotherapeutics, the "other" drug specified was not an OTC drug.	The R was inferred to have received (or be receiving) treatment for the use of that drug. A code of 3 (Yes LOGICALLY ASSIGNED) was assigned to the corresponding edited drug variable. For example, Rs who did not report receiving treatment for prescription stimulants but reported receiving treatment for street stimulants were considered to qualify as having received treatment for prescription-type stimulants (i.e., those that were not available as OTCs, which would include street drugs). In this example, the edited variable TXLTYSTM would be assigned a code of 3.
The R did not report receiving treatment for a particular drug during his or her last (or current) treatment episode but indicated that this drug was the primary drug for which he or she last received treatment (or was currently receiving treatment).	The R was inferred to have received (or be receiving) treatment for the use of that drug.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R reported receiving treatment only for alcohol in the past 12 months, but questions about treatment for specific drugs during the last or current treatment episode had missing values (i.e., "don't know," "refused," bad data, or blank).	The R was logically inferred not to have received treatment for that drug during the last or current episode. The missing value in that drug's variable was replaced with a special code of 4 (No LOGICALLY ASSIGNED).
The R was routed to questions about the last or current treatment episode but did not have any indication of treatment for any of the drugs that he or she ever used.	The following edits were implemented: • If the R reported receiving treatment only for alcohol in the past 12 months, a special logically inferred "yes" code of 5 was assigned to the variable for alcohol treatment during the last treatment episode (TXLTYALC). • If the R reported receiving treatment only for drugs other than alcohol in the past 12 months, a special code of 5 was assigned to the "some other drug" variable (TXLTYSOD) to indicate that the drug for which the R received treatment was unknown. • Otherwise, a special code of 7 was assigned to TXLTYSOD, the "some other drug" variable, to indicate that treatment for alcohol or other drugs was unknown. (Prior to 2002, these edits required Rs to have denied receiving treatment for all drugs they had ever used. Since 2002, the above edits also have been implemented if Rs did not report treatment for any specific drugs, and missing data existed in the questions about treatment for specific drugs.)
The R was logically inferred to have received treatment for alcohol during the last or current treatment episode (TXLTYALC = 5), and question TX36 about treatment for any other drug (TXLTYSOD) was answered as "no."	The edited "OTHER, Specify" variables TXLTYA through TXLTYE were assigned legitimate skip codes.
The R refused to report in question TX36 whether he or she received treatment for any other drug.	The refusal was propagated onto the edited "OTHER, Specify" variables TXLTYA through TXLTYE. Since 2002, this edit has been implemented regardless of whether the R had reported treatment for at least one drug in questions TX26 through TX35. (Prior to 2002, this edit required at least one response of "yes" in TX26 through TX35.)
The R reported treatment only for "some other drug," but the only substances specified were tobacco products (i.e., cigarettes, chewing tobacco, snuff, cigars, pipe tobacco).	The variables specifying treatment for tobacco products were assigned bad data codes. In addition, other variables pertaining to the last (or current) treatment episode were assigned bad data codes if the items had been answered. The following variables were affected: TXLTYMN (i.e., main place where the R was last treated); TXLTYOUT (i.e., outcome of the last treatment episode); variables beginning with TXPY (i.e., payment sources for the last [or current] treatment episode); and TXLTYDUR (i.e., length of the last or current treatment). The rationale for these edits was that anything pertaining to the last treatment (e.g., payment sources for the last or current treatment) would logically be assumed to pertain to treatment only for tobacco.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
If the R reported in the alcohol module that he or she never used alcohol (AL01 = 2), the R would be skipped out of question TX26, pertaining to receipt of treatment for alcohol. However, the R also could report in question TX03 that he or she received treatment for "alcohol only" or "alcohol and drugs" in the past 12 months.	The edited variable pertaining to receipt of alcohol treatment during the last or current episode (TXLTYALC) retained a code of 98 (blank).
Question TX37 pertaining to the main drug for which the R last received (or was currently receiving) treatment was skipped because the R reported receipt of treatment for only one drug during the last or current treatment episode. That includes situations in which the only drug for which the R reported receiving treatment was "some other drug."	The edited variable TXLTYPRM was assigned a legitimate skip code, provided that none of the edited variables about treatment for alcohol through prescription sedatives (TXLTYALC through TXLTYSED) had a code of 98 (blank). Otherwise, TXLTYPRM retained a code of 98.
The R reported being in treatment for 366 days in the past 12 months.	The edited variable TXLTYDUR was edited to 365 days.
The length of time that the R reported currently being in treatment or being in treatment the last time translated to a number of years greater than the R's age.	The edited variable TXLTYDUR was assigned a bad data code.
The R reported receiving treatment in the past 12 months and reported receiving treatment in the past 12 months for alcohol only or drugs only. However, this response was inconsistent with the responses to questions on the drugs for which the R was treated (or was being treated) during the last (or current) treatment episode. For example, the R reported being treated in the past 12 months only for alcohol but reported last being treated for use of one or more other drugs.	Logically, the last or current treatment episode would fall within the 12-month period prior to the interview. Therefore, the variable pertaining to receipt of treatment for alcohol, other drugs, or both in the past 12 months (TXYRADG) was edited as follows: • If the R originally indicated treatment for alcohol only (i.e., a code of 1 in question TX03), with treatment for other drugs also having been indicated during the last episode, a special code of 11 was assigned to TXYRADG. • If the R originally indicated treatment for drugs only (i.e., a code of 2 in question TX03), with treatment for alcohol also having been indicated during the last episode, a special code of 12 was assigned to TXYRADG. The edits were done in this manner because the subsequent fill pattern for specific locations where the R received treatment in the past 12 months was based on the R's original answer for receipt of treatment only for alcohol, only for other drugs, or both.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R reported receiving treatment in the past 12 months but did not know or refused to report whether he or she received treatment only for alcohol, only for other drugs, or for both. However, data were provided on the drugs for which the R was treated during his or her last (or current) treatment episode.	Data on the drugs for which the R was last treated (or was currently being treated) were used to indicate the *minimum* for which the R could have been treated in the past 12 months: • If the R indicated last (or currently) being treated for alcohol but did not indicate treatment for other drugs during the last (or current) treatment episode, it was possible to infer in TXYRADG that the R was at least treated for alcohol in the past 12 months in TXYRADG (but the R also may have been treated for other drugs at some point during that period). A special code of 4 was assigned to TXYRADG. • If the R indicated last (or currently) being treated for one or more drugs other than alcohol but did not indicate treatment for alcohol, it was possible to infer in TXYRADG that the R was at least treated for drugs other than alcohol in the past 12 months. A special code of 5 was assigned to TXYRADG. • If the R reported last (or currently) being treated both for alcohol and for other drugs, it was possible to infer in TXYRADG that the R was treated for both alcohol and other drugs in the past 12 months. A special code of 6 was assigned to TXYRADG.
The R reported receiving treatment in the past 12 months and did not report receiving treatment in a particular location in the past 12 months, but this location was specified as treatment in "some other place" in the past 12 months.	The R was logically inferred to have received treatment in that location in the past 12 months. A code of 3 (Yes LOGICALLY ASSIGNED) was given to the edited treatment location variable in this situation.
The R reported receiving treatment in the past 12 months (or was inferred to have received treatment in the past 12 months) and did not report receiving treatment in a particular location that he or she subsequently reported was the main place that he or she received treatment the last time (or the main place where he or she was currently receiving treatment).	The R was logically inferred to have received treatment in that location in the past 12 months. A special logically assigned "yes" code of 5 was assigned to indicate that the affirmative response came from the data on the main location where the R last received (or was currently receiving) treatment. If the R reported that the main location where he or she received treatment was "some other place" and specified a valid response in question TX25SP, that "OTHER, Specify" response also was moved over to the "OTHER, Specify" variable TXYROTSP pertaining to treatment locations in the past 12 months.
The R reported receiving treatment in the past 12 months but answered "no" to every item about particular locations for treatment in that period, including "some other place."	The edited variable pertaining to "some other place" (TXYRSOP) was assigned a special code of 7 to indicate that the treatment location was unknown.
The R reported receiving treatment in the past 12 months and did not initially indicate receiving treatment in a hospital emergency room in that period. However, the R subsequently reported receiving treatment in the past 12 months in an emergency room for use of marijuana, cocaine, heroin, LSD, PCP, or methamphetamine.	The variable that did not indicate treatment in an emergency room (TXYRTXER) was edited to infer that the R had received treatment in that location in the past 12 months.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R reported receiving treatment in the past 12 months in every specific location that was asked about (i.e., except for treatment in "some other place").	No editing was done if the R reported being or having been in treatment for 15 days or more. If the R reported being or having been in treatment for fewer than 15 days, however, responses of "yes" in the entire list of edited past year treatment location variables were replaced with bad data codes. If treatment in "some other place" also was reported, the edited variable TXYRSOP was assigned a bad data code. In the variable TXYROTSP (i.e., the other treatment location that was specified), any responses were replaced with bad data codes. If the R also reported that he or she was still in treatment (TX07 = 1), the edited variable TXRCVNOW also was assigned a bad data code.
Rs could report still being in treatment in question TX07 but may report that they received treatment only in jail in the past 12 months.	When Rs reported receiving treatment only in jail in the past 12 months, they were logically inferred not to be currently in treatment (TXRCVNOW = 4). This edit also applied when Rs reported that the main treatment location was jail, and this was the only specific location reported for treatment in the past 12 months.
The R did not report a particular payment source for his or her last episode of treatment but specified this payment source as "some other source."	The R was inferred to have used that particular payment source for treatment. A code of 3 (Yes LOGICALLY ASSIGNED) was assigned to the edited variable for that payment source.
The R answered all items about payment sources for treatment as "no," including the item indicating that the last treatment was free.	A special code of 5 was assigned to the edited "some other source" variable (TXPYSOS) to indicate that the payment source was unknown.
The R reported that every specific payment source that was asked about paid for his or her last episode of treatment (i.e., except for "some other source" and free payment, the latter of which would have been skipped).	All source of payment variables that the R had answered as "yes" were assigned a bad data code. That included situations in which "some other source" of payment also was reported. In the variable TXPYSP (i.e., the other payment source that was specified), any responses were replaced with bad data codes.
The R reported *all* of the following: • receipt of treatment in every specific location in the past 12 months (i.e., except for treatment in "some other place"), and • payment of the last treatment by every specific payment source (i.e., except for "some other source" and free treatment).	When this specific pattern occurred, data from additional variables also were assumed to be questionable. Responses entered for the following variables were replaced with bad data codes: TXYRADG (i.e., treatment for alcohol, drugs, or both in the past 12 months); TXYRVSER (treatment in an emergency room for marijuana, cocaine, heroin, LSD, PCP, or methamphetamine in the past 12 months); TXYRNMER (number of times the R visited an emergency room for treatment of the above drugs); TXLTYMN (the main place the R received treatment the last time); drugs that the R was asked about for the last treatment episode (including the main drug for which the R received treatment, if applicable); and TXLTYDUR (length of time in treatment currently or the last time).

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R indicated that "some other source" paid for the last treatment, but then specified that this treatment was free.	If no other payment source was indicated, then it was logically inferred that the R's last treatment was free (i.e., TXPYFRE = 3). Otherwise, if one or more payment sources had been indicated previously (e.g., private health insurance, the R's own funds), then it was inferred that "some other source" had not paid for the last treatment. In this situation, the response of free treatment that had been specified also was wiped out in the edited "OTHER, Specify" variable (TXPYSP).
The R reported in question TX44 that the only treatment he or she received in the past 12 months was for detoxification (or answered TX44 as "don't know" or "refused"), but the R also reported attending self-help groups in the past 12 months. Self-help groups typically are not places where people go to receive detoxification.	The response was accepted that the R received treatment in a self-help group in the past 12 months, and the R was logically inferred to have received treatment other than detoxification in that period. The edited variable corresponding to question TX44 (TXYRDTXO) was assigned a code of 4 (No LOGICALLY ASSIGNED).

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004

Issue	Edits Implemented
The respondent (R) reported receiving substance treatment in the past 12 months (i.e., TXYREVER = 1).	The following lifetime treatment variables were assigned legitimate skip codes: TXALONEV, TXALONAG, TXDRONEV, TXDRONAG, TXALCDEV, TXALCDAG, TXDRGAEV, and TXDRGAAG. These variables corresponded to questions TX45, TX45A, TX46, TX46A, TX47, TX47A, TX48, and TX48A, respectively.
The R was logically inferred to have received treatment in the past 12 months (i.e., TXYREVER = 3). Consequently, the R was routed into the lifetime treatment questions. Had the R answered the past year treatment question TX02 as "yes," however, the R would have been skipped out of the lifetime treatment questions.	Responses of 1 and 2 (i.e., "yes" and "no," respectively) in TXALONEV, TXDRONEV, TXALCDEV, and TXDRGAEV were bumped by a value of 10 (i.e., set to values of 11 or 12). No editing was done to these variables if they had values of 94 ("don't know") or 97 ("refused").
The R received treatment in his or her lifetime but not in the past 12 months, and the R never used alcohol (i.e., ALCEVER = 2). The R also reported using at least one other drug in the marijuana through sedatives modules or in the special drugs module for methamphetamine.	The following variables pertaining to lifetime alcohol treatment were assigned legitimate skip codes: TXALONEV, TXALONAG, TXALCDEV, TXALCDAG, TXDRGAEV, and TXDRGAAG.
The R received treatment in his or her lifetime but not in the past 12 months, and the R indicated unambiguously in the marijuana through sedatives modules or in the special drugs module for methamphetamine that he or she never used any of these drugs. The R also reported using alcohol.	All the variables pertaining to drugs are skip filled, namely, TXDRONEV, TXALCDEV, TXDRGAEV and the age variables TXDRONAG, TXALCDAG, and TXDRGAAG.
The R reported first receiving substance treatment for alcohol or other drugs (or both) at an age that was later than his or her current age.	Consistent with standard editing practice, any ages for first receipt of treatment that were inconsistent with the R's current age were set to bad data.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
The R reported first receiving treatment for alcohol or other drugs at an age that was earlier than the age-at-first-use (AFU) data from the core drug use modules or the special drugs module for methamphetamine. • For alcohol, the R reported first receiving treatment for alcohol at an age that earlier than ALCTRY, the AFU for alcohol from the core alcohol module. • For other drugs, the R reported first receiving treatment for his or her use of drugs at an age that was earlier than the minimum AFU for the drugs that the R reported using.	No editing was done to the age at first treatment data that were inconsistent with the AFU data from the core substance use modules. However, flags were created to indicate the degree of deviation from (a) ALCTRY (TXFGALAG), (b) the minimum AFU for other drugs (TXFGDGAG), and (c) the minimum AFU from both ALCTRY and the AFUs for other drugs (TXFGADAG). Values in these flags had the following meanings: 0 = First treatment age was consistent with minimum AFU 1 = First treatment age differs from minimum AFU by 1 year 2 = First treatment age differs from minimum AFU by 2 years 3 = First treatment age differs from minimum AFU by 3 or 4 years 4 = First treatment age differs from minimum AFU by 5 or more years. The flag for both alcohol and drugs (TXFGADAG) was created from the values in TXFGALAG and TXFGDGAG. The final value in TXFGADAG was chosen according to whatever value in TXFGALAG and TXFGDGAG indicated the greatest inconsistency between treatment age and core AFU data. If both TXFGALAG and TXFGDGAG were blank, then TXFGADAG was set to blank as well.
The R could report receiving treatment for alcohol in the past 12 months and could give an age when he or she first received alcohol treatment. However, the R previously reported in the core alcohol module that he or she never used alcohol. Similarly, the R could report receiving treatment for other drugs in the past 12 months and could give an age when he or she first received drug treatment. However, the R previously reported never using drugs in the core marijuana through sedatives modules or in the special drugs module for methamphetamine.	Again, no editing was done to the inconsistent treatment age data. However, the flag variables mentioned above were assigned a code of 5 when this situation occurred. This code of 5 had the following meaning: 5 = Never used alcohol in core but reported an alcohol treatment age 5 = Never used drugs in core but reported a drug treatment age. If TXFGALAG or TXFGDGAG had a value of 5, then TXFGADAG also was assigned a value of 5, where 5 = Any lifetime nonuse with a related treatment age.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
The R answered the lifetime treatment question TX01 (edited variable TXEVER) as "don't know" or "refused."	The following edits were implemented for the lifetime treatment variables TXALONEV through TXDRGAAG according to what the R reported in the core substance use modules for lifetime use (or nonuse) of alcohol or drugs. • If the R was a lifetime user only of alcohol (i.e., had definitely never used other drugs that were covered in the core modules), the lifetime variables pertaining to treatment for drugs (TXDRONEV, TXDRONAG, TXALCDEV, TXALCDAG, TXDRGAEV, and TXDRGAAG) were assigned legitimate skip codes; even if the R had reported lifetime receipt of treatment, these variables still would have been skipped because the R never used any of the drugs covered in the core module. In addition, if TXEVER was refused, that refusal was propagated to TXALONEV and TXALONAG. • If the R was a lifetime user of other drugs but had never used alcohol, the lifetime variables pertaining to treatment for alcohol (TXALONEV, TXALONAG, TXALCDEV, TXALCDAG, TXDRGAEV, and TXDRGAAG) were assigned legitimate skip codes; even if the R had reported lifetime receipt of treatment, these variables still would have been skipped because the R never used alcohol. In addition, if TXEVER was refused, that refusal was propagated to TXDRONEV and TXDRONAG.
The R had the following pattern in his or her data: 1. Lifetime use of alcohol (i.e., ALCEVER = 1). 2. Initially skipped out of past year treatment questions but was logically inferred to have received treatment in the past 12 months (i.e., TXYREVER = 3; see Table B.22). 3. Never used marijuana through sedatives, or methamphetamine (based on answers from special drugs). 4. Question TX45 (edited variable TXALONEV) was answered as "no," meaning that the R had never received treatment or counseling for alcohol. Logically, the R had to have received treatment for something, and alcohol was the only substance that the R reported using.	TXALONEV was bumped by 20, such that TXALONEV would show a value of 22. This was done instead of inferring that the R received alcohol treatment. Nevertheless, a value of 22 in TXALONEV would alert analysts to an inconsistency in the data and give them the option of deciding how to handle these cases in an analysis.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
The R had the following pattern in his or her data: 1. Lifetime use of at least one drug in the marijuana through sedatives modules or methamphetamine use from the special drugs module. 2. Initially skipped out of past year treatment questions but was logically inferred to have received treatment in the past 12 months (i.e., TXYREVER = 3). 3. Never used alcohol. 4. Question TX46 (edited variable TXDGONEV) was answered as "no," meaning that the R had never received treatment or counseling for drugs other than alcohol. Logically, the R had to have received treatment for something, and illicit drugs (but not alcohol) were the only substances that the R reported using.	TXDRONEV was bumped by 20, such that TXDRONEV would show a value of 22. This was done instead of inferring that the R received drug treatment. Nevertheless, a value of 22 in TXDRONEV would alert analysts to an inconsistency in the data and give them the option of deciding how to handle these cases in an analysis.
The R reported lifetime use of both alcohol and other drugs. However, both TXALCDEV = 2 and TXDRGAEV = 2, meaning that the R never got treatment for either alcohol or drugs. Logically, if the R answered the lifetime treatment question TX01 (edited variable TXEVER) as "yes," the R had to have gotten treatment for something.	Both TXALCDEV and TXDRGAEV were bumped by 20, such that the edited value would be 22. Again, this would give analysts the option of deciding how to handle these cases in an analysis.
A lifetime lead variable (i.e., TXALONEV, TXDRONEV, TXALCDEV, and TXDRGAEV) was coded as 2 (i.e., "no") or was coded as 12 (based on the second issue described in this table).	The corresponding age-at-first-treatment variables TXALONAG, TXDRONAG, TXALCDAG, and TXDRGAAG were assigned legitimate skip codes.
A past year lead variable (i.e., TXYALDRG, TXYDRALC) was coded as 2 (i.e., "no").	The corresponding age-at-first-treatment variables TXYALDAG and TXYDRAAG were assigned legitimate skip codes.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
The past year treatment variable TXYRADG (corresponding to question TX03) indicated that the R was treated in the past year only for alcohol or only for drugs.	If TXYRADG = 1 (i.e., treatment in the past year only for alcohol), the following variables were assigned legitimate skip codes: TXYDRAGE (age at first treatment for drugs, if the R had been treated in the past year only for drugs), TXYDRALC (whether the R ever was treated for alcohol use, if the R had ever used alcohol and had been treated in the past year only for drugs), TXYDRAAG (the treatment age variable corresponding to TXYDRALC), TXYADAAG (age at first treatment for alcohol, if the R had been treated in the past year for both alcohol and drugs), and TXYADDAG (age at first treatment for drugs, if the R had been treated in the past year for both alcohol and drugs). Similarly, if TXYRADG = 2 (i.e., treatment in the past year only for drugs), the following variables were assigned legitimate skip codes: TXYALAGE (age at first treatment for alcohol, if the R had been treated in the past year only for drugs), TXYALDRG (whether the R ever had been treated for drug use, if the R had ever used marijuana through sedatives and had been treated in the past year only for alcohol), TXYALDAG (the treatment age variable corresponding to TXYALDRG), TXYADAAG (age at first treatment for alcohol, if the R had been treated in the past year for both alcohol and drugs), and TXYADDAG (age at first treatment for drugs, if the R had been treated in the past year for both alcohol and drugs).
TXYRADG indicated that the R was treated in the past year for both alcohol and drugs, including situations in which the R was logically inferred to have received treatment for both alcohol and drugs in the past year.	The following variables were assigned legitimate skip codes: TXYALAGE, TXYALDRG, TXYALDAG, TXYDRAGE, TXYDRALC, and TXYDRAAG. If TXYRADG = 6 (i.e., logically inferred to have received treatment for both alcohol and drugs in the past year; see Table B.22), the skipped variables TXYADAAG and TXYADDAG retained codes of blank.
TXYRADG had been assigned a code of 4 (i.e., received treatment for alcohol in the past year but treatment for drugs during this period was unknown; see Table B.22).	The following variables pertaining to receipt of treatment only for drugs were assigned legitimate skip codes: TXYDRAGE, TXYDRALC, and TXYDRAAG. The skipped variables pertaining to treatment for alcohol in the past year, with or without treatment for drugs (i.e., TXYALAGE, TXYALDRG, TXYALDAG, TXYADAAG, and TXYADDAG), retained codes of blank.
TXYRADG had been assigned a code of 5 (i.e., received treatment for alcohol in the past year but treatment for drugs during this period was unknown; see Table B.22).	The following variables pertaining to receipt of treatment only for alcohol were assigned legitimate skip codes: TXYALAGE, TXYALDRG, and TXYALDAG. The skipped variables pertaining to treatment for drugs in the past year, with or without treatment for alcohol (i.e., TXYDRAGE, TXYDRALC, TXYDRAAG, TXYADAAG, and TXYADDAG), retained codes of blank.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
TXYRADG had been assigned a code of 11 (i.e., the R had reported receiving treatment only for alcohol in the past 12 months but some treatment for drugs also was indicated; see Table B.22). TXYALDRG did not indicate that the R had ever received treatment for his or her use of drugs.	TXYALDRG was assigned a code of 3 (i.e., Yes LOGICALLY ASSIGNED).
TXYRADG had been assigned a code of 12 (i.e., the R had reported receiving treatment only for drugs in the past 12 months but some treatment for alcohol also was indicated; see Table B.22). TXYDRALC did not indicate that the R had ever received treatment for his or her use of alcohol.	TXYDRALC was assigned a code of 3 (i.e., Yes LOGICALLY ASSIGNED).
TXYRADG had been set to bad data.	Any nonblank values in variables that were dependent on TXYRADG were set to bad data. This edit was relevant to the following questions (corresponding edited variables shown in parentheses): TX49 (TXYALAGE), TX49A (TXYALDRG), TX49B (TXYALDAG), TX50 (TXYDRAGE), TX50A (TXYDRALC), TX50B (TXYDRAAG), TX51 (TXYADAAG), and TX51A (TXYADDAG).
TXYRADG was answered as "don't know" or "refused."	All of the variables TXYALAGE through TXYADDAG that had been skipped retained codes of blank.

Table B.24 Edits Pertaining to the Perceived Need for Treatment Variables

Issue	Edits Implemented
The only indication(s) of lifetime drug use that routed the respondent (R) into the substance treatment questions had been set to bad data because only over-the-counter (OTC) drug use had been reported in the core.	Nonblank values in the edited variables pertaining to perceived need for substance treatment were replaced with bad data codes.
The R specified the need for treatment for an OTC psychotherapeutic medication (e.g., aspirin).	This information on OTC drugs was not used to infer need for treatment for any of the psychotherapeutic drugs because the questions about perceived need for treatment for psychotherapeutic drugs referred specifically to prescription-type medications (i.e., and not OTCs).
The R did not report needing treatment for a particular drug in the past 12 months, but need for treatment for this drug was specified as a treatment need for "some other drug." In the case of the psychotherapeutics, the other drug specified was not an OTC drug.	The R was inferred to perceive the need for treatment for the use of that drug. For example, Rs who did not report needing treatment for prescription stimulants but reported needing treatment for street stimulants were considered to qualify as perceiving the need for treatment for prescription-type stimulants (i.e., those that were not available over the counter, which would include street drugs). The edited variable NDTXSTMR was assigned a code of 3 (Yes LOGICALLY ASSIGNED). This code of 3 could be edited further, as discussed below.
The R reported needing treatment in the past 12 months for the use of alcohol or other drugs, but questions about the perceived need for treatment for all specific drugs that the R had ever used were answered as "no."	A special code was assigned to the "some other drug" variable (NDTXSOD) to indicate that the specific drug for which the R thought that he or she needed treatment was unknown.
Question TX10, pertaining to the perceived need for additional treatment, is an "enter all that apply" type of question. That is, Rs could report needing additional treatment for more than one drug shown in the list in TX10. However, Rs could report needing additional treatment for drugs that they had reported never using in the corresponding core module (e.g., reported never using heroin but reported needing additional treatment for heroin). In contrast, Rs would not get asked questions TX11 through TX21 (regarding perceived need for treatment for specific drugs) unless they were lifetime users of a particular drug.	No editing was done when this pattern occurred. Consequently, these noncore data would be inconsistent with the core data.

Table B.24 Edits Pertaining to the Perceived Need for Treatment Variables (continued)

Issue	Edits Implemented
The R was logically inferred to have received treatment in the past 12 months (TXYREVER = 3). Because the R did not originally answer question TX02 as "yes," the computer-assisted interviewing (CAI) program routed the Rs to questions about whether they thought they needed treatment for their use of alcohol or specific drugs (i.e., question TX08 and questions TX11 through TX22).	The following edits were done when TXYREVER = 3: • If a question was originally answered as "yes," then the corresponding edited variable was assigned a code of 11 (Yes [TXYREVER = 3]). For example, if the R reported needing treatment for alcohol or other drugs (TX08 = 1), then the edited variable NDTXNEDR was assigned a code of 11. Similarly, if the R reported needing treatment for a specific drug (e.g., prescription stimulants), then the edited variable (e.g., NDTXSTMR) was assigned a code of 11. • If a question was originally answered as "no," then the corresponding edited variable was assigned a code of 12 (No [TXYREVER = 3]). For example, if TX08 had been answered as "no" (TX08 = 2), then NDTXNEDR was assigned a code of 12. (If NDTXNEDR was set to 12, then subsequent variables continued to be assigned legitimate skip codes.) Similarly, if a question about the need for treatment for a specific drug had been answered as" no," then the edited variable was assigned a code of 12. • If the R was inferred to perceive the need for treatment for a drug based on "OTHER, Specify" data, the edited variable was assigned a code of 13. Suppose, for example, that NDTXSTMR had already been coded as 3 because the R had specified prescription-type stimulants as "some other drug" for which the R needed treatment (but question TX19 had not been answered as "yes"). If the R was logically inferred to have received treatment in the past 12 months (TXYREVER = 3), then NDTXSTMR was subsequently coded as 13 (Yes LOGICALLY ASSIGNED [TXYREVER = 3]). • If the R was a lifetime nonuser of a drug, the edits continued to assign a legitimate skip code. For example, if the R had never used prescription-type stimulants, then NDTXSTMR continued to receive a code of 99 when TXYREVER = 3. • Codes for any reasons that respondents reported for why they did not get treatment were bumped by 10; the resulting codes were 11 or 13. The rationale for these edits was that Rs would not have been asked questions about their perceived need for treatment for alcohol or specific other drugs if they had originally reported that they received treatment in the past 12 months. The above edits were done to conserve respondents' answers, as opposed to wiping out the data.
The R reported making an effort to get treatment (question TX22 answered as "yes"), but the R reported not needing treatment for every specific drug that he or she was asked about.	The edited variable NDTXEFTR was assigned a code of 11. The same edits described above for other variables that applied when TXYREVER = 3 also were performed when NDTXEFTR was assigned a code of 11 due to this issue.

Table B.25 Edits Pertaining to the Adult Mental Health Service Utilization Variables

Issue	Edits Implemented
Respondents (Rs) did not choose an outpatient treatment location from the list of locations in question ADMT14, but that location was specified as a source of outpatient mental health treatment in the past 12 months in AUOPYRSP.	The edited variable corresponding to receipt of outpatient treatment at that location was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). For example, if an R did not report receiving outpatient mental health counseling at the office of a private therapist, reported receiving outpatient counseling in "some other place," and specified something to indicate that he or she received counseling from a private therapist, the edited variable AUOPTHER was assigned a code of 3.
Rs reported receiving mental health services in every inpatient or outpatient location in a list.	For inpatient treatment, all of the variables corresponding to the service locations were set to bad data, including the numbers of nights that Rs reported spending at these various inpatient treatment settings. For outpatient treatment, the edits depended on what Rs specified for the "other" outpatient location where they received mental health treatment. • If a valid "other" outpatient location was not specified, the entire series of outpatient variables (including the reported numbers of visits) was set to bad data. • If the R reported a valid "other" outpatient location where he or she received mental health services in the past 12 months, the data were retained to indicate that the R received services in this location. However, the remaining variables pertaining to receipt of outpatient mental health treatment were set to bad data.
Rs reported at least one of the following: (a) they stayed overnight as an inpatient for mental health treatment in a particular type of facility for 365 or 366 days in the past 12 months, or (b) they stayed overnight as an inpatient in more than one type of facility, and the total number of nights that they stayed as inpatients summed to 365 or more.	If Rs reported inpatient treatment in a particular location for 366 days in the past 12 months, the corresponding edited variable (e.g., AUNMPSYH for the number of nights hospitalized in a psychiatric hospital) was reset to 365. No other editing was done when these patterns occurred.
Rs did not choose a payment source for their mental health treatment but subsequently indicated that this was (or would be) the principal payment source.	The edited payment source variable was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). For example, if an R did not report that private health insurance paid or would pay for outpatient treatment but then reported that private insurance was (or would be) the principal source of payment, the edited variable AUPOPINS (private health insurance paid/will pay for any outpatient mental health treatment) was assigned a code of 3.
Rs reported a specific source of payment for their services but also reported that "No one paid because the treatment was free."	No editing was done because these responses were not necessarily inconsistent. Rs could have received services in more than one setting or from more than one provider, with some services being free and other services requiring payment.

Table B.25 Edits Pertaining to the Adult Mental Health Service Utilization Variables (continued)

Issue	Edits Implemented
Rs did not report a specific reason in question ADMT27 or ADMT27A for why they did not receive mental health treatment in the past 12 months, but they specified this as "some other reason."	The edited variable associated with that particular reason for not receiving mental health treatment was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). For example, if Rs specified that they did not get mental health treatment because they could not afford the cost and they had not chosen that reason in question ADMT27, the edited variable AUUNCOST (no mental health treatment because the R could not afford the cost) was assigned a code of 3. Similarly, if Rs specified that they did not get treatment because they thought they could handle the problem without treatment but had not chosen that reason in ADMT27A, the edited variable AUUNHNDL (no mental health treatment because the R thought he or she could handle the problem without treatment) was assigned a code of 3.
Rs did not choose a particular alternative service provider from the list of providers in question ADMT29b, but that provider was specified as a source of alternative mental health treatment or support in the past 12 months.	The edited variable corresponding to receipt of alternative treatment from that type of provider was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). For example, if an R did not report receiving treatment or support from a chiropractor in question ADMT29b, reported receiving treatment from some other provider, and specified something to indicate that he or she received treatment or support from a chiropractor, the edited variable AUALCHIR was assigned a code of 3.
Rs did not report receiving inpatient mental health treatment, or they reported receiving inpatient treatment but not at a location listed in question ADMT02. However, the Rs also specified treatment in a particular inpatient location as some other source of "alternative" treatment in AUALOTSP.	If a specific inpatient treatment location had not been reported in AUINPSYH through AUINRESD but the Rs specified treatment in that location in AUALOTSP, the edited inpatient variable was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from AUALOTSP). In addition, if AUINPYR, pertaining to receipt of any inpatient mental health treatment in the past 12 months, was not answered as "yes," AUINPYR was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED.
Rs did not report receiving outpatient mental health treatment, or they reported receiving outpatient treatment but not at a location listed in question ADMT14. However, the Rs also specified treatment in a particular outpatient location as some other source of "alternative" treatment in AUALOTSP.	If a specific outpatient treatment location had not been reported in AUOPMENT through AUOPDTMT but the Rs specified treatment in that location in AUALOTSP, the edited outpatient variable was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from AUALOTSP). This code of 5 was designed to allow analysts to distinguish between logical inferences based on AUOPYRSP (see above) and those based on AUALOTSP. In contrast, if the R did not report receiving mental health treatment in an outpatient medical clinic, for example, a code of 3 in AUOPCLNC would mean that the R had not reported treatment in an outpatient medical clinic but reported receiving treatment in that location as some other outpatient location, from AUOPYRSP. In addition, if AUOPTYR, pertaining to receipt of any outpatient mental health treatment in the past 12 months, was not answered as "yes," AUOPTYR was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED.

Table B.25 Edits Pertaining to the Adult Mental Health Service Utilization Variables (continued)

Issue	Edits Implemented
Rs did not report taking medication in the past 12 months that was prescribed for a mental health condition. However, the Rs also specified that they took medication as some other source of "alternative" treatment in AUALOTSP.	The edited variable AURXYR, pertaining to taking prescribed medication in the past 12 months, was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED.
Rs did not report receiving treatment from any alternative providers in the past 12 months, or they reported receiving alternative treatment but not from a provider listed in question ADMT29B. However, the Rs also specified treatment from a particular alternative provider as some other source of "outpatient" treatment in AUOPYRSP.	If treatment from a specific alternative provider had not been reported in AUALACUP through AUALMASG but the Rs specified treatment from that location in AUOPYRSP, the edited alternative treatment variable was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from AUOPYRSP). This code of 5 was designed to allow analysts to distinguish between logical inferences based on AUALOTSP (see above) and those based on AUOPYRSP. For example, if the R did not report receiving mental health treatment from a chiropractor but reported receiving outpatient treatment from a chiropractor in AUOPYRSP, the edited variable AUALCHIR (mental health treatment from a chiropractor in the past 12 months) was assigned a code of 5. In comparison, a code of 3 in AUALCHIR would mean that the R had not reported treatment from a chiropractor but reported receiving that treatment from some other provider in AUALOTSP. In addition, if AUALTYR, pertaining to receipt of any alternative mental health treatment in the past 12 months, was not answered as "yes," AUALTYR was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED.

Table B.26 Prioritization of Specific Skip Logic Edits for the Consumption of Alcohol Module

Response Pattern	Edit
Data from the core alcohol module indicated that the respondent (R) never used alcohol.	A code of 91 (or 991, 9991) was assigned to the edited variables in this module. This edit took precedence over situations in which questions may not have been applicable for other reason (e.g., if Rs were male).
Even if the R had used alcohol, data from remaining core modules indicated that the R had never used drugs other than alcohol.	A code of 91 was assigned to the edited variables corresponding to questions CA08 and CA09, pertaining to alcohol use in combination with other drugs in the past 30 days (edited variables CADRKDRG through CADRKSED). The code of 91 had the following meaning in these variables: NEVER USED ALCOHOL OR NEVER USED OTHER DRUGS. Similarly, if the R had not used illicit drugs and the R had been logically inferred to be a nonuser of at least one drug, then these edited variables were assigned a code of 81, where 81 = NEVER USED ALCOHOL OR NEVER USED DRUGS Logically assigned.
Data from the core alcohol module indicated that the R had used alcohol, but definitely not in the past 30 days.	A code of 93 (or 993, 9993) was assigned to the edited variables in this module to indicate that the R did not use alcohol in the past 30 days.
Even if the R had used alcohol in the past 30 days, data from remaining core modules indicated that the R had used some other drugs in the lifetime but not in the past 30 days.	A code of 93 was assigned to the edited variables corresponding to questions CA08 and CA09. The code of 93 had the following meaning in these variables: DID NOT USE ALCOHOL OR DID NOT USE DRUGS IN THE PAST 30 DAYS. In addition, data were overwritten with a code of 83 (DID NOT USE ALCOHOL OR DID NOT USE DRUGS IN THE PAST 30 DAYS Logically assigned) in CADRKDRG (corresponding to question CA08) if the only reports of drug use in the past 30 days had been changed to indicate logical inference of lifetime nonuse because the only drugs that the R had ever reported using in the past 30 days were over-the-counter (OTC) psychotherapeutic drugs.
The R had used alcohol in the past 30 days (or was potentially a past month user, based on ALCREC = 8 from the core alcohol module), but the R was aged 21 or older.	The edited variables corresponding to questions CA02 through CA07SP about underage alcohol use were assigned a code of 99 (LEGITIMATE SKIP) if they were blank or otherwise were assigned a code of 89 (LEGITIMATE SKIP Logically assigned).

Table B.26 Prioritization of Specific Skip Logic Edits for the Consumption of Alcohol Module (continued)

<table>
<tr><th>Response Pattern</th><th>Edit</th></tr>
<tr><td>The R had used alcohol in the past 30 days (or was potentially a past month user) and was aged 12 to 20.</td><td>If the R did not indicate in question CA02B that he or she drank the alcohol in some other place the last time, then the edited variable CADROTHSP (corresponding to question CA02SP) was assigned a legitimate skip code (i.e., 99 if blank or 89 otherwise).
If the R indicated in question CA03 (edited variable CABUYFRE) that the R paid for the last alcohol that he or she drank:
Edited variables corresponding to questions CA07 (edited variable CAFREWHO) and CA07SP (edited variable CAFRESP) were assigned a legitimate skip code.
If the R paid for the last alcoholic beverage that he or she drank but the R gave money to someone else who bought it (question CA04 = 2, corresponding to the edited variable CAGVMONY), then the edited variables corresponding to questions CA05 (CABUYWHO), CA05A (CABPLACE), and CA05B (CABUNDAG) were assigned a legitimate skip code.
If the R paid for the last alcoholic beverage that he or she drank and also bought it (CAGVMONY = 1), then the edited variable corresponding to question CA06 (CAGVWHO) was assigned a legitimate skip code. In addition, if the R reported in CA05 that he or she bought the alcohol in a store, in a restaurant, in a club, or at an event (CA05 = 1), then CABUNDAGE was assigned a legitimate skip code. If the R reported in question CA05 that he or she bought the alcohol from another person (CA05 = 2), then CABPLACE was assigned a legitimate skip code.
If the R indicated in question CA03 that he or she did not pay for the last alcohol that he or she drank:
CAGVMONY, CABUYWHO, CABPLACE, CABUNDAG, and CAGVWHO were assigned a legitimate skip code.
If the R did not report in question CA07 that he or she got the last alcoholic beverage some other way, then the edited variable CAFRESP (corresponding to CA07SP) was assigned a legitimate skip code.</td></tr>
</table>

Table B.26 Prioritization of Specific Skip Logic Edits for the Consumption of Alcohol Module (continued)

Response Pattern	Edit
The R had used alcohol in the past 30 days and also had used other drugs in that period.	• If the R reported in question CA08 (edited variable CADRKDRG) that he or she did not use other drugs at the same time or within a couple of hours of the last time when the R drank alcohol in the past 30 days, then the edited variables corresponding to the drugs in question CA09 (i.e., edited variables CADRKMRJ through CADRKSED) were assigned legitimate skip codes. • Otherwise (e.g., if CA08 was answered as "yes"), if the core data indicated that the R had not used a particular drug in the past 30 days but the R had used some other drug in the past 30 days, then the variable corresponding to the drug that the R had not used was assigned a legitimate skip code. Suppose, for example, that the R used other drugs at the same time or within a couple of hours of when the R last used alcohol and the R had used marijuana but not cocaine in the past 30 days. In this situation, CADRKCOC (i.e., use of cocaine in combination with alcohol) was assigned a legitimate skip code.
The R reported using other drugs at the same time or within a couple of hours of when he or she last used alcohol. However, question CA09 about use of alcohol in combination with specific drugs was skipped because the R reported use of only one illicit drug in the past 30 days.	The edited variable corresponding to the one illicit drug that the R had used in the past 30 days was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from skip pattern). For all other drugs that the R had definitely not used in the past 30 days, the edited variables were assigned legitimate skip codes. For example, if the R answered question CA08 as "yes," and the only drug that the R reported using in the past 30 days was marijuana, then it could be logically inferred that the R used marijuana at the same time or within a couple of hours of when he or she last used alcohol, and the edited variable CADRKMRJ was assigned a code of 5.
The R was a lifetime alcohol user but was male.	The edited variables corresponding to questions CA12 through CA14d about females' consumption of four or more drinks on a single occasion were assigned a code of 99 (or 999, 9999) if they were blank or otherwise were assigned a code of 89 (or 989, 9989).
The lifetime binge alcohol question CA10 (for consumption of five or more drinks on a single occasion) had been skipped because either of the following occurred: • The core data indicated some consumption of five or more drinks on a single occasion in the past 30 days. • Question CA01 (edited variable CADRLAST) indicated that the R had five or more drinks the last time that he or she drank in the past 30 days.	The edited variable CABNGEV (corresponding to question CA10) was set to a value of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).
Question CA10 was answered as "no," indicating that the R had never had five or more drinks on a single occasion.	The edited variables CABNGAGE (corresponding to question CA11), CABNGYFU (derived from questions CA11a through CA11c), and CABNGMFU (derived from questions CA11c and CA11d) were assigned legitimate skip codes.

Table B.26 Prioritization of Specific Skip Logic Edits for the Consumption of Alcohol Module (continued)

Response Pattern	Edit
The lifetime binge alcohol question CA12 for females (for consumption of four or more drinks on a single occasion) had been skipped because any of the following occurred: • The core data indicated that the R usually had four or more drinks on those days when she drank in the past 30 days. • CADRLAST indicated that the R had four or more drinks the last time she drank in the past 30 days. • Question CA10 (edited variable CABNGEV) was answered as "yes," indicating that the R had consumed five or more drinks on a single occasion at least once in her lifetime.	The edited variable CA4FDEV (corresponding to question CA12) was set to a value of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).
For female Rs, question CA12 was answered as "no," indicating that these Rs had never had four or more drinks on a single occasion. There also was no indication from CA10 that the R had ever had five or more drinks on a single occasion.	• The edited variables CA4FDAGE (corresponding to question CA14), CA4FDYFU (derived from questions CA14a through CA14C), and CA4FDMFU (derived from questions CA14C and CA14D) were assigned legitimate skip codes. • If the R had used alcohol in the past 30 days (or the most recent period of alcohol use was ambiguous and could have included the past 30 days), and there was no other indication that she had ever had four or more drinks on a single occasion, then the edited variable CA4FDDYS (corresponding to question CA13) was assigned a legitimate skip code. Otherwise, if the R was a lifetime alcohol user but definitely had not used alcohol in the past 30 days, then the edits described above took precedence, and CA4FDDYS was assigned a code of 93.

Table B.27 Edits in the Consumption of Alcohol Module Pertaining to the Binge Alcohol Use Variables

Issue	Edits Implemented
The respondent (R) indicated in the final verification check for the age when he or she first engaged in binge alcohol use (BACC06 for five or more drinks and WBACC06 for four or more drinks) that the age at first binge alcohol use based on the new month and year of first use (MFU and YFU) was correct (i.e., BACC06 = 4 or WBACC06 = 4). The computer-assisted interviewing (CAI) program updated the age at first use (AFU; AGE1STBA or AGE1WBA for female respondents) with the value of the age calculated from the MFU and YFU (i.e., MYR1STBA or MYR1WBA). However, the new value for the AFU indicated that the R was more than 1 year younger than his or her current age at the time the R first engaged in binge alcohol use. Had the R initially reported this AFU, the R would not have been routed to the relevant MFU and YFU questions.	The updated value was retained in the edited AFU variables CABNGAGE or CA4FDAGE. Based on the updated value for the AFU, it was logically inferred that the R should have skipped the MFU and YFU items. A code of 9989 was assigned to the relevant YFU variable (CABNGYFU or CA4FDFYU), and a code of 89 was assigned to the corresponding MFU variable (CABNGMFU or CA4FDMFU).
The final verification check (BACC06 or WBACC06) was skipped because the R entered revised data for the relevant MFU and YFU that made MFU and YFU for first binge alcohol use consistent with the AFU.	No editing was done because the R was considered to have resolved the inconsistency.
The final verification check (BACC06 or WBACC06) was skipped because the R entered a new MFU that was the same as the R's birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, the revised MFU and YFU were set to bad data if they could never be consistent with the AFU.
The R entered a new MFU or YFU that differed from what the R previously reported. The age based on the revised MFU and YFU (updated in MYR1STBA or MYR1WBA) still mismatched the AFU, but the R indicated in the final verification check that the new value from the relevant MYR variable was correct.	No editing was done in this situation. The CAI program automatically updated the value for the AFU to be consistent with the updated values reported for the MFU and YFU.
The consistency check was triggered between the "AGE1" AFU variable stored by the CAI program (e.g., AGE1STBA for consumption of five or more drinks) and the corresponding MYR variable (e.g., MYR1STBA). However, the R answered the first consistency check (e.g., BACC03, regarding whether the value in MYR1STBA was correct) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The AFU value that had been stored in the AGE1 variable (e.g., in AGE1STBA) was retained, but the corresponding MFU and YFU variables were set to bad data.
The consistency check was triggered between the AGE1 and corresponding MYR variables. However, the R entered the same values for the YFU and MFU that triggered the inconsistency with the AFU in the first place.	No editing was done to the AFU, but the YFU and MFU variables were set to bad data.

Table B.27 Edits in the Consumption of Alcohol Module Pertaining to the Binge Alcohol Use Variables (continued)

Issue	Edits Implemented
The R entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for the MYR variable. However, the R failed to resolve the inconsistency between the AGE1 variable and the updated value in the MYR variable. The R also reported either in the first verification check that the MYR was not correct (e.g., BACC03 = 6) or reported in the second verification check that the AFU from the AGE1 variable was correct (e.g., BACC04 = 2).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: • The default edit was to set the MFU and YFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the MFU and YFU were assigned the code that corresponded to the answer in the final verification check.
The R entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for the MYR. However, the R failed to resolve the inconsistency between the AGE1 variable and the updated value in the MYR variable. The R also reported in the second verification check that neither answer was correct for what was originally captured in the AGE1 and MYR variables (e.g., i.e., BACC04 = 3).	The following edits were implemented for the AFU, MFU, and YFU: • The default edit was to set the AFU, MFU, and YFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the AFU, MFU, and YFU were assigned the code that corresponded to the answer in the final verification check.
The AFU was inconsistent with the value from the MYR (based on the MFU and YFU). The R also reported that the age from the MYR was correct (e.g., BACC03 = 4). However, the R answered the second consistency check (BACC04) as "don't know" or "refused." Consequently, the R did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for "don't know" or "refused" were assigned to the AFU, MFU, and YFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R's correct age when he or she first engaged in binge alcohol use, or whether the MFU and YFU indicated the R's correct age at initiation of binge alcohol use. Therefore, the AFU, MFU, and YFU all were set to missing values.
The R reported first binge alcohol use at an age that was earlier than what was reported for his or her first use of any alcohol or at an age that was greater than the R's current age.	The edited AFU variable (e.g., CABNGAGE) was set to bad data. Any nonmissing values in the YFU and MFU also were set to bad data.
The R reported first binge alcohol use in a year that was earlier than what the R reported for the year in which he or she first used any alcohol.	The edited YFU variable was set to bad data. Any nonmissing values in the MFU also were set to bad data.
The R reported first binge alcohol use in the same year in which the R first used any alcohol. However, the R reported first binge alcohol use in a month that was earlier than the month in which the R first used any alcohol.	The edited MFU variable was set to bad data.
The R reported first binge alcohol use in the same month as his or her birth month. However, the R could never have first engaged in binge alcohol use at the age he or she reported and also have first engaged in binge alcohol use in the year that the R reported.	The relevant YFU and MFU variables were set to bad data.

Table B.27 Edits in the Consumption of Alcohol Module Pertaining to the Binge Alcohol Use Variables (continued)

Issue	Edits Implemented
An edited AFU variable for binge alcohol use had a missing value. However, the R first used alcohol at his or her current age.	• The binge alcohol AFU variable with the missing value was set to the R's current age. • The corresponding editing indicator (EI) variable was set to a value of 2, where 2 = Logically assigned data. • If the YFU and MFU had been set to bad data but the unedited values that the R reported for the YFU and MFU were the same as those in ALCYFU and ALCMFU, then the relevant binge YFU and MFU variables were reset to their unedited values.
An edited YFU variable for binge alcohol use had a missing value. However, the R first used alcohol in the current survey year.	• The binge alcohol YFU variable with the missing value was set to the current year. • If the edited YFU was not the same as the unedited value, then the corresponding EI variable was set to a value of 2, where 2 = Logically assigned data. • If the MFU had a missing value and ALCMFU indicated that the R first used alcohol in the interview month, then the binge MFU was set to the same value as ALCMFU. If the edited MFU value was not the same as the unedited value, then the corresponding EI variable for the MFU (e.g., EIBNGMFU) was set to a value of 2.
An edited MFU variable for binge alcohol use had a missing value. However, the YFU data and interview date supplied information on when the R logically first engaged in binge alcohol use.	• The missing value in the MFU was replaced with the definite month in which the R logically first engaged in binge alcohol use. For example, if the R reported first use of any alcohol and first binge alcohol use in the same year and the R first used any alcohol in December of that year, then December logically was the only month in that year in which the R could have initiated binge alcohol use. • If the edited MFU value was not the same as the unedited value, then the corresponding EI variable for the MFU (e.g., EIBNGMFU) was set to a value of 2.
For females, the R reported first consumption of five or more drinks on a single occasion that was earlier than the age she reported for first consumption of four or more drinks.	• The earlier AFU from CABNGAGE was assigned to CA4FDAGE, and EI4FDAGE was set to a value of 2. • If CABNGYFU had a nonmissing value, then data from CABNGYFU and CABNGMFU also were transferred to CA4FDYFU and CA4FDMFU. • If the edited values for CA4FDYFU and CA4FDMFU differed from the corresponding unedited values, then EI4FDYFU and EI4FDMFU were assigned a value of 2.

Table B.27 Edits in the Consumption of Alcohol Module Pertaining to the Binge Alcohol Use Variables (continued)

Issue	Edits Implemented
For females, the R reported first consumption of five or more drinks on a single occasion in a year that was earlier than the year she reported for first consumption of four or more drinks.	• The earlier YFU from CABNGYFU was assigned to CA4FDYFU, and EI4FDYFU was set to a value of 2. • Data from CABNGMFU also were transferred to CA4FDMFU. • If the edited value for CA4FDMFU differed from the corresponding unedited value, then EI4FDMFU was assigned a value of 2.
For females, the R reported first consumption of five or more drinks on a single occasion and four or more drinks on a single occasion in the same year. However, the R reported first consumption of five or more drinks in a month that was earlier than what she reported for first consumption of four or more drinks.	• Data from CABNGMFU were transferred to CA4FDMFU. • If the edited value for CA4FDMFU differed from the corresponding unedited value, then EI4FDMFU was assigned a value of 2.
For females, CA4FDDYS (corresponding to question CA13) had a missing value. However, the edited variable DR5DAY from the core alcohol module indicated that these Rs had five or more drinks on a single occasion on all 30 days in the past month.	CA4FDDYS was assigned a value of 30, and EI4FDDYS was assigned a value of 2.
For females, CA4FDDYS had a nonmissing value that was less than the number of days from DR5DAY.	CA4FDDYS was assigned the value from DR5DAY, and EI4FDDYS was assigned a value of 2.
For females, the R reported that she had four or more drinks on a single occasion on 0 days in the past 30 days. However, CADRLAST or the core alcohol data indicated some consumption of four or more drinks in that period.	CA4FDDYS was set to bad data.

Appendix C: Race and Hispanic/Latino Group Alpha Codes

This page intentionally left blank

Appendix C: Race and Hispanic/Latino Group Alpha Codes

C.1 Introduction

For the 2014 National Survey on Drug Use and Health (NSDUH), it was not uncommon for a respondent to feel that the categories for race or Hispanicity given in the questionnaire did not apply to him or her. In these situations, interviewers were given the opportunity to manually enter (type) a category that the respondent felt best described him or her. The manually entered responses were called "OTHER, specify" or "alpha-specify" responses because they were typed in a part of the question that asked the interviewer to specify an alphabetic response. These alpha-specify responses were then matched to codes to describe the responses, which were collected and maintained in a file known as a "dictionary." Other-specify responses from each survey year were matched against this file, and any responses without codes were given new codes and added to the dictionary. Consequently, the size of the dictionary file has increased each survey year.

In most cases, new unmatched responses were just new misspellings of an already established category, such as a response of "Porto Rican" instead of "Puerto Rican." If an interviewer entered both a geographical entity and a race in the "OTHER, specify" response, such as "Japanese Peruvian," the geographical entity was ignored and the respondent was coded as "Japanese." The geographical entity was recorded only if no other information was available, either in the "OTHER, specify" response or in the non-OTHER, specify response. As discussed in Chapter 4, many respondents provided a race in the alpha-specify response to the Hispanic/Latino group question and vice versa, so responses to both questions were examined in the creation of each variable. This appendix summarizes the procedures that were implemented to assign race and Hispanic/Latino values to respondents based on alpha-specify responses to the questionnaire.

Once a racial category was selected that represented the "OTHER, specify" response, this was combined with information that was provided in the non-OTHER, specify categories. If the information provided in the "OTHER, specify" response was so general that formal imputation seemed to be required, and more specific information was available in the non-OTHER, specify categories, then the final assignment of a racial category was made using only the information from the non-OTHER, specify category(ies) and the "OTHER, specify" information was ignored.

C.2 Race

In the 2014 questionnaire, two core questions (QD05 and QD05ASIA) focused on the respondent's race. Respondents were permitted to select more than one race in QD05. If they selected "Asian" as one of their races, they were routed to QD05ASIA, where they were permitted to select more than one answer. Respondents had the opportunity to direct the interviewer to select "Other" as the race in both QD05 and (if applicable) QD05ASIA, whereby the interviewer then typed the alphabetic response given by the respondent. The alpha-specify responses to these two questions were considered together. Two new levels, Guamanian or Chamorro and Samoan, were added to QD05 in 2013. The race questions used in the 2014 survey were as follows:

QD05: Which of these groups describes you?

1	White
2	Black/African American
3	American Indian/Alaska Native (American Indian includes North American, Central American, and South American Indians)
4	Native Hawaiian
5	Guamanian or Chamorro
6	Samoan
7	Other Pacific Islander
8	Asian (Including: Asian Indian, Chinese, Filipino, Japanese, Korean and Vietnamese)
9	Other (Specify)

QD05ASIA: (Asked only if level 8 of QD05 was selected.) Which of these Asian groups describes you?

1	Asian Indian
2	Chinese
3	Filipino
4	Japanese
5	Korean
6	Vietnamese
7	Other (Specify)

The Hispanic/Latino group question (QD04), discussed in Section C.3, also has an "OTHER, specify" response. Whenever race information was not available from QD05 or QD05ASIA, the response to QD04 was examined to determine whether any race information was available.

C.2.1 Race Alpha Responses

The four types of race "OTHER, specify" responses are described in Chapter 4. Abbreviated descriptions are repeated here for convenience.

Directly Mapped Codes. Directly mapped codes were codes mapped to one or more of the categories given in the questionnaire. There were two types of directly mapped codes: (1) racial category codes and (2) geographic category codes. Racial category codes were exactly equivalent to one or more categories in QD05 or QD05ASIA. For example, a response such as "Han" mapped directly to a category in QD05ASIA (Chinese), and a response such as "mestizo" mapped directly to two categories in QD05 (white and American Indian/Alaska Native). Geographic category codes corresponded to a country where census data indicated a racially homogeneous society. For example, an entry of "Polish" mapped to white because the Polish census data indicated that nearly all Poles were white.

Indirectly Mapped Codes. Codes that were indirectly mapped also corresponded to countries where census data were used, but for indirect mapping, the countries were racially heterogeneous. A racial category from among the 11 categories given in the questionnaire was chosen by generating a random number and allocating the race based on a comparison of the random number with the proportions of races in the geographical entity's (country's) census. For example, an entry of "St. Thomas" would have a 76.2 percent chance of being allocated to the black/African-American category, because the latest census indicated that 76.2 percent of individuals from the U.S. Virgin Islands were black. If two or three heterogeneous countries were entered in the "OTHER, specify" response, the final race was allocated using the following procedure: (1) randomly assign races based on the proportions for each country mentioned, and (2) combine the results. Exceptions to these rules occurred with the categories Mexican, Puerto Rican, Cuban, Central or South American, Dominican, and Spanish (from Spain).

Starting with the 2006 imputation process, the handling of indirectly mapped codes obtained from QD05ASIA has been simplified. In earlier survey years, these types of write-in responses were mapped to a race through country census information. Since the 2006 NSDUH, however, all census-based write-in responses to the Asian race question were mapped directly to the "Other Asian" racial category.

Informative Codes for Formal Imputation Procedures. Some "OTHER, specify" responses did not lead to definitive information about the respondent's race but were used to limit the final imputation. With these informative codes, the final imputation was restricted according to the information that was available. No imputation was required if more specific information was available from responses to the non-OTHER, specify categories. For example, a response of "mixed" resulted in an imputation among donors with two or more races, and a response of "brown" resulted in an imputation among donors who were not single-race white.

Noninformative Codes. Finally, a noninformative response that was not accompanied by a response to one of the given (non-OTHER, specify) categories resulted in an unrestricted imputation. Religious identifications (e.g., Muslim) were considered noninformative, even if the religion was usually associated with a particular ethnic group (e.g., Shinto is usually associated with Japanese).

Table C.1 lists all the race codes used in the 2014 survey, along with supplementary information related to race codes. Special situations associated with the four types of codes described in this section are discussed in the following sections. For most codes, the final assignment depended upon whether the response was given in QD05 or QD05ASIA. For informative codes described above, the six Hispanic/Latino codes—Mexican, Puerto Rican, Cuban, Central or South American, Dominican, and Spanish (from Spain)—were treated differently depending upon whether they were listed in conjunction with other racial or geographical entities.

Codes with an asterisk were those that caused the Hispanic/Latino indicator to be edited to a "yes." That is, if QD03 was either missing or "no," and any of these codes appeared as an "OTHER, specify" response to QD05 or QD05ASIA, the edited Hispanic/Latino indicator (EDHOIND) was set to 1 to denote "Hispanic/Latino," and the imputation indicator for the Hispanic/Latino indicator (IIHOIND) was set to 2 to indicate "logically assigned." See Chapter 4 for more details on the edited Hispanic/Latino indicator. Note that EDHOIND also could be edited to a "no." This is discussed in Section C.3.1.

C.2.1.1 Handling of Directly Mapped Codes

For codes that were directly mapped, the final column of Table C.1 indicates to which race the code was mapped. With some exceptions, the handling of directly mapped codes that were racial categories or Asian geographic categories did not depend upon whether the response was observed in QD05 or QD05ASIA. The exceptions to this rule occurred if the response included a reference to "Indian," which was mapped to "American Indian/Alaska Native" if the response was given in QD05 and to "Asian Indian" if the response was given in QD05ASIA. On the other hand, for directly mapped codes that were non-Asian geographic categories, the final mapping always depended upon whether the response was observed in QD05 or QD05ASIA. In this case, if the code was observed in QD05ASIA, the code was always mapped to "Other Asian."

Most of the directly mapped cases were mapped directly to a single category regardless of whether the response was in QD05 or QD05ASIA. However, sometimes the category to which the code was mapped in these cases is indicated only for QD05 in the final column in Table C.1. In these instances, it was assumed that the directly mapped code for QD05ASIA was "Other Asian" (this is not shown in the table for space-saving reasons). For codes that corresponded to multiple-race respondents, individual Asian categories were not tracked.

In general, if the respondent selected one or more non-OTHER, specify categories in QD05 and/or QD05ASIA, racial category codes were recognized, but geographic category codes were ignored. This is the primary difference in the handling of the two types of directly mapped codes. For example, if the interviewer selected the category for "black/African American" for the respondent and also wrote in "Polish," it was assumed that the respondent was a black Pole, and for racial identification purposes, the respondent was considered single-race black/African American. This was true even though the Polish census did not identify significant numbers of nonwhite people in the Polish population.

C.2.1.2 Handling of Indirectly Mapped Codes

In most cases, indirectly mapped codes refer to heterogeneous countries where census data were used. In these cases, the race was assigned by comparing a randomly generated number to the proportion of each racial category in that country's census. As with the directly mapped codes, the final mapping of the indirectly mapped codes also depended upon whether the response was in QD05 or QD05ASIA, unless the heterogeneous countries listed were all Asian. In a similar manner to the directly mapped QD05 cases, if the code was observed in QD05ASIA, it was mapped to "Other Asian," provided none of the entries observed were Asian racial categories, Asian countries, or countries with an Asian minority. (Codes that were indirectly mapped if the response was in QD05, but were directly mapped to "Other Asian" if the response was in QD05ASIA, are denoted by "QD05ASIA: O.A." in the fourth column of the table.) Codes where there was at least one Asian minority in a specified heterogeneous country that was not all Asian, and the response was given in QD05ASIA, were handled on a case-by-case basis. The resulting strategy was either a different indirect mapping than that given if the response was in QD05 or a direct mapping.

When census data were used, it was not uncommon to find that a small proportion of the population was identified as "Other." In the rare instance that the randomly generated number indicated the respondent belonged to this "Other" group, then the selected race was determined by imputation. Codes where this was possible are identified with a superscript "I" in the third column of Table C.1. Rather than an "Other" indication, the census sometimes gave general information (e.g., Asian) where more specific information needed to be determined through imputation. In the case where the imputation was limited to Asian categories, the superscript "IA" was used.

Generally, if two entries (countries or racial categories) were observed, the race for each entity was determined first (either through a direct map or a random assignment using census data), and then the two races were combined. In some cases, a racial category was listed along with a geographical entity. As stated earlier, in most cases the geographical entity was ignored because it was usually assumed that the respondent was a resident of the listed country who also happened to be identified with the given racial category. However, it became clear on occasion that the respondent had parentage that belonged to the racial category and different parentage that came from the listed country. In these instances, the racial category was treated in the same manner as a homogeneous country of that race, and the determination of a final race was conducted in the same manner as if two countries had been listed. If one of the races listed was an Asian racial category, for example, then the response was treated in the same manner whether it was observed in QD05 or QD05ASIA. If the final assignment depended upon the census data of two indirectly mapped codes or an indirectly mapped code and a racial category, "double census" is parenthetically indicated in the third column of Table C.1. If three indirectly mapped codes were indicated by the respondent, "triple census" is indicated.[1]

Details about how to handle census information for each indirectly mapped code are shown in Table C.2. Note that the racial categories for each country listed in Table C.2 have been modified to conform to the racial categories specified by the questionnaire. For example, the black racial category from other countries has been modified to the black/African-American category. Every category and restricted imputation level with a nonzero probability of selection is listed. If a code had an indirect map (using census data) for QD05, but had a direct map for QD05ASIA, this is not specified in Table C.2. Instead, this information must be obtained from Table C.1. Explanations of the categories that are not self-explanatory are listed below.

- White or Mestizo: Imputation was restricted to respondents who were either white or Mestizo (i.e., white and American Indian/Alaska Native only). See Chapter 4 for the explanation of level 18 of EDRACE.
- Not American Indian: Imputation was restricted to respondents who were a single race other than American Indian/Alaska Native or were multiple race and American

---

[1] When an indirectly mapped code with superscript I or IA appeared as a component in a double census or triple census code, the probability associated with the "other" category was distributed among the races appearing in the census. This was the simplest way to preserve race information from all the component indirectly mapped codes. A more complicated alternative would be to impute race information for each component country, even if the "other" category was selected at random for one or more of the race categories in the census. See the entry for "Costa Rica" (code 78) in Table C.2 for an example.

Indian/Alaska Native was not one of their component races. See Chapter 4 for the explanation of level 19 of EDRACE.

- Multiple: Imputation was restricted to respondents who were multiple race. See Chapter 4 for the explanation of level 16 of EDRACE.

C.2.1.3 Handling of Codes Informative for Formal Imputation Procedures

For six Hispanic/Latino codes that were highly prevalent in the data, census data were not used to assign the final racial category. (These are the six categories listed in QD04.) Instead, the final racial category for respondents who responded "Mexican," "Puerto Rican," "Central or South American," "Cuban," "Dominican," or "Spanish" was determined by a restricted imputation with donors who indicated one of these categories in QD04. Furthermore, if a respondent indicated any combination of these six categories, the final racial category was determined using a restricted imputation with donors who were from the geographical entities listed. On the other hand, if any of these six Hispanic/Latino groups was listed along with a second code that was not among these six, census data were used along with the census data from the second country listed. More details about how specific restricted imputations are conducted are shown in Table C.3.

If the code was observed in QD05ASIA, then the imputation was not only restricted by what was written in the "OTHER, specify" response but also by the Asian categories that had the necessary attributes. Again, the information was ignored if more specific information was available from responses to the non-OTHER, specify categories.

C.2.1.4 Noninformative Codes

For noninformative codes, a final race could still have been assigned based on responses to other categories in QD05. If no other categories were listed in QD05, race was imputed, where the imputation was restricted to a Hispanic/Latino group if the respondent gave Hispanic/Latino information in QD04. Otherwise, the final race was determined through an unrestricted imputation.

C.3 Hispanicity

As with the race questions, Hispanic/Latino respondents[2] had the opportunity to specify a Hispanic/Latino group by giving the response "Other" to QD04, the Hispanic/Latino group question. Also, respondents were permitted to select multiple Hispanic/Latino groups in response to QD04. Below is the Hispanic/Latino group question.

QD04: Which of these Hispanic, Latino, or Spanish groups best describes you?

1 Mexican/Mexican American/Mexicano/Chicano
2 Puerto Rican
3 Central or South American

[2] For the purposes of the instrument question-routing, Hispanic/Latino respondents were identified by their response to question QD03: "Are you of Hispanic, Latino, or Spanish origin or descent?"

4 Cuban/Cuban American
5 Dominican (from Dominican Republic)
6 Spanish (from Spain)
7 Other (Specify)

Levels 5 and 6 were added to QD04 after the 2004 survey. They were included because there was a large number of "OTHER, specify" responses for these categories in previous years.

The QD05 and QD05ASIA questions are discussed in Section C.2. They also have "OTHER, specify" responses, which were gleaned for Hispanic/Latino group information whenever no Hispanic/Latino group information was available from QD04.

C.3.1 Hispanic/Latino Group Alpha Responses

There were only two types of Hispanic/Latino group "OTHER, specify" responses: (1) those that mapped to one or more EDQD04*xx*[3] variables and (2) those that were ignored. There were no census-based routines and no responses that caused the imputation to be restricted. The imputation of a Hispanic/Latino group was restricted only when race information was available.

Table C.4 lists all the Hispanic/Latino group codes used in the 2014 survey and the Hispanic/Latino groups to which they mapped. Note that these mappings utilized the arbitrary priority rule provided in Chapter 4. This rule was used to create EDHOGRP, which skipped the intermediate step of recording the Hispanic/Latino groups indicated in QD04. These are recorded in the EDQD04*xx* variables, which are described in Chapter 4, along with the creation of EDHOGRP. The Hispanic/Latino code 600, "Stated Clearly as Not Hispanic/Latino," was unique in that it could be used to edit the Hispanic/Latino indicator, if needed. If QD03 was missing or 1, then EDHOIND was edited to a 2 if this code appeared in QD04, QD05, or QD05ASIA.

[3] See Chapter 4 for a discussion of EDQD04*xx*.

Table C.1 **Descriptions of Race Codes and the Categories to Which They Mapped**

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
21	White	Directly mapped (racial category)	White
22	Black/African American	Directly mapped (racial category)	Black/African American
23	American Indian/Alaska Native	Directly mapped (racial category)	American Indian/Alaska Native
24	Native Hawaiian	Directly mapped (racial category)	Native Hawaiian
25	Other Pacific Islander	Directly mapped (racial category)	Other Pacific Islander
26	Asian Indian	Directly mapped (racial category)	Asian Indian
27	Chinese	Directly mapped (racial category)	Chinese
28	Filipino	Directly mapped (racial category)	Filipino
29	Japanese	Directly mapped (racial category)	Japanese
30	Korean	Directly mapped (racial category)	Korean
31	Vietnamese	Directly mapped (racial category)	Vietnamese
32	Other Asian	Directly mapped (racial category)	Other Asian
33	Asian (nonspecific)	Codes informative for formal imputation procedures	Not a Direct Map
34	Guamanian	Directly mapped (geographic category)	Other Pacific Islander
35	Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: American Indian/Alaska Native QD05ASIA: Asian Indian
50	Belize	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.[2]
51	Guyana	QD05: Indirectly mapped[I] QD05ASIA: Directly mapped (geographic category)	QD05ASIA: Asian Indian
52	Suriname	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
53	Haiti	Indirectly mapped (QD05)	QD05ASIA: O.A.
54	Trinidad and Tobago	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)	Not a Direct Map
55	Jamaica	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
56	Virgin Islands (St. Thomas, St. Croix)	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[IA]	Not a Direct Map
57	Bahamas	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[IA]	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
58	Barbados	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
59	Grenada	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
60	St. Lucia	Indirectly mapped (QD05)	QD05ASIA: O.A.
61	St. Vincent & the Grenadines	Directly mapped (geographic category)	Black/African American
62	Dominica	Directly mapped (geographic category)	Black/African American
63	Other West Indies	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)	Not a Direct Map
64	Brazil	Indirectly mapped (QD05)	QD05ASIA: Japanese
65	Canada	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	QD05ASIA: O.A.
66	Bahamas & Haiti	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA) (double census)[IA]	QD05ASIA: O.A.
67	Brazil & Portugal	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA) (double census)[IA]	QD05ASIA: O.A.
70	Mexico	Codes informative for formal imputation procedures	QD05ASIA: O.A.
71	Puerto Rico	Codes informative for formal imputation procedures	QD05ASIA: O.A.
72	Cuba	Codes informative for formal imputation procedures	QD05ASIA: O.A.
73	Dominican Republic	Codes informative for formal imputation procedures	QD05ASIA: O.A.
74	Guatemala	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
75	Honduras	Indirectly mapped (QD05)	QD05ASIA: O.A.
76	El Salvador	Indirectly mapped (QD05)	QD05ASIA: O.A.
77	Nicaragua	Indirectly mapped (QD05)	QD05ASIA: O.A.
78	Costa Rica	Indirectly mapped (QD05)[I]	QD05ASIA: Chinese
79	Panama	Indirectly mapped (QD05)	QD05ASIA: O.A.
80	Colombia	Indirectly mapped (QD05)	QD05ASIA: O.A.
81	Venezuela	Indirectly mapped (QD05)	QD05ASIA: O.A.
82	Ecuador	Indirectly mapped (QD05)	QD05ASIA: O.A.
83	Peru	Indirectly mapped (QD05)	QD05ASIA: Japanese
84	Bolivia	Indirectly mapped (QD05)	QD05ASIA: O.A.
85	Chile	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
86	Argentina	Indirectly mapped (QD05)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
87	Paraguay	Indirectly mapped (QD05)	QD05ASIA: O.A.
88	Uruguay	Indirectly mapped (QD05)	QD05ASIA: O.A.
89	Mexico & Puerto Rico	Codes informative for formal imputation procedures	Not a Direct Map
90	Mexico & Cuba	Codes informative for formal imputation procedures	Not a Direct Map
91	Mexico & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
92	Mexico & Spain	Codes informative for formal imputation procedures	Not a Direct Map
93	Puerto Rico & Cuba	Codes informative for formal imputation procedures	Not a Direct Map
94	Puerto Rico & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
95	Puerto Rico & Spain	Codes informative for formal imputation procedures	Not a Direct Map
96	Cuban & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
97	Cuban & Spain	Codes informative for formal imputation procedures	Not a Direct Map
98	Dominican & Spain	Codes informative for formal imputation procedures	Not a Direct Map
100	Norway	Directly mapped (geographic category)	QD05: White
101	Sweden	Directly mapped (geographic category)	QD05: White
102	Denmark	Directly mapped (geographic category)	QD05: White
103	United Kingdom	Indirectly mapped (QD05)	QD05ASIA: Asian Indian
104	Ireland	Directly mapped (geographic category)	QD05: White
105	Portugal	Directly mapped (geographic category)	QD05: White
106	Spain	Codes informative for formal imputation procedures	QD05: White
107	Germany	Directly mapped (geographic category)	QD05: White
108	France	Directly mapped (geographic category)	QD05: White
109	Italy	Directly mapped (geographic category)	QD05: White
110	Netherlands	Directly mapped (geographic category)	QD05: White
111	Belgium	Directly mapped (geographic category)	QD05: White
112	Greece	Directly mapped (geographic category)	QD05: White
113	Russia	Directly mapped (geographic category)	QD05: White
114	Ukraine	Directly mapped (geographic category)	QD05: White
115	Turkey	Directly mapped (geographic category)	QD05: White

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
116	Other Western Europe	Directly mapped (geographic category)	QD05: White
117	Other Eastern Europe	Directly mapped (geographic category)	QD05: White
118	Other Southern Europe	Directly mapped (geographic category)	QD05: White
119	Morocco	Directly mapped (geographic category)	QD05: White
120	Algeria	Directly mapped (geographic category)	QD05: White
121	Tunisia	Directly mapped (geographic category)	QD05: White
122	Libya	Directly mapped (geographic category)	QD05: White
123	Egypt	Directly mapped (geographic category)	QD05: White
124	Other North Africa	Directly mapped (geographic category)	QD05: White
125	Saudi Arabia	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)	Not a Direct Map
126	Yemen	Directly mapped (geographic category)	QD05: White
127	Oman	Directly mapped (geographic category)	QD05: White
128	UAE	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
129	Qatar	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
130	Bahrain	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[IA]	Not a Direct Map
131	Israel	Directly mapped (geographic category)	QD05: White QD05ASIA: O.A.
132	Iraq	Directly mapped (geographic category)	QD05: White QD05ASIA: O.A.
133	Kuwait	QD05: Directly mapped (geographic category) QD05ASIA: Indirectly mapped	QD05: White
134	Iran	Directly mapped (geographic category)	Other Asian
135	Other Middle East	Directly mapped (geographic category)	QD05: White QD05ASIA: O.A.
136	Armenia	Directly mapped (geographic category)	QD05: White
137	Georgia	Directly mapped (geographic category)	QD05: White
138	Azerbaijan	Directly mapped (geographic category)	QD05: White

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
139	Russia Asian people groups (Tatar, Chechen, Dagestan, etc.)	Directly mapped (racial category)	Other Asian
140	Kazakhstan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
141	Uzbekistan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
142	Tadjikistan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
143	Kyrgizstan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
144	Turkmenistan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
145	Other Central Asia (includes Afghanistan)	Directly mapped (geographic category)	Other Asian
150	Sri Lanka	Directly mapped (geographic category)	Other Asian
151	India	Directly mapped (geographic category)	Asian Indian
152	Other South Asia (includes Pakistan, Bangladesh, Himalayan countries)	Directly mapped (geographic category)	Other Asian
153	Burma/Myanmar	Directly mapped (geographic category)	Other Asian
154	Laos/Hmong/Iu Mienh	Directly mapped (geographic category)	Other Asian
155	Cambodia/ Kampuchea	Directly mapped (geographic category)	Other Asian
156	Indonesia/Bali/Java	Directly mapped (geographic category)	Other Asian
157	Malaysia	Indirectly mapped[IA]	Not a Direct Map
158	Malay	Directly mapped (racial category)	QD05ASIA: O.A.
159	Singapore	Indirectly mapped[I]	Not a Direct Map
160	Thailand	Directly mapped (geographic category)	QD05ASIA: O.A.
161	Thai	Directly mapped (racial category)	QD05ASIA: O.A.
162	Mongolia	Directly mapped (geographic category)	QD05ASIA: O.A.
163	Tibet	Directly mapped (geographic category)	QD05ASIA: O.A.
164	Other East Asia	Directly mapped (geographic category)	QD05ASIA: O.A.
165	Djibouti	Indirectly mapped (QD05)	QD05ASIA: O.A.
166	Sudan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
167	Other Eastern Africa	Directly mapped (geographic category)	QD05ASIA: Asian Indian
168	South Africa	Indirectly mapped (QD05)	QD05ASIA: Asian Indian
169	Namibia	Indirectly mapped (QD05)	QD05ASIA: O.A.
170	Zimbabwe	Indirectly mapped (QD05)	QD05ASIA: Asian Indian
171	Zambia	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
172	Botswana	Directly mapped (geographic category)	QD05ASIA: O.A.
173	Angola	Indirectly mapped (QD05)	QD05ASIA: O.A.
174	Mozambique	Directly mapped (geographic category)	QD05ASIA: O.A.
175	Mauritius	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)	Not a Direct Map
176	Other Southern Africa	Directly mapped (geographic category)	QD05ASIA: O.A.
177	Cape Verde	Indirectly mapped (QD05)	QD05ASIA: O.A.
178	Sao Tome	Directly mapped (geographic category)	QD05ASIA: O.A.
179	Mauritania	Indirectly mapped (QD05)	QD05ASIA: O.A.
180	Mali	Indirectly mapped (QD05)	QD05ASIA: O.A.
181	Niger	Indirectly mapped (QD05)	QD05ASIA: O.A.
182	Other Western Africa	Directly mapped (geographic category)	QD05ASIA: O.A.
183	Chad	Directly mapped (geographic category)	QD05ASIA: O.A.
184	Other Central Africa	Directly mapped (geographic category)	QD05ASIA: O.A.
185	African/Africa	Directly mapped (geographic category)	QD05ASIA: Asian Indian
186	Australia	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
187	New Zealand	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[IA]	Not a Direct Map
188	Fiji	Directly mapped (geographic category)	Other Pacific Islander
189	Nauru	Directly mapped (geographic category)	QD05ASIA: Chinese
190	Samoa	Indirectly mapped (QD05)	QD05ASIA: O.A.
191	Samoan	Directly mapped (racial category)	OD05ASIA: O.A.
192	Other Oceania	Directly mapped (geographic category)	QD05ASIA: O.A.
193	European (nonspecific)	Directly mapped (geographic category)	QD05ASIA: O.A.
194	Cape Verde & Portuguese	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
195	Cape Verde & Mexican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
196	Cape Verde & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
201	Biracial (nonspecific)	Codes informative for formal imputation procedures	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
202	White & Black/African American	Directly mapped (racial category)	White & Black/African American
203	White & American Indian/Alaska Native (including mestizo)	Directly mapped (racial category)	White & American Indian/Alaska Native
204	White & Native Hawaiian	Directly mapped (racial category)	White & Native Hawaiian
205	White & Other Pacific Islander	Directly mapped (racial category)	White & Other Pacific Islander
206	White & Asian Indian	Directly mapped (racial category)	White & Asian Indian
207	White & Chinese	Directly mapped (racial category)	White & Chinese
208	White & Filipino	Directly mapped (racial category)	White & Filipino
209	White & Japanese	Directly mapped (racial category)	White & Japanese
210	White & Korean	Directly mapped (racial category)	White & Korean
211	White & Vietnamese	Directly mapped (racial category)	White & Vietnamese
212	White & Other Asian	Directly mapped (racial category)	White & Other Asian
213	White & Asian (nonspecific)	Directly mapped (racial category)	White & Asian Nonspecific
214	White & Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White & American Indian/Alaska Native QD05ASIA: White & Asian Indian
223	Black/African American & American Indian/Alaska Native	Directly mapped (racial category)	Black/African American & American Indian/Alaska Native
224	Black/African American & Native Hawaiian	Directly mapped (racial category)	Black/African American & Native Hawaiian
225	Black/African American & Other Pacific Islander	Directly mapped (racial category)	Black/African American & Other Pacific Islander
226	Black/African American & Asian Indian	Directly mapped (racial category)	Black/African American & Asian Indian
227	Black/African American & Chinese	Directly mapped (racial category)	Black/African American & Chinese
228	Black/African American & Filipino	Directly mapped (racial category)	Black/African American & Filipino

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
229	Black/African American & Japanese	Directly mapped (racial category)	Black/African American & Japanese
230	Black/African American & Korean	Directly mapped (racial category)	Black/African American & Korean
231	Black/African American & Vietnamese	Directly mapped (racial category)	Black/African American & Vietnamese
232	Black/African American & Other Asian	Directly mapped (racial category)	Black/African American & Other Asian
233	Black/African American & Asian (nonspecific)	Directly mapped (racial category)	Black/African American & Asian Nonspecific
234	Black/African American & Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: Black/African American & American Indian QD05ASIA: Black/African American & Asian Indian
244	American Indian/Alaska Native & Native Hawaiian	Directly mapped (racial category)	American Indian/Alaska Native & Native Hawaiian
245	American Indian/Alaska Native & Other Pacific Islander	Directly mapped (racial category)	American Indian/Alaska Native & Other Pacific Islander
246	American Indian/Alaska Native & Asian Indian	Directly mapped (racial category)	American Indian/Alaska Native & Asian Indian
247	American Indian/Alaska Native & Chinese	Directly mapped (racial category)	American Indian/Alaska Native & Chinese
248	American Indian/Alaska Native & Filipino	Directly mapped (racial category)	American Indian/Alaska Native & Filipino
249	American Indian/Alaska Native & Japanese	Directly mapped (racial category)	American Indian/Alaska Native & Japanese
250	American Indian/Alaska Native & Korean	Directly mapped (racial category)	American Indian/Alaska Native & Korean

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
251	American Indian/Alaska Native & Vietnamese	Directly mapped (racial category)	American Indian/Alaska Native & Vietnamese
252	American Indian/Alaska Native & Other Asian	Directly mapped (racial category)	American Indian/Alaska Native & Other Asian
253	American Indian/Alaska Native & Asian (nonspecific)	Directly mapped (racial category)	American Indian/Alaska Native & Asian Nonspecific
265	Native Hawaiian & Other Pacific Islander	Directly mapped (racial category)	Native Hawaiian & Other Pacific Islander
266	Native Hawaiian & Asian Indian	Directly mapped (racial category)	Native Hawaiian & Asian Indian
267	Native Hawaiian & Chinese	Directly mapped (racial category)	Native Hawaiian & Chinese
268	Native Hawaiian & Filipino	Directly mapped (racial category)	Native Hawaiian & Filipino
269	Native Hawaiian & Japanese	Directly mapped (racial category)	Native Hawaiian & Japanese
270	Native Hawaiian & Korean	Directly mapped (racial category)	Native Hawaiian & Korean
271	Native Hawaiian & Vietnamese	Directly mapped (racial category)	Native Hawaiian & Vietnamese
272	Native Hawaiian & Other Asian	Directly mapped (racial category)	Native Hawaiian & Other Asian
273	Native Hawaiian & Asian (nonspecific)	Directly mapped (racial category)	Native Hawaiian & Asian Nonspecific
286	Other Pacific Islander & Asian Indian	Directly mapped (racial category)	Other Pacific Islander & Asian Indian
287	Other Pacific Islander & Chinese	Directly mapped (racial category)	Other Pacific Islander & Chinese
288	Other Pacific Islander & Filipino	Directly mapped (racial category)	Other Pacific Islander & Filipino
289	Other Pacific Islander & Japanese	Directly mapped (racial category)	Other Pacific Islander & Japanese
290	Other Pacific Islander & Korean	Directly mapped (racial category)	Other Pacific Islander & Korean
291	Other Pacific Islander & Vietnamese	Directly mapped (racial category)	Other Pacific Islander & Vietnamese
292	Other Pacific Islander & Other Asian	Directly mapped (racial category)	Other Pacific Islander & Other Asian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
293	Other Pacific Islander & Asian (nonspecific)	Directly mapped (racial category)	Other Pacific Islander & Asian Nonspecific
307	Asian Indian & Chinese	Directly mapped (racial category)	Asian Indian & Chinese
308	Asian Indian & Filipino	Directly mapped (racial category)	Asian Indian & Filipino
309	Asian Indian & Japanese	Directly mapped (racial category)	Asian Indian & Japanese
310	Asian Indian & Korean	Directly mapped (racial category)	Asian Indian & Korean
311	Asian Indian & Vietnamese	Directly mapped (racial category)	Asian Indian & Vietnamese
312	Asian Indian & Other Asian	Directly mapped (racial category)	Asian Indian & Other Asian
328	Chinese & Filipino	Directly mapped (racial category)	Chinese & Filipino
329	Chinese & Japanese	Directly mapped (racial category)	Chinese & Japanese
330	Chinese & Korean	Directly mapped (racial category)	Chinese & Korean
331	Chinese & Vietnamese	Directly mapped (racial category)	Chinese & Vietnamese
332	Chinese & Other Asian	Directly mapped (racial category)	Chinese & Other Asian
349	Filipino & Japanese	Directly mapped (racial category)	Filipino & Japanese
350	Filipino & Korean	Directly mapped (racial category)	Filipino & Korean
351	Filipino & Vietnamese	Directly mapped (racial category)	Filipino & Vietnamese
352	Filipino & Other Asian	Directly mapped (racial category)	Filipino & Other Asian
360	Japanese & Korean	Directly mapped (racial category)	Japanese & Korean
361	Japanese & Vietnamese	Directly mapped (racial category)	Japanese & Vietnamese
362	Japanese & Other Asian	Directly mapped (racial category)	Japanese & Other Asian
371	Korean & Vietnamese	Directly mapped (racial category)	Korean & Vietnamese
372	Korean & Other Asian	Directly mapped (racial category)	Korean & Other Asian
382	Vietnamese & Other Asian	Directly mapped (racial category)	Vietnamese & Other Asian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
383	Indian (Asian or American unclear) & Native Hawaiian	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Native Hawaiian QD05ASIA: Asian Indian & Native Hawaiian
384	Indian (Asian or American unclear) & Other Pacific Islander	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Other Pacific Islander QD05ASIA: Asian Indian & Other Pacific Islander
385	Indian (Asian or American unclear) & Chinese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Asian Indian & Chinese
386	Indian (Asian or American unclear) & Filipino	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Asian Indian & Filipino
387	Indian (Asian or American unclear) & Japanese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Asian Indian & Japanese
388	Indian (Asian or American unclear) & Korean	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Asian Indian & Korean
389	Indian (Asian or American unclear) & Vietnamese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Asian Indian & Vietnamese
390	Indian (Asian or American unclear) & Other Asian	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Asian Indian & Other Asian
401	White, Black/African American, American Indian/Alaska Native	Directly mapped (racial category)	White, Black/African American, American Indian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
402	White, Black/African American, Native Hawaiian	Directly mapped (racial category)	White, Black/African American, Native Hawaiian
403	White, Black/African American, Other Pacific Islander	Directly mapped (racial category)	White, Black/African American, Other Pacific Islander
404	White, Black/African American, Asian Indian	Directly mapped (racial category)	White, Black/African American, Asian Indian
405	White, Black/African American, Chinese	Directly mapped (racial category)	White, Black/African American, Chinese
406	White, Black/African American, Filipino	Directly mapped (racial category)	White, Black/African American, Filipino
407	White, Black/African American, Japanese	Directly mapped (racial category)	White, Black/African American, Japanese
408	White, Black/African American, Korean	Directly mapped (racial category)	White, Black/African American, Korean
409	White, Black/African American, Vietnamese	Directly mapped (racial category)	White, Black/African American, Vietnamese
410	White, Black/African American, Other Asian	Directly mapped (racial category)	White, Black/African American, Other Asian
411	White, Black/African American, Asian (nonspecific)	Directly mapped (racial category)	White, Black/African American, Asian Nonspecific
412	White, American Indian/Alaska Native, Native Hawaiian	Directly mapped (racial category)	White, American Indian/Alaska Native, Native Hawaiian
413	White, American Indian/Alaska Native, Other Pacific Islander	Directly mapped (racial category)	White, American Indian/Alaska Native, Other Pacific Islander
414	White, American Indian/Alaska Native, Asian Indian	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian Indian
415	White, American Indian/Alaska Native, Chinese	Directly mapped (racial category)	White, American Indian/Alaska Native, Chinese
416	White, American Indian/Alaska Native, Filipino	Directly mapped (racial category)	White, American Indian/Alaska Native, Filipino

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
417	White, American Indian/Alaska Native, Japanese	Directly mapped (racial category)	White, American Indian/Alaska Native, Japanese
418	White, American Indian/Alaska Native, Korean	Directly mapped (racial category)	White, American Indian/Alaska Native, Korean
419	White, American Indian/Alaska Native, Vietnamese	Directly mapped (racial category)	White, American Indian/Alaska Native, Vietnamese
420	White, American Indian/Alaska Native, Other Asian	Directly mapped (racial category)	White, American Indian/Alaska Native, Other Asian
421	White, American Indian/Alaska Native, Asian (nonspecific)	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian Nonspecific
422	White, Black/African American, Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White, Black/African American, American Indian/Alaska Native QD05ASIA: White, Black/African American, Asian Indian
423	White, Native Hawaiian, Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White, Native Hawaiian, American Indian/Alaska Native QD05ASIA: White, Native Hawaiian, Asian Indian
424	White, Other Pacific Islander, Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White, Other Pacific Islander, American Indian/Alaska Native QD05ASIA: White, Other Pacific Islander, Asian
600	Mexican & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
601	Mexican & El Salvadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
602	Mexican & Honduran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
603	Mexican & Nicaraguan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
604	Mexican & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
605	Mexican & Panamanian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
606	Mexican & Colombian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
607	Mexican & Venezuelan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
608	Mexican & Ecuadorian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
609	Mexican & Peruvian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
610	Mexican & Bolivian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
611	Mexican & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
612	Mexican & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
613	Mexican & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
614	Mexican & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
615	Mexican & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
616	Puerto Rican & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
617	Puerto Rican & El Salvadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
618	Puerto Rican & Honduran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
619	Puerto Rican & Nicaraguan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
620	Puerto Rican & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
621	Puerto Rican & Panamanian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
622	Puerto Rican & Colombian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
623	Puerto Rican & Venezuelan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
624	Puerto Rican & Ecuadorian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
625	Puerto Rican & Peruvian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
626	Puerto Rican & Bolivian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
627	Puerto Rican & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
628	Puerto Rican & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
629	Puerto Rican & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
630	Puerto Rican & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
631	Puerto Rican & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
632	Cuban & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
633	Cuban & El Salvadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
634	Cuban & Honduran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
635	Cuban & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
636	Cuban & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
637	Cuban & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
638	Cuban & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
639	Cuban & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
640	Cuban & Ecuadorian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
641	Cuban & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
642	Cuban & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
643	Cuban & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
644	Cuban & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
645	Cuban & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
646	Cuban & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
647	Cuban & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
648	Dominican & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
649	Dominican & El Salvadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
650	Dominican & Honduran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
651	Dominican & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
652	Dominican & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
653	Dominican & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
654	Dominican & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
655	Dominican & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
656	Dominican & Ecuadorian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
657	Dominican & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
658	Dominican & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
659	Dominican & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
660	Dominican & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
661	Dominican & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
662	Dominican & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
663	Dominican & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
664	Spanish (from Spain) & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
665	Spanish (from Spain) & El Salvadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
666	Spanish (from Spain) & Honduran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
667	Spanish (from Spain) & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
668	Spanish (from Spain) & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
669	Spanish (from Spain) & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
670	Spanish (from Spain) & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
671	Spanish (from Spain) & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
672	Spanish (from Spain) & Ecuadorian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
673	Spanish (from Spain) & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
674	Spanish (from Spain) & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
675	Spanish (from Spain) & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
676	Spanish (from Spain) & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
677	Spanish (from Spain) & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
678	Spanish (from Spain) & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
679	Spanish (from Spain) & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
680	Colombian & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
681	Colombian & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
682	Colombian & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
683	Colombian & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
684	Colombian & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
685	Colombian & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
686	Colombian & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
687	Colombian & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
688	Colombian & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
689	Venezuelan & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
690	Venezuelan & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
691	Venezuelan & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
692	Venezuelan & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
693	Venezuelan & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
694	Venezuelan & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
695	Venezuelan & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
696	Venezuelan & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
697	Ecuadoran & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
698	Ecuadoran & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
699	Ecuadoran & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
700	Ecuadoran & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
701	Ecuadoran & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
702	Ecuadoran & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
703	Ecuadoran & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
704	Peruvian & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
705	Peruvian & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
706	Peruvian & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
707	Peruvian & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
708	Peruvian & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
709	Peruvian & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
710	Bolivian & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
711	Bolivian & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
712	Bolivian & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
713	Bolivian & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
714	Bolivian & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
715	Chilean & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
716	Chilean & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
717	Chilean & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
718	Chilean & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
719	Argentine & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
720	Argentine & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
721	Argentine & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
722	Paraguayan & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
723	Paraguayan & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
724	Uruguayan & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
725	Guatemalan & Colombian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
726	El Salvadoran & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
727	Honduran & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
728	Nicaraguan & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
729	Costa Rican & Colombian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
730	Panamanian & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
731	Guatemalan & Venezuelan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
732	El Salvadoran & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
733	Honduran & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
734	Nicaraguan & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
735	Costa Rican & Venezuelan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
736	Panamanian & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
737	Guatemalan & Ecuadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
738	El Salvadoran & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
739	Honduran & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
740	Nicaraguan & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
741	Costa Rican & Ecuadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
742	Panamanian & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
743	Guatemalan & Peruvian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
744	El Salvadoran & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
745	Honduran & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
746	Nicaraguan & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
747	Costa Rican & Peruvian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
748	Panamanian & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
749	Guatemalan & Bolivian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
750	El Salvadoran & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
751	Honduran & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
752	Nicaraguan & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
753	Costa Rican & Bolivian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
754	Panamanian & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
755	Guatemalan & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
756	El Salvadoran & Chilean	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
757	Honduran & Chilean	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
758	Nicaraguan & Chilean	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
759	Costa Rican & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
760	Panamanian & Chilean	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
761	Guatemalan & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
762	El Salvadoran & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
763	Honduran & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
764	Nicaraguan & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
765	Costa Rican & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
766	Panamanian & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
767	Guatemalan & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
768	El Salvadoran & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
769	Honduran & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
770	Nicaraguan & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
771	Costa Rican & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
772	Panamanian & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
773	Guatemalan & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
774	El Salvadoran & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
775	Honduran & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
776	Nicaraguan & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
777	Costa Rican & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
778	Panamanian & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
779	Guatemalan & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
780	El Salvadoran & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
781	Honduran & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
782	Nicaraguan & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
783	Costa Rican & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
784	Panamanian & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
785	Guatemalan & El Salvadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
786	Guatemalan & Honduran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
787	Guatemalan & Nicaraguan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
788	Guatemalan & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
789	Guatemalan & Panamanian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
790	El Salvadoran & Honduran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
791	El Salvadoran & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
792	El Salvadoran & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
793	El Salvadoran & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
794	Honduran & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
795	Honduran & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
796	Honduran & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
797	Nicaraguan & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
798	Nicaraguan & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
799	Costa Rican & Panamanian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
800	Mexican & Jamaican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
801	Puerto Rican & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
802	Cuban & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
803	Dominican & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
804	Spanish (from Spain) & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
805	Mexican & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
806	Puerto Rican & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
807	Cuban & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
808	Dominican & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
809	Spanish (from Spain) & Other European	Directly mapped (geographic category)	QD05: White
810	Trinidadian & Mexican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
811	Trinidadian & Puerto Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
812	Trinidadian & Cuban	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
813	Trinidadian & Dominican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
814	Trinidadian & Spanish (from Spain)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
815	Guatemalan & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
816	El Salvador & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
817	Honduran & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
818	Nicaraguan & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
819	Costa Rican & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
820	Panamanian & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
821	Colombian & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
822	Venezuelan & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
823	Ecuadoran & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
824	Peruvian & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
825	Bolivian & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
826	Chilean & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
827	Argentine & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
828	Paraguay & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
829	Uruguayan & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
830	Brazil & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
831	(part) Mexican, ½ (part) White	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
832	(part) Mexican, ½ (part) Black/African American	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
833	(part) Mexican, ½ (part) American Indian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
834	(part) Mexican, ½ (part) Hawaiian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
835	(part) Mexican, ½ (part) Other Pacific Islander	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
836	(part) Mexican, ½ (part) Asian Indian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
837	(part) Mexican, ½ (part) Chinese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
838	(part) Mexican, ½ (part) Filipino	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
839	(part) Mexican, ½ (part) Japanese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
840	(part) Mexican, ½ (part) Korean	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
841	(part) Mexican, ½ (part) Vietnamese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
842	(part) Mexican, ½ (part) Other Asian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
843	(part) Puerto Rican, ½ (part) White	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
844	(part) Puerto Rican, ½ (part) Black/African American	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
845	(part) Puerto Rican, ½ (part) American Indian/Alaska Native	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
846	(part) Puerto Rican, ½ (part) Native Hawaiian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
847	(part) Puerto Rican, ½ (part) Other Pacific Islander	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
848	(part) Puerto Rican, ½ (part) Asian Indian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
849	(part) Puerto Rican, ½ (part) Chinese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
850	(part) Puerto Rican, ½ (part) Filipino	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
851	(part) Puerto Rican, ½ (part) Japanese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
852	(part) Puerto Rican, ½ (part) Korean	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
853	(part) Puerto Rican, ½ (part) Vietnamese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
854	(part) Puerto Rican, ½ (part) Other Asian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
855*	(part) Hispanic/Latino, ½ (part) White	Directly mapped (racial category)	White

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
856*	(part) Hispanic/Latino, ½ (part) Black/African American	Directly mapped (racial category)	Black/African American
857*	(part) Hispanic/Latino, ½ (part) American Indian/Alaska Native	Directly mapped (racial category)	American Indian/Alaska Native
858*	(part) Hispanic/Latino, ½ (part) Native Hawaiian	Directly mapped (racial category)	Native Hawaiian
859*	(part) Hispanic/Latino, ½ (part) Other Pacific Islander	Directly mapped (racial category)	Other Pacific Islander
860*	(part) Hispanic/Latino, ½ (part) Asian Indian	Directly mapped (racial category)	Asian Indian
861*	(part) Hispanic/Latino, ½ (part) Chinese	Directly mapped (racial category)	Chinese
862*	(part) Hispanic/Latino, ½ (part) Filipino	Directly mapped (racial category)	Filipino
863*	(part) Hispanic/Latino, ½ (part) Japanese	Directly mapped (racial category)	Japanese
864*	(part) Hispanic/Latino, ½ (part) Korean	Directly mapped (racial category)	Korean
865*	(part) Hispanic/Latino, ½ (part) Vietnamese	Directly mapped (racial category)	Vietnamese
866*	(part) Hispanic/Latino, ½ (part) Other Asian	Directly mapped (racial category)	Other Asian
869	Haitian & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
870	Haitian & Dominican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
871	Honduran & Haitian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
872	Guatemalan & Iranian	Indirectly mapped (QD05) (double census)[1]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
873	Panamanian & Jamaican	Indirectly mapped (QD05) (double census)I	QD05ASIA: O.A.
874	Cuban & Thai	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
875	Venezuelan & Trinidadian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
876	Puerto Rican & Arab	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
878	Mexican & Samoan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
879	Salvadoran & Egyptian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
880	Costa Rican & Haitian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
881	Mexican & Iranian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
882	Spanish & Barbadian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
883	Peruvian & Other Middle Eastern	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
884	Puerto Rican & African	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
885	Jamaican & Egyptian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
886	Argentine & Turkish	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
887	Mexican & Egyptian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
888	Guatemalan & Canadian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
889	Haitian & European	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
890	Argentine, Cuban, & Spanish	Indirectly mapped (QD05) (triple census)	QD05ASIA: O.A.
891	Mexican, Cuban, & France	Indirectly mapped (QD05) (triple census)[I]	QD05ASIA: O.A.
892	Mexican, Puerto Rican, & European	Indirectly mapped (QD05) (triple census)[I]	QD05ASIA: O.A.
893	Haiti & Trinidad	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
894	Belize & Honduras	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
895	Trinidad & European	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
896	Puerto Rican & Haitian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
897	Puerto Rican & Virgin Islander	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
900*	Definitely Hispanic/Latino (Hispanic, Latino/a, Chicano/a, etc., not Spain or Dominican Republic)	Codes informative for formal imputation procedures	Not a Direct Map
901*	Definitely Hispanic/Latino (Hispanic Spanish, Español, etc.)	Codes informative for formal imputation procedures	Not a Direct Map
902*	Definitely Hispanic/Latino (Hispanic Dominican Republic, Dominicano, etc.)	Indirectly mapped (QD05)	QD05ASIA: O.A
903	Central/South American (no country)	Codes informative for formal imputation procedures	Not a Direct Map
904	Nonwhite (nonspecific/brown)	Codes informative for formal imputation procedures	Not a Direct Map
905*	Hispanic/Latino nonwhite (including trigueno = "dark," moreno)	Codes informative for formal imputation procedures	Not a Direct Map
906*	Mezclado, Mezclada (Hispanic/Latino mixed)	Codes informative for formal imputation procedures	Not a Direct Map
907	Mixed	Codes informative for formal imputation procedures	Not a Direct Map
908	Olive	Directly mapped (geographic category)	White
909	Creole	Indirectly mapped	QD05ASIA: O.A.
910	Arab	Directly mapped (geographic category)	QD05: White QD05ASIA: Other Asian
911	Jewish	Directly mapped (geographic category)	White
912	Kurd	Directly mapped (geographic category)	Other Asian
913	Chaldean/Caldanian/ Assyrian	Directly mapped (geographic category)	Other Asian
914	Romany/Gypsy	Directly mapped (geographic category)	White
915	Central/South American & West Indies	Indirectly mapped	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
916	Central/South American & Mexican	Codes informative for formal imputation procedures	Not a Direct Map
917	Central/South American & Puerto Rican	Codes informative for formal imputation procedures	Not a Direct Map
918	Central/South American & Cuban	Codes informative for formal imputation procedures	Not a Direct Map
919	Central/South American & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
920	Central/South American & Spanish	Codes informative for formal imputation procedures	Not a Direct Map
921	Arab/Asian	QD05: Directly mapped (racial category) QD05ASIA: Directly mapped (geographic category)	QD05: White & Asian QD05ASIA: Other Asian
922	Arab/European	Directly mapped (geographic category)	QD05: White QD05ASIA: White & Other Asian
923	Arab/African	Directly mapped (geographic category)	QD05: White & Black/African American QD05ASIA: Other Asian & Black/African American
924	Arab/Chaldean	Directly mapped (racial category)	QD05: White & Other Asian QD05ASIA: O.A.
925	European & Asian Indian	Directly mapped (racial category)	White & Asian Indian
926	West Indies & Belize	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
927	West Indies & Cape Verde	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
928	Arab & Cape Verde	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
929	Aryan	Directly mapped (geographic category)	QD05: White QD05ASIA: Asian Indian
930	Turkish & Lebanese	Directly mapped (racial category)	White
932	Puerto Rican & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
933	Puerto Rican & Spanish (from Spain)	Codes informative for formal imputation procedures	Not a Direct Map
934	Dominican & Asian Indian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
935	Cuban & Spanish (from Spain)	Codes informative for formal imputation procedures	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
936	Dominican & Spanish (from Spain)	Codes informative for formal imputation procedures	Not a Direct Map
951	White and something else	Directly mapped (racial category)	White (Multiple Race)
952	Black/African American and something else	Directly mapped (racial category)	Black/African American (Multiple Race)
953	American Indian/Alaska Native and something else	Directly mapped (racial category)	American Indian/Alaska Native (Multiple Race)
954	Native Hawaiian and something else	Directly mapped (racial category)	Native Hawaiian (Multiple Race)
955	Other Pacific Islander and something else	Directly mapped (racial category)	Other Pacific Islander (Multiple Race)
956	Asian Indian and something else	Directly mapped (racial category)	Asian Indian (Multiple Race)
957	Chinese and something else	Directly mapped (racial category)	Chinese (Multiple Race)
958	Filipino and something else	Directly mapped (racial category)	Filipino (Multiple Race)
959	Japanese and something else	Directly mapped (racial category)	Japanese (Multiple Race)
960	Korean and something else	Directly mapped (racial category)	Korean (Multiple Race)
961	Vietnamese and something else	Directly mapped (racial category)	Vietnamese (Multiple Race)
962	Other Asian and something else	Directly mapped (racial category)	Other Asian (Multiple Race)
963	Asian (nonspecific) and something else	Codes useful for formal imputation procedures	Not a Direct Map
964	Indian (Asian or American unclear) and something else	Directly mapped (racial category)	QD05: American Indian/Alaska Native (Multiple Race) QD05ASIA: Asian Indian (Multiple Race)
965	Brown & White	Directly mapped (racial category)	White & Black/African American
985	Bad data	Noninformative code	Not a Direct Map
994	Unknown/"Don't Know"	Noninformative code	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
997	Rather Not Say/"Refused" ("American" or "All of Them")	Noninformative code	Not a Direct Map

* These codes caused the Hispanic/Latino indicator to be edited to a "yes" if QD03 was missing or "no." The code that caused the Hispanic/Latino indicator to be edited to a "no" was a Hispanic/Latino code (600) and is listed in Table C.4.

[1] Among the indirectly mapped codes, codes where an imputation was possible based on census information are indicated by the superscript "I." If the imputation was limited to Asians in these cases, the superscript "IA" is used. See Section C.2.1.2 for details.

[2] The abbreviation "O.A." is equivalent to "Other Asian."

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes

Race Code	Race Name	Probabilities
50	Belize	6.1% Black/African American, 10.6% American Indian/Alaska Native, 24.9% White and Black/African American, 48.7% White and American Indian/Alaska Native, 9.7% Unrestricted Imputation QD05 when in combination with another race: 6.1% Black/African American, 13.8% American Indian/Alaska Native, 28.1% White and Black/African American, 51.9% White and American Indian/Alaska Native
51	Guyana	QD05: 36% Black/African American, 7% American Indian/Alaska Native, 50% Asian Indian, 7% Unrestricted Imputation
52	Suriname	QD05: 1% White, 10% Black/African American, 2% American Indian/Alaska Native, 37% Asian Indian, 2% Chinese, 15% Other Asian, 31% White and Black/African American, 2% Unrestricted Imputation QD05ASIA: 71% Asian Indian, 29% Other Asian
53	Haiti	95% Black/African American, 5% White and Black/African American
54	Trinidad and Tobago	QD05: 34.2% Black/African American, 35.4% Asian Indian, 7.7% Black/African American and Asian Indian, 15.3% Impute among those with multiple races, 7.4% Unrestricted Imputation QD05ASIA: 82% Asian Indian, 18% Black/African American and Asian Indian QD05 when in combination with another race: 36.7% Black/African American, 37.9% Asian Indian, 25.5% Black/African American and Asian Indian
55	Jamaica	QD05: 92.1% Black/African American, 0.8% Asian Indian, 6.1% White and Black/African American,1% Unrestricted Imputation QD05 when in combination with another race: 92.4% Black/African American, 1.1% Asian Indian, 6.4% White and Black/African American
56	Virgin Is (St Thomas, St Croix)	QD05: 13.1% White, 76.2% Black/African American, 1.1% Asian Nonspecific, 3.5% White and Black/African American, 6.1% Unrestricted Imputation QD05ASIA: Impute among Asians QD05 when in combination with another race: 15.1% White, 78.2% Black/African American, 1.1% Asian Nonspecific, 5.5% White and Black/African American
57	Bahamas	QD05: 4.7% White, 90.6% Black/African American, 2.1% White and Black/African American, 2.6% Unrestricted Imputation QD05ASIA: Impute among Asians
58	Barbados	2.7% White, 92.4% Black/African American, 3.1% White and Black/African American, 1.3% Asian Indian, 0.5% Unrestricted Imputation QD05 when in combination with another race: 2.9% White, 92.6% Black/African American, 3.3% White and Black/African American, 1.3% Asian Indian

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
59	Grenada	82% Black/African American, 13% White and Black/African American, 5% Unrestricted Imputation
60	St. Lucia	QD05: 85.3% Black/African American, 2.2% Asian Indian, 5.45% White and Black/African American, 5.45% Black/African American and Asian Indian, 1.7% Unrestricted Imputation QD05ASIA: 29% Asian Indian 71% Black/African American and Asian Indian
63	Other West Indies	QD05: 80% Black/African American, 14% Asian Nonspecific, 6% Unrestricted Imputation QD05ASIA: Impute among Asians QD05 when in combination with another race: 85.1% Black/African American, 14.9% Asian Nonspecific
64	Brazil	47.7% White, 7.6% Black/African American, 0.5% American Indian/Alaska Native, 43.1% White and Black/African American, 1.1% Asian Nonspecific
65	Canada	QD05: 91.3% White, 4.2% American Indian/Alaska Native, 4.5% Chinese QD05ASIA: Impute among Asians
70	Mexico	QD05 when in combination with another race: 9.3% White, 30.3% American Indian/Alaska Native, 60.3% White and American Indian/Alaska Native
71	Puerto Rico	QD05 when in combination with another race: 75.8% White, 12.4% Black/African American, 2.8% American Indian/Alaska Native, 2.8% Native Hawaiian, 2.8% Other Pacific Islander, 1.1% White and Black/African American, 1.1% White and American Indian/Alaska Native, 1.1% Black/African American and American Indian/Alaska Native
72	Cuba	QD05 when in combination with another race: 64.1% White, 9.3% Black/African American, 26.6% White and American Indian/Alaska Native
73	Dominican Republic	16% White, 11% Black/African American, 73% White and Black/African American
74	Guatemala	40.5% American Indian/Alaska Native, 59.4% White and American Indian/Alaska Native, 0.1% Unrestricted Imputation QD05 when in combination with another race: 40.5% American Indian/Alaska Native, 59.5% White and American Indian/Alaska Native
75	Honduras	1% White, 2% Black/African American, 7% American Indian/Alaska Native, 90% White and American Indian/Alaska Native
76	El Salvador	12.7% White, 1% American Indian/Alaska Native, 86.3% White and American Indian/Alaska Native
77	Nicaragua	17% White, 9% Black/African American, 5% American Indian/Alaska Native, 69% White and American Indian/Alaska Native

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
78	Costa Rica	QD05: 1.1% Black/African American, 2.4% American Indian/Alaska Native, 6.7% White and Black/African American, 83.6% Imputed among White or White and American Indian, 6.2% Unrestricted Imputation QD05 when in combination with another race: 43.9% White, 1.1% Black/African American, 2.4% American Indian/Alaska Native, 8.8% White and Black/African American, 43.9% White and American Indian/Alaska Native
79	Panama	10% White, 14% Black/African American, 6% American Indian/Alaska Native, 70% White and American Indian/Alaska Native
80	Colombia	20% White, 4% Black/African American, 1% American Indian/Alaska Native, 14% White and Black/African American, 58% White and American Indian/Alaska Native, 3% Black/African American and American Indian/Alaska Native
81	Venezuela	21% White, 10% Black/African American, 2% American Indian/Alaska Native, 67% White and American Indian/Alaska Native
82	Ecuador	6.1% White, 7.2% Black/African American, 7% American Indian/Alaska Native, 79.3% White and American Indian/Alaska Native, 0.4% Unrestricted Imputation
83	Peru	15% White, 1% Black/African American, 45% American Indian/Alaska Native, 1% Chinese, 1% Japanese, 37% White and American Indian/Alaska Native
84	Bolivia	15% White, 55% American Indian/Alaska Native, 30% White and American Indian/Alaska Native
85	Chile	10.8% American Indian/Alaska Native, 88.9% White, 0.3% Unrestricted Imputation QD05 when in combination with another race: 10.8% American Indian/Alaska Native, 89.2% White
86	Argentina	97% White, 3% White and American Indian/Alaska Native
87	Paraguay	2.5% White, 2.5% American Indian/Alaska Native, 95% White and American Indian/Alaska Native
88	Uruguay	88% White, 4% Black/African American, 8% White and American Indian/Alaska Native
103	United Kingdom	87.2% White, 3% Black/African American, 4.2% Asian Indian, 2% Impute Among Multiple Race, 3.5% Unrestricted Imputation
125	Saudi Arabia	QD05: 90% White, 10% Asian Indian QD05ASIA: 90% Other Asian, 10% Asian Indian
128	UAE	QD05: 30.5% White, 50% Asian Indian, 11.5% Other Asian, 8% Not American Indian/Alaska Native QD05ASIA: 50% Asian Indian, 50% Other Asian
129	Qatar	QD05: 40% White, 36% Asian Indian, 10% Other Asian, 14% Not American Indian/Alaska Native QD05ASIA: 36% Asian Indian, 64% Other Asian

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
130	Bahrain	QD05: 51.7% White, 1.6% Black/African American, 45.5% Asian Indian, 1.2% Impute among those who are not American Indian/Alaska Native
133	Kuwait	QD05ASIA: 9% Asian Indian, 91% Other Asian
140	Kazakhstan	26.9% White, 68.6% Other Asian, 4.5% Impute among those who are not American Indian/Alaska Native
141	Uzbekistan	5.5% White, 92% Other Asian, 2.5% Not American Indian/Alaska Native
142	Tadjikistan	3.5% White, 89.9% Other Asian, 6.6% Not American Indian/Alaska Native
143	Kyrgizstan	22.9% White, 65.3% Other Asian, 11.8% Not American Indian/Alaska Native
144	Turkmenistan	6.7% White, 88.2% Other Asian, 5.1% Not American Indian/Alaska Native
157	Malaysia	6.7% Asian Indian, 22.6% Chinese, 61.9% Other Asian, 8.8% Impute among those who are not American Indian/Alaska Native
159	Singapore	9.2% Asian Indian, 74.2% Chinese, 13.3% Other Asian, 3.3% Impute among those who are not American Indian/Alaska Native
165	Djibouti	2.5% White, 97.5% Black/African American
166	Sudan	39% White, 58% Black/African American, 3% Impute among those who are not American Indian/Alaska Native
168	South Africa	8.9% White, 79.2% Black/African American, 2.5% Asian Indian, 8.9% White and Black/African American, 0.5% Impute among those who are not American Indian/Alaska Native
169	Namibia	6% White, 87.5% Black/African American, 6.5% White and Black/African American
170	Zimbabwe	1% White, 98% Black/African American, .5% Asian Indian, .5% White and Black/African American
171	Zambia	99% Black/African American, 1% Impute among those who are not American Indian/Alaska Native
173	Angola	1% White, 97% Black/African American, 2% White and Black/African American
175	Mauritius	QD05: 2% White, 68% Asian Indian, 3% Chinese, 27% White and Black/African American QD05ASIA: 96% Asian Indian, 4% Chinese
177	Cape Verde	1% White, 28% Black/African American, 71% White and Black/African American
179	Mauritania	30% White, 30% Black/African American, 40% White and Black/African American
180	Mali	10% White, 90% Black/African American
181	Niger	9.3% White, 89.5% Black/African American, 1.2% Impute among those who are not American Indian/Alaska Native
186	Australia	QD05: 92% White, 7% Asian Nonspecific, 1% Not American Indian/Alaska Native QD05ASIA: Impute among Asians

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
187	New Zealand	QD05: 67% White, 21.7% Other Pacific Islander, 11.3% Other Asian QD05ASIA: Impute among Asians
190	Samoa	.4% White, 92.6% Other Pacific Islander, 7% White and Other Pacific Islander
902	Definitely Hispanic/Latino (Hispanic Dominican Republic, Dominicano, etc.)	16% White, 11% Black/African American, 73% White and Black/African American
909	Creole	50% White, 50% White and Black/African American
915	Central/South American & West Indies	50% White and Black/African American, 50% Black/African American and American Indian/Alaska Native
920	Central/South American & Spanish	50% White, 50% White and American Indian/Alaska Native

Table C.3 Procedures for Restricted Imputation for Codes Informative for Formal Imputation Procedures

Race Code	Race Name	Restriction on Donors in Formal Imputation
33	Asian (nonspecific)	Donors were Asian: impute specific Asian group
70	Mexico	Donors were Mexican[1]
71	Puerto Rico	Donors were Puerto Rican
72	Cuba	Donors were Cuban
78	Costa Rica (QD05: 94% White or Mestizo)	For this 94%, donors were white or white and American Indian/Alaska Native
89	Mexico & Puerto Rico	Donors were Mexican, Puerto Rican, or both
90	Mexico & Cuba	Donors were Mexican, Cuban, or both
91	Mexico & Dominican Republic	Donors were Mexican, Dominican, or both
92	Mexico & Spain	Donors were Mexican, Spanish, or both
93	Puerto Rico & Cuba	Donors were Puerto Rican, Cuban, or both
94	Puerto Rico & Dominican Republic	Donors were Puerto Rican, Dominican, or both
95	Puerto Rico & Spain	Donors were Puerto Rican, Spanish, or both
96	Cuba & Dominican Republic	Donors were Cuban, Dominican, or both
97	Cuba & Spain	Donors were Cuban, Spanish, or both
98	Dominican Republic & Spain	Donors were Dominican, Spanish, or both
128	UAE (QD05: 8% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
129	Qatar (QD05: 14% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
140	Kazakhstan (QD05: 6.6% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
141	Uzbekistan (QD05: 2.5% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
142	Tadjikistan (QD05: 6.6% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
143	Kyrgizstan (QD05: 11.8% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
144	Turkmenistan (QD05: 5.1% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
159	Singapore (QD05: 1.4% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
166	Sudan (QD05: 3% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
171	Zambia (QD05: 0.2% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
186	Australia (QD05: 1% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
201	Biracial (nonspecific)	Donors were multiple race: imputed constituent races[2]

Table C.3 Procedures for Restricted Imputation for Codes Informative for Formal Imputation Procedures (continued)

Race Code	Race Name	Restriction on Donors in Formal Imputation
900	Definitely Hispanic/Latino (Hispanic, Latino/a, Chicano/a, etc., not Spain, D.R.)	Donors were Hispanic/Latino
901	Definitely Hispanic/Latino (Hispanic Spanish, Español, etc.)	Donors were Hispanic/Latino
903	Central/South American (no country)	Donors were Central/South American
904	Nonwhite (nonspecific/brown)	Donors were any race but single-race white
905	Hispanic/Latino nonwhite (including trigueno = "dark," moreno)	Donors were Hispanic/Latino who were any race but single-race white
906	Mezclado, Mezclada (Hispanic/Latino mixed)	Donors were multiple race and Hispanic/Latino: imputed constituent races
907	Mixed	Donors were multiple race: imputed constituent races
916	Central/South American & Mexican	Donors were Central/South American, Mexican, or both
917	Central/South American & Puerto Rican	Donors were Central/South American, Puerto Rican, or both
918	Central/South American & Cuban	Donors were Central/South American, Cuban, or both
919	Central/South American & Dominican	Donors were Central/South American, Dominican, or both
920	Central/South American & Spanish	Donors were Central/South American, Spanish, or both
932	Puerto Rican & Dominican	Donors were Puerto Rican, Dominican, or both
933	Puerto Rican & Spanish (from Spain)	Donors were Puerto Rican, Spanish, or both
935	Cuban & Spanish (from Spain)	Donors were Cuban, Spanish, or both
936	Dominican & Spanish (from Spain)	Donors were Dominican, Spanish, or both

[1] Even though a recipient may not be Hispanic/Latino, he or she may still have indicated "Mexican" in the QD05 "OTHER, specify" response. Donors in this case included both Hispanic/Latino and (though extremely rare) non-Hispanic/Latino Mexicans.

[2] Because most multiple-race respondents have only two constituent races, any respondent with this code and nothing else is likely to be assigned a biracial donor. However, for the sake of simplicity, respondents with this code were not treated any differently than respondents with code 907 ("Mixed").

Table C.4 Mapping of Hispanic/Latino Group Codes

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
11	Mexican/Mexican American/Mexicano/Chicano	Mexican
12	Puerto Rican	Puerto Rican
13	Central or South American	Central or South American
14	Cuban/Cuban American	Cuban
15	Dominican (Dominican Republic)	Dominican
16	Spanish (from Spain)	Spanish
17	Caribbean Hispanic/Latino (not specified as Dominican)	Other Hispanic
21	Mexican & Puerto Rican	Mexican
22	Mexican & Central or South American	Mexican
23	Mexican & Cuban	Mexican
24	Mexican & Dominican	Mexican
25	Mexican & Spanish (from Spain)	Mexican
26	Puerto Rican & Central or South American	Puerto Rican
27	Puerto Rican & Cuban	Cuban
28	Puerto Rican & Dominican	Puerto Rican
29	Puerto Rican & Spanish (from Spain)	Puerto Rican
30	Central or South American & Cuban	Cuban
31	Central or South American & Dominican	Central or South American
32	Central or South American & Spanish (from Spain)	Central or South American
33	Cuban & Dominican	Cuban
34	Cuban & Spanish (from Spain)	Cuban
35	Dominican & Spanish (from Spain)	Dominican
36	Mexican, Puerto Rican, & Central or South American	Mexican
37	Mexican, Puerto Rican, & Cuban	Mexican
38	Mexican, Puerto Rican, & Dominican	Mexican
39	Mexican, Puerto Rican, & Spanish (from Spain)	Mexican
40	Mexican, Central or South American, & Cuban	Mexican
41	Mexican, Central or South American, & Dominican	Mexican
42	Mexican, Central or South American, & Spanish (from Spain)	Mexican
43	Mexican, Cuban, & Dominican	Mexican
44	Mexican, Cuban, & Spanish (from Spain)	Mexican
45	Mexican, Dominican, & Spanish (from Spain)	Mexican
46	Puerto Rican, Central or South American, & Cuban	Cuban
47	Puerto Rican, Central or South American, & Dominican	Puerto Rican

Table C.4 Mapping of Hispanic/Latino Group Codes (continued)

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
48	Puerto Rican, Central or South American, & Spanish (from Spain)	Puerto Rican
49	Puerto Rican, Cuban, & Dominican	Cuban
50	Puerto Rican, Cuban, & Spanish (from Spain)	Cuban
51	Puerto Rican, Dominican, & Spanish (from Spain)	Puerto Rican
52	Central or South American, Cuban, & Dominican	Cuban
53	Central or South American, Cuban, & Spanish (from Spain)	Cuban
54	Central or South American, Dominican, & Spanish (from Spain)	Central or South American
55	Cuban, Dominican, and Spanish (from Spain)	Cuban
56	Portuguese & Mexican	Mexican
57	Portuguese & Puerto Rican	Puerto Rican
58	Portuguese & Cuban	Cuban
59	Portuguese & Central or South American	Central or South American
60	Portuguese & Dominican	Dominican
61	Portuguese & Spanish (from Spain)	Spanish
100	Brazilian	Central or South American
101	Portuguese	Other Hispanic/Latino
102	Cape Verde	Other Hispanic/Latino
103	Belizean (formerly British Honduras)	Central or South American
104	Guyana	Central or South American
105	Jamaican	Other Hispanic/Latino
106	Other Caribbean (possibly Hispanic)	Other Hispanic/Latino
107	Philippines & Guam	Other Hispanic/Latino
108	Brazilian & Portuguese	Central or South American
109	Cape Verde & Portuguese	Other Hispanic/Latino
110	Haitian/Jamaican	Other Hispanic/Latino
200	Mexican/Jamaican	Mexican
201	Puerto Rican/Jamaican	Puerto Rican
202	Central or South American/Jamaican	Central or South American
203	Cuban/Jamaican	Cuban
204	Dominican/Jamaican	Dominican
205	Spanish (from Spain)/Jamaican	Spanish
206	Mexican/West Indies	Mexican
207	Puerto Rican/West Indies	Puerto Rican
208	Central or South American/West Indies	Central or South American

Table C.4 Mapping of Hispanic/Latino Group Codes (continued)

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
209	Cuban/West Indies	Cuban
210	Dominican/West Indies	Dominican
211	Spanish (from Spain)/West Indies	Spanish
212	Mexican/Haitian	Mexican
213	Puerto Rican/Haitian	Puerto Rican
214	Central or South American/Haitian	Central or South American
215	Cuban/Haitian	Cuban
216	Dominican/Haitian	Dominican
217	Spanish (from Spain)/Haitian	Spanish
500	Hispanic/Latino	Hispanic/Latino group imputed
501	Hispanic/Latino Mixed/Mezclada	Hispanic group imputed
502	Hispanic Creole	Other Hispanic
600*	Stated clearly as Not Hispanic/Latino	Hispanic/Latino indicator edited to "no"
800	Non-Hispanic/Latino country	Other Hispanic/Latino
801	Racial category (white, black/African American, etc.)	Hispanic group imputed
802	Combination race and non-Hispanic country	Other Hispanic/Latino
985	Bad Data/"Mixed"	Hispanic/Latino group imputed
994	Unknown/"Don't Know"	Hispanic/Latino group imputed
997	American or "All of Them"	Hispanic/Latino group imputed

* This code caused the Hispanic/Latino indicator to be edited to a "no." Codes that caused the Hispanic/Latino indicator to be edited to a "yes" are listed in Table C.1.

Appendix D: Model Summaries

This page intentionally left blank

Appendix D: Model Summaries

D.1 Introduction

The tables in this appendix list the covariates used in all the imputation models that were run in the 2014 National Survey on Drug Use and Health (NSDUH). For each variable or set of variables to which the predictive mean neighborhood (PMN) imputation method was applied, two models were run: one to adjust the weights for item nonresponse (response propensity models) and a second to calculate predicted means. Imputation was usually done separately among age groups. Therefore, most of the tables within this appendix display only one age group at a time.

The models for the demographic variables are presented in Section D.3 and the models for the drug variables are presented in Section D.4. With the exception of the lifetime usage models, separate tables are provided in Section D.4 for each drug age group combination. Tables that present the models for each age group for the household composition variables, which are derived from the questionnaire roster items, are provided in Section D.5. Section D.6 presents the models for the income variables and Section D.7 presents the models for the health insurance variables. Chapter 9 provides a more detailed description of these two methods. Section D.8 presents the models for the roster pair variables.

The definition of terms and variables in the models of the various sections can be found in the chapters that correspond to those sections as follows: Section D.3 (Chapters 3 & 4), Section D.4 (Chapters 5 & 6), Section D.5 (Chapter 7), Section D.6 (Chapter 8), Section D.7 (Chapter 9), and Section D.8 (Chapter 10).

In the tables, the variable "age" is the mean-centered age, where the age was "centered" by subtracting the mean age and where the mean was calculated across all respondents within the age group who were used to build the given model. The variables "age squared" and "age cubed" represent the square and cube, respectively, of this mean-centered age variable. Also in the tables, when an asterisk "*" is given, it represents an interaction between two variables.

D.2 Screener and Segment-Level Variables

In the PMN procedure, statistical modeling was performed to adjust weights for item nonresponse and also to calculate predicted means in the imputation models. Descriptions of questionnaire-derived variables are described in the main body of this report, with the exception of screener and segment-level variables, which are described below. These variables were often used as covariates in both types of models for the PMN procedures.

D.2.1 Household Type

Household type was a three-level race/ethnicity variable based on screener data. It was created by recoding the race/ethnicity of the screening head of the household to one of three levels: Hispanic/Latino, non-Hispanic/Latino black/African American, or non-Hispanic/Latino non-black/African American.

D.2.2 Census Region

Census region was a four-level geographic variable recoded from the respondent's state of residence. The four levels were Northeast, Midwest, South, and West. The Northeast includes Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The Midwest consists of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia compose the South. Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming compose the West.

D.2.3 Population Density

The population density variable classifies respondents according to their living situation, whether it be in a rural or urban area and, if urban, the size of the urban area. It was used to categorize segments where the respondents lived according to modified 2000 census data, which were adjusted to reflect population increases between census years by intercensal projections by Claritas, Inc.[1] This variable had five levels: segment in core-based statistical area (CBSA)[2] with 1 million or more people; segment in CBSA with 250,000 to 999,999 people; segment in CBSA with fewer than 250,000 people; segment not in CBSA and not in rural area; and segment not in CBSA and in rural area.

D.2.4 Percentage Hispanic/Latino in Segment

The "Percentage Hispanic/Latino in Segment" variable was used to categorize segments according to the concentration of Hispanics/Latinos in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 20 percent, 20 to less than 71 percent, and 71 percent or more.

D.2.5 Percentage Owner Occupied in Segment

The "Percentage Owner Occupied in Segment" variable was used to categorize segments according to the concentration of owner-occupied households in the segments in which the respondents lived, using the adjusted 2000 census data. It was used as a surrogate for income because wealthy segments tend to have many homeowners, while poor segments tend to have many renters. It had three levels: less than 10 percent, 10 to less than 50 percent, and 50 percent or more.

D.2.6 Percentage Black/African American in Segment

The "Percentage Black/African American in Segment" variable was used to categorize segments according to the concentration of black/African-American households in the segments

[1] Claritas, Inc. is a market research firm headquartered in San Diego, California.

[2] CBSAs, developed in response to standards put forth by the Office of Management and Budget (OMB), are metropolitan and micropolitan areas that were designated using data from the 2000 census. More information about CBSAs can be retrieved from http://www.census.gov/hhes/www/housing/resseg/cbsa.html.

in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 10 percent, 10 to less than 40 percent, and 40 percent or more.

D.2.7 Percentage Asian/Other Pacific Islander in Segment

The "Percentage Asian/Other Pacific Islander in Segment" variable was used to categorize segments according to the concentration of Asian/Other Pacific Islander households in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 5 percent, 5 to less than 10 percent, and 10 percent or more.

D.2.8 Percentage American Indian/Alaska Native in Segment

The "Percentage American Indian/Alaska Native in Segment" variable was used to categorize segments according to the concentration of American Indian/Alaska Native households in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 1 percent, 1 to less than 3 percent, and 3 percent or more.

D.2.9 Percentage Non-Hispanic/Latino Black/African American in Segment

The "Percentage Non-Hispanic/Latino Black/African American in Segment" variable was used to categorize segments according to the concentration of non-Hispanic/Latino black/African-American households in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 10 percent, 10 percent to less than 50 percent, and 50 percent or more.

D.3 Demographic Variables

For justifications of the aggregation of age groups for certain imputation steps, see Chapters 3 and 4.

Exhibit D.1 Definitions of Levels for Categorical Variables

Age Category
Y: 12-17, T: 18-25, A1: 26-34, A2: 35-49, A3: 50+[1]
Percentage Asian/Other Pacific Islander in Segment
H: ≥ 10%, M: [5,10)%, L: < 5%[1]
Percentage Black/African American in Segment
H: ≥ 40%, M: [10,40)%, L: < 10%[1]
Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]
Percentage American Indian/Alaska Native in Segment
H: ≥ 3%, M: [1,3)%, L: < 1%[1]
Hispanic/Latino Origin Indicator
H: Hispanic/Latino, NH: Not Hispanic/Latino[1]

Exhibit D.1 Definitions of Levels for Categorical Variables (continued)

Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]

Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced or Separated, NM: Never Been Married[1]

Race/Ethnicity of Householder
B: Black/African American, H: Hispanic/Latino, W: White[1]

Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]

MSA (3 Levels)
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]

Population Density (5 Levels)
LC: Segment in a CBSA with 1 Million or More People, MC: Segment in a CBSA from 250,000 to 999,999 People, SC: Segment in a CBSA with Fewer than 250,000 People, NC: Segment Not in a CBSA and Not in a Rural Area, R: Segment Not in a CBSA and in a Rural Area[1]

Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only,[1] B: Not Hispanic/Latino and Black/African American Only, O: Not Hispanic/Latino and Other, H: Hispanic/Latino

Race/Hispanic Recode (5 Levels)
W: White Only,[1] B: Black/African American Only, AI: American Indian/Alaska Native Only, A: Asian/Other Pacific Islander Only, MR: Multiple Race

Race/Hispanic Recode (8 Levels)
W: Not Hispanic/Latino and White,[1] B: Not Hispanic/Latino and Black/African American, AI: Not Hispanic/Latino and American Indian/Alaska Native, A: Not Hispanic/Latino and Asian/Other Pacific Islander, MR: Not Hispanic/Latino and Multiple Race, PR: Hispanic/Latino and Puerto Rican, M: Hispanic/Latino and Mexican, O: Not Hispanic/Latino and Other

Census Region
N: Northeast, M: Midwest, S: South, W: West[1]

Gender
M: Male, F: Female[1]

Age Category * Gender
MY: Male 12-17, MT: Male 18-25, MA1: Male 26-34, MA2: Male 35-49, MA3: Male 50+,[1] FY: Female 12-17,[1] FT: Female 18-25,[1] FA1: Female 26-34,[1] FA2: Female 35-49,[1] FA3: Female 50+[1]

Race/Hispanic Recode (8 Levels) * Gender
MW: Male Not Hispanic/Latino and White,[1] MB: Male Not Hispanic/Latino and Black/African American, MA: Male Not Hispanic/Latino and Asian/Other Pacific Islander, MAI: Male Not Hispanic/Latino and American Indian/Alaska Native, MMR: Male Not Hispanic/Latino and Multiple Race, MPR: Male Hispanic/Latino and Puerto Rican, MM: Male Hispanic/Latino and Mexican, MO: Male Not Hispanic/Latino and Other, FW: Female Not Hispanic/Latino and White,[1] FB: Female Not Hispanic/Latino and Black/African American,[1] FA: Female Not Hispanic/Latino and Asian/Other Pacific Islander,[1] FAI: Female Not Hispanic/Latino and American Indian/Alaska Native,[1] FMR: Female Not Hispanic/Latino and Multiple Race,[1] FPR: Female Hispanic/Latino and Puerto Rican,[1] FM: Female Hispanic/Latino and Mexican,[1] FO: Female Not Hispanic/Latino and Other[1]

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area; MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.1 Summaries for Response Propensity Models

Imputation Step	Variables Included in Response Propensity Model
Marital Status	Age Category (Y, T, A1, A2); Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (5 Levels) (LC, MC, SC, NC); Census Region (N, M, S); Gender (M); Age Category * Gender (MY, MT, MA1, MA2)
Race 12-17	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Census Region (N, M, S)
Race 18-25	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Marital Status (3 Levels) (M, WM); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Census Region (N, M, S)
Race 26+	Age Category (A1, A2); Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Marital Status (4 Levels) (M, W, DS); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Census Region (N, M, S)
Hispanic/Latino Origin 12-17	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Census Region (N, M, S)
Hispanic/Latino Origin 18-25	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Census Region (N, M, S)
Hispanic/Latino Origin 26+	Age Category (A1, A2); Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Census Region (N, M, S)
Hispanic/Latino Group	Age Category (Y, T, A1, A2); Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Census Region (N, M, S); Gender (M); Age Category * Gender (MY, MT, MA1, MA2)
Education Level 12-17	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)
Education Level 18+	Age Category (T, A1, A2); Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age Category * Gender (MT, MA1, MA2)

Table D.1 Summaries for Response Propensity Models (continued)

Imputation Step	Variables Included in Response Propensity Model
Employment Status 12-25	Age Category (Y); Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) (L, HS, SC); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age Category * Gender (MY)
Employment Status 26+	Age Category (A1, A2); Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) (L, HS, SC); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age Category * Gender (MA1, MA2)
Born in US 12-17	Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (8 Levels) (B, AI, A, MR, PR, M, O); Census Region (N, M, S); Gender (M)
Born in US 18-25	Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (8 Levels) (B, AI, A, MR, PR, M, O); Census Region (N, M, S); Gender (M)
Born in US 26+	Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (8 Levels) (B, AI, A, MR, PR, M, O); Census Region (N, M, S); Gender (M); Age Category * Gender (MA1, MA2)
Age of Entry	Age Category (Y, T, A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (8 Levels) (B, AI, A, MR, PR, M, O); Census Region (N, M, S); Gender (M); Age Category * Gender (MY, MT, MA1, MA2); Race/Hispanic Recode (8 Levels) * Gender (MB, MA, MAI, MMR, MPR, MM, MO)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.1 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.2 Summaries for Predictive Mean Models

Imputation Step	Variables Included in Predictive Mean Model
Marital Status	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; **Age Cubed**; **Age Squared**; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (5 Levels) (LC, MC, SC, NC); Census Region (N, M, S); Gender (M); Age * Gender
Race 12-17	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; Age Squared; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); **Race/Ethnicity of Householder (B**, **H)**; Percentage Owner Occupied in Segment (H, M); Census Region (N, M, S)
Race 18-25	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Marital Status (3 Levels) (M, WM); **Race/Ethnicity of Householder (B**, H**)**; Percentage Owner Occupied in Segment (H, M); Census Region (N, M, S)
Race 26+	**Percentage Asian/Other Pacific Islander in Segment (H**, **M)**; Percentage Black/African American in Segment (H, M); Age; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Marital Status (4 Levels) (M, W, DS); **Race/Ethnicity of Householder (B**, **H)**; Percentage Owner Occupied in Segment (H, M); Census Region (N, M, S)
Hispanic/Latino Origin 12-17	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; Age Squared; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Census Region (N, M, S)
Hispanic/Latino Origin 18-25	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; Age Cubed; Age Squared; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Marital Status (3 Levels) (M, WM); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Census Region (N, M, S)
Hispanic/Latino Origin 26+	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; Age Cubed; Age Squared; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Marital Status (4 Levels) (M, W, DS); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Census Region (N, M, S)
Hispanic/Latino Group	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; Age Cubed; Age Squared; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender

Table D.2 Summaries for Predictive Mean Models (continued)

Imputation Step	Variables Included in Predictive Mean Model
Education Level 12-17	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); **Age**; **Age Squared**; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)
Education Level 18+	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; Age Cubed; Age Squared; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender
Employment Status 12-25	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender
Employment Status 26+	Percentage Asian/Other Pacific Islander in Segment (H, M); Percentage Black/African American in Segment (H, M); Age; **Age Cubed**; **Age Squared**; **Education Level (4 Levels) (L, HS, SC)**; Percentage Hispanic/Latino in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**
Born in US 12-17	Age; Age Squared; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (8 Levels) (B, AI, A, MR, PR, M, O); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender
Born in US 18-25	Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (8 Levels) (B, AI, A, MR, PR, M, O); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender
Born in US 26+	Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (8 Levels) (B, AI, A, MR, PR, M, O); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender
Age of Entry	Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (8 Levels) (B, AI, A, MR, PR, M, O); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Race/Hispanic Recode (8 Levels) * Gender (MB, MA, MAI, MMR, MPR, MM, MO)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.1 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

D.4 Drug Variables

Exhibit D.2 Definitions of Levels for Categorical Variables

Intermediate Drug-Specific Past Month Indicator
Y: Yes, N: No[1]
Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)
PM: Past Month, PY: Past Year but Not Past Month, LF: Lifetime but Not Past Year, NU: Lifetime Nonuser[1]
Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels)
PM: Past Month, PY: Past Year but Not Past Month, LF: Lifetime but Not Past Year[1]
Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)
PM: Past Month, PY: Past Year but Not Past Month, P3Y: Past 3 Years but Not Past Year, LF: Lifetime but Not Past 3 Years, NU: Lifetime Nonuser[1]
Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels)
PM: Past Month, PY: Past Year but Not Past Month, P3Y: Past 3 Years but Not Past Year, LF: Lifetime but Not Past 3 Years[1]
Age Category
A1: 26-34, A2: 35-49, A3: 50+[1]
Age Squared * Race/Hispanic Recode (4 Levels)
W: Age Squared * Not Hispanic/Latino and White Only,[1] B: Age Squared * Not Hispanic/Latino and Black/African American Only, O: Age Squared * Not Hispanic/Latino and Other, H: Age Squared * Hispanic/Latino
Age * Race/Hispanic Recode (4 Levels)[2]
W: Age * Not Hispanic/Latino and White Only,[1] B: Age * Not Hispanic/Latino and Black/African American Only, O: Age * Not Hispanic/Latino and Other,[1] H: Age * Hispanic/Latino
Cigarette Lifetime Indicator
Y: Yes, N: No[1]
Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Intermediate Drug-Specific Lifetime Indicator
Y: Yes, N: No[1]
Drug-Specific Lifetime Indicator
Y: Yes, N: No[1]
Lifetime State Rank
L: Low State Rank (lowest tertile), M: Middle State Rank (middle tertile), H: High State Rank (highest tertile)[1]
Marital Status (2 Levels)
M: Married, NM: Not Married[1]
Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]
Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]
MSA (3 Levels)
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]
Pipe Recency
PM: Past Month, LF: Lifetime but Not Past Month, NU: Lifetime Nonuser[1]
Race/Hispanic Recode (2 Levels)
W: Not Hispanic/Latino and White Only, O: Hispanic/Latino or Not White Only[1]
Race/Hispanic Recode (4 Levels)[2]
W: Not Hispanic/Latino and White Only,[1] B: Not Hispanic/Latino and Black/African American Only, O: Not Hispanic/Latino and Other,[1] H: Hispanic/Latino
Recency State Rank
L: Low State Rank (lowest tertile), M: Middle State Rank (middle tertile), H: High State Rank (highest tertile)[1]

Exhibit D.2 Definitions of Levels for Categorical Variables (continued)

Census Region N: Northeast, M: Midwest, S: South, W: West[1]
Gender M: Male, F: Female[1]
Gender * Race/Hispanic Recode (4 Levels)[2] MW: Male Not Hispanic/Latino and White Only,[1] MB: Male Not Hispanic/Latino and Black/African American Only, MO: Male Not Hispanic/Latino and Other,[1] MH: Male Hispanic/Latino, FW: Female Not Hispanic/Latino and White Only,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FO: Female Not Hispanic/Latino and Other,[1] FH: Female Hispanic/Latino[1]

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

[2] For the vast majority of drug models, the reference cell for the four-level race/Hispanic recode was "Not Hispanic/Latino and White Only." For a few models, the reference cell was "Not Hispanic/Latino and Other." The atypical models were those for which statisticians were encouraged to include the "Not Hispanic/Latino and White Only" level as a good covariate, because of past performance. In general, the choice of "Not Hispanic/Latino and White Only" as the reference cell is slightly preferable because it facilitates comparisons to the most populous category in the United States. Interactions involving the race/Hispanic recode are handled similarly.

Table D.3 Lifetime Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12-17	Cigarette Lifetime Indicator (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
18-25	Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
26+	Age Category (A1, A2); Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.4 Cigarettes: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Daily Users: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH) Nondaily Users: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.4 Cigarettes: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigarettes; Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Ever Daily Used	Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigarettes; Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Age at First Use for Cigarettes; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.5 Cigarettes: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Daily Users: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.5 Cigarettes: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Nondaily Users: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Age at First Use	Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigarettes; Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.5 Cigarettes: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigarettes; Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Age at First Use for Cigarettes; Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.6 Cigarettes: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	Age Category (A1, A2); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Daily Users: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.6 Cigarettes: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Nondaily Users: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Age at First Use	Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigarettes; Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.6 Cigarettes: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigarettes; Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y); Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Age at First Use for Cigarettes; Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.7 Smokeless Tobacco (Chewing Tobacco and Snuff): People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Lifetime State Rank (L, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Smokeless Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Smokeless Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Age Squared; Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**
	Chewing Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Chewing Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Age Squared; Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**

Table D.7 Smokeless Tobacco (Chewing Tobacco and Snuff): People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Snuff: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Age Squared; Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**
12-Month Frequency	N/A	N/A
30-Day Frequency	Chewing Tobacco: **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and **Pipes (Y)**; MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, **H)**; Census Region (N, M, S); Gender (M)	Chewing Tobacco Daily Users: **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**; **Age**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; **Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), **Hallucinogens (Y)**, Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)** Chewing Tobacco Nondaily Users: Age; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Drug-Specific Lifetime Indicator for Alcohol (Y), Heroin (Y), and Hallucinogens (Y); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)

Table D.7 Smokeless Tobacco (Chewing Tobacco and Snuff): People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**; **Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Census Region (N, M, S); Gender (M)	Snuff Daily Users: **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**; Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; Age Squared; **Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (**W, **B**, H**)**; Recency State Rank (L, M); Census Region (N, M, S); **Gender (M)**; **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)** Snuff Nondaily Users: **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Age Squared; **Drug-Specific Lifetime Indicator for Cigars (Y)**, Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and **Pipes (Y)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); **Gender (M)**; **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**

Table D.7 Smokeless Tobacco (Chewing Tobacco and Snuff): People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Chewing Tobacco (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, and Daily Cigarette Use; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.8 Smokeless Tobacco (Chewing Tobacco and Snuff): People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Smokeless Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Smokeless Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.8 Smokeless Tobacco (Chewing Tobacco and Snuff): People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Chewing Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Chewing Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
	Snuff: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Snuff: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A

Table D.8 Smokeless Tobacco (Chewing Tobacco and Snuff): People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Chewing Tobacco: **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**; **Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, Sedatives (Y), Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Census Region (N, M, S); Gender (M)	Chewing Tobacco Daily Users: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); **Marital Status (3 Levels) (**M, **WM)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); **Gender (M)**; **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)** Chewing Tobacco Nondaily Users: Age; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Drug-Specific Lifetime Indicator for Hallucinogens (Y); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); **Gender (M)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
	Snuff: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); **Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Census Region (N, M, S); Gender (M)	Snuff Daily Users: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**

Table D.8 Smokeless Tobacco (Chewing Tobacco and Snuff): People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Snuff Nondaily Users: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); **Gender (M)**; **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Chewing Tobacco (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, and Daily Cigarette Use; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.9 Smokeless Tobacco (Chewing Tobacco and Snuff): People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Smokeless Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH) Chewing Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, PT, **UN)**; Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Smokeless Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH) Chewing Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (**B, O, **H)**; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (**B, O, **H)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (**MB, MO, **MH)**

Table D.9 Smokeless Tobacco (Chewing Tobacco and Snuff): People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Snuff: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	Chewing Tobacco: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, H); **Census Region (**N, M, **S)**; Gender (M)	Chewing Tobacco Daily Users: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (**MB, MO, **MH)** Chewing Tobacco Nondaily Users: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Race/Hispanic Recode (4 Levels) (B, O, H)

Table D.9 Smokeless Tobacco (Chewing Tobacco and Snuff): People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); **Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, Sedatives (Y), **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); **MSA (3 Levels) (**R, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Census Region (N, M, S); Gender (M)	Snuff Daily Users: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (**MB, MO, **MH)** Snuff Nondaily Users: Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (**MB, MO, **MH)**

Table D.9 Smokeless Tobacco (Chewing Tobacco and Snuff): People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Chewing Tobacco (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, and Daily Cigarette Use; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.10 Cigars: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; **Cigarette Lifetime Indicator (Y)**; Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y); Lifetime State Rank (L, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**; Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; Age Squared; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), **Heroin (Y)**, and Pipes (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.10 Cigars: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigars; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y); Drug-Specific Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Alcohol (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.11 Cigars: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A

Table D.11 Cigars: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigars; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y); Drug-Specific Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Alcohol (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.12 Cigars: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A

Table D.12 Cigars: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**; Age Category (A1, A2); **Drug-Specific Lifetime Indicator for** Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 30-Day Frequency for Cigars; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y); Drug-Specific Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Alcohol (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.13 Pipes: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y); Lifetime State Rank (L, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF**); **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, Alcohol (Y), Inhalants (Y), Marijuana (Y), **Hallucinogens (Y)**, Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels)** (B, **O**, **H)**; **Recency State Rank (L, M)**; **Census Region (N**, **M**, **S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.14 Pipes: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Education Level (4 Levels)** (L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Alcohol (Y), **Inhalants (Y)**, Marijuana (Y), Hallucinogens (Y), **Pain Relievers (Y)**, **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), Crack (Y), **Heroin (Y)**, and Cigars (Y); **Marital Status (3 Levels) (M**, **WM)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, O, **H)**; Recency State Rank (L, M); **Census Region (N**, **M**, **S)**; **Gender (M)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.15 Pipes: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**; **Age Category (A1**, **A2)**; **Education Level (4 Levels)** (L, **HS**, **SC)**; **Employment Status** (FT, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **Marital Status (4 Levels) (M**, **W**, **DS)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels)** (B, **O**, **H)**; **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.16 Alcohol: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.16 Alcohol: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Binge Drinking Frequency	N/A	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Intermediate Alcohol 30-Day Frequency; MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.16 Alcohol: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific 30-Day Frequency for Alcohol; Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.17 Alcohol: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	No Model Used; No Nonrespondents
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.17 Alcohol: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Binge Drinking Frequency	N/A	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Intermediate Alcohol 30-Day Frequency; MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.17 Alcohol: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific 30-Day Frequency for Alcohol; Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Marijuana (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.18 Alcohol: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.18 Alcohol: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.18 Alcohol: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Binge Drinking Frequency	N/A	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Intermediate Alcohol 30-Day Frequency; MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.18 Alcohol: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific 30-Day Frequency for Alcohol; Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Marijuana (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.19 Inhalants: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Month vs. Past Year Not Past Month	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.19 Inhalants: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and Cigarettes (PM, PY, P3Y, LF); **Drug-Specific Lifetime Indicator for** Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Hallucinogens (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, PY, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for** Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Hallucinogens (Y); MSA (3 Levels) (R, SM); **Pipe Recency (PM**, **LF)**; Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	**Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants**; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**, and **Cigarettes (**PM, PY, **P3Y**, **LF)**; **Age**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); MSA (3 Levels) (R, SM); **Pipe Recency (PM**, **LF)**; Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**

Table D.19 Inhalants: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific 30-Day Frequency for Inhalants; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Hallucinogens (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.20 Inhalants: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Marital Status (3 Levels)** (M, **WM)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.20 Inhalants: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), Heroin (Y), and Cigars (Y); **Marital Status (3 Levels) (M**, **WM)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (**B, O, **H)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; Age Squared; Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, **PT**, **UN)**; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Marital Status (3 Levels) (**M, **WM)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; Recency State Rank (L, M); **Census Region (N**, **M**, **S)**; Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (**MW, **MB**, **MH)**
12-Month Frequency	**Intermediate Drug-Specific Past Month Indicator for Inhalants (Y)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, PY, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for** Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, Heroin (Y), and Hallucinogens (Y); **MSA (3 Levels) (R**, **SM)**; Pipe Recency (PM, LF); **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); **Marital Status (3 Levels) (**M, **WM)**; MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.20 Inhalants: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for Marijuana (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Hallucinogens (Y); **MSA (3 Levels) (R**, SM); **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (N**, **M**, **S)**; Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; Age; Education Level (4 Levels) (L, HS, SC); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S)

Table D.20 Inhalants: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific 30-Day Frequency for Inhalants; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Hallucinogens (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.21 Inhalants: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); **Age Squared**; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); **Employment Status (FT, PT, UN)**; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); **Marital Status (4 Levels) (M, W, DS)**; **MSA (3 Levels) (R, SM)**; Race/Hispanic Recode (4 Levels) (B, O, H); **Census Region (N, M, S)**; Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Marital Status (4 Levels) (M, W, DS)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**

Table D.21 Inhalants: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF**); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, Pipes (Y), **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Cigars (Y)**; **Marital Status (4 Levels)** (M, **W**, DS**)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels)** (B, **O**, **H)**; Recency State Rank (L, M); **Census Region (N**, **M**, **S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**; Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; Education Level (4 Levels) (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Marital Status (4 Levels) (**M, **W**, DS**)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); **Recency State Rank (L**, **M)**; Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); **Drug-Specific Lifetime Indicator for Marijuana (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and **Hallucinogens (Y)**; **MSA (3 Levels) (R**, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Age; Education Level (4 Levels) (L); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Marijuana (Y), and Pain Relievers (Y); Race/Hispanic Recode (2 Levels) (W)

Table D.21 Inhalants: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Age Category (A1**, **A2)**; **Drug-Specific Lifetime Indicator for** Marijuana (Y), **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Hallucinogens (Y)**; **MSA (3 Levels) (R**, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (N**, **M**, **S)**; Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; Age; Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Pain Relievers (Y), and Crack (Y)

Table D.21 Inhalants: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific 30-Day Frequency for Inhalants; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Hallucinogens (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.22 Marijuana: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.22 Marijuana: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, P3Y, **LF)**, **Smokeless Tobacco (**PM, PY, P3Y, **LF)**, and **Cigarettes (**PM, PY, P3Y, **LF)**; **Drug-Specific Lifetime Indicator for** Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.22 Marijuana: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific 30-Day Frequency for Marijuana; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.23 Marijuana: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.23 Marijuana: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.23 Marijuana: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific 30-Day Frequency for Marijuana; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pain Relievers (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.24 Marijuana: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; Age Squared; **Cigarette Lifetime Indicator (Y)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); **Lifetime State Rank (L, M)**; Marital Status (4 Levels) (M, W, DS); **MSA (3 Levels) (R, SM)**; Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.24 Marijuana: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Inhalants (PM, PY, LF)**; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.24 Marijuana: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific 30-Day Frequency for Marijuana; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pain Relievers (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.25 Hallucinogens: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); **Age Squared**; **Cigarette Lifetime Indicator (Y)**; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, Alcohol (Y), Inhalants (Y), Marijuana (Y), **Chewing Tobacco (Y)**, **Snuff (Y)**, and **Pipes (Y)**; **Lifetime State Rank (L, M)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Month vs. Past Year Not Past Month	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; **Recency State Rank (L, M)**; Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**

Table D.25 Hallucinogens: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for** Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Tranquilizers (Y); **MSA (3 Levels) (**R, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and **Cigarettes (**PM, PY, P3Y, **LF)**; Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); **Recency State Rank (L**, M**)**; Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**

Table D.25 Hallucinogens: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific 30-Day Frequency for Hallucinogens; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.26 Hallucinogens: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.26 Hallucinogens: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.26 Hallucinogens: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, Marijuana (PM, PY, LF), and **Inhalants (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigars (PM, PY, P3Y, LF), **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Census Region (N, M, S); Gender (M)	**Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Age Squared; Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, PT, **UN)**; Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); **Marital Status (3 Levels) (**M, **WM)**; **MSA (3 Levels) (**R, **SM)**; Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.26 Hallucinogens: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific 30-Day Frequency for Hallucinogens; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Tranquilizers (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.27 Hallucinogens: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; **Cigarette Lifetime Indicator (Y)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, **Chewing Tobacco (Y)**, Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); **Marital Status (4 Levels) (M, W, DS)**; **MSA (3 Levels) (R, SM)**; Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.27 Hallucinogens: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), **Cocaine (Y)**, **Crack (Y)**, Heroin (Y), and Cigars (Y); **Marital Status (4 Levels) (M**, **W**, **DS)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); **Age Category (A1**, **A2)**; **Drug-Specific Lifetime Indicator for** Pain Relievers (Y), Stimulants (Y), Sedatives (Y), **Cocaine (Y)**, Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY, LF); Education Level (4 Levels) (L, HS, SC); Drug-Specific Lifetime Indicator for Tranquilizers (Y), and Stimulants (Y); Race/Hispanic Recode (4 Levels) (B, O, H)

Table D.27 Hallucinogens: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Marijuana (**PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); **Drug-Specific Lifetime Indicator for** Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Tranquilizers (Y); **MSA (3 Levels) (R**, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (**N, **M**, **S)**; Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; Age; Age Cubed; Age Squared; Marital Status (2 Levels) (M); Pipe Recency (PM); Race/Hispanic Recode (2 Levels) (W)

Table D.27 Hallucinogens: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific 30-Day Frequency for Hallucinogens; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Tranquilizers (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.28 Pain Relievers: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.28 Pain Relievers: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Inhalants (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, and Hallucinogens (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (**PM, PY, P3Y, **LF)**, and Cigarettes (PM, PY, P3Y, LF); **Drug-Specific Lifetime Indicator for Tranquilizers (Y)**, Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and **Stimulants (Y)**; MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	N/A	N/A

Table D.28 Pain Relievers: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.29 Pain Relievers: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.29 Pain Relievers: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	N/A	N/A

Table D.29 Pain Relievers: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.30 Pain Relievers: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.30 Pain Relievers: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	N/A	N/A

Table D.30 Pain Relievers: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Hallucinogens (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.31 Tranquilizers: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.31 Tranquilizers: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and **Marijuana (**PM, PY, **LF)**; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Age Squared; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); **Recency State Rank (L**, **M)**; Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and **Pain Relievers (**PM, PY, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, P3Y, **LF)**, **Smokeless Tobacco (**PM, PY, P3Y, **LF)**, and **Cigarettes (**PM, PY, P3Y, **LF)**; Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	N/A	N/A

Table D.31 Tranquilizers: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.32 **Tranquilizers: People 18 to 25 Years Old**

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.32 Tranquilizers: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Month vs. Past Year Not Past Month	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.32 Tranquilizers: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	N/A	N/A

Table D.32 Tranquilizers: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.33 Tranquilizers: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.33 Tranquilizers: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); **Age Category (**A1, **A2)**; Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, PT, **UN)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), **Inhalants (**PM, **PY**, **LF)**, Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); MSA (3 Levels) (R, SM); **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (**B, **O**, H**)**; Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	N/A	N/A

Table D.33 Tranquilizers: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.34 Stimulants: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; **Cigarette Lifetime Indicator (Y)**; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); **Lifetime State Rank (L, M)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.34 Stimulants: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, **Inhalants (Y)**, Hallucinogens (Y), **Pain Relievers (Y)**, Tranquilizers (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (**B, O, **H)**; Recency State Rank (L, M); **Census Region (N**, **M**, **S)**; **Gender (M)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (**B, **O**, H**)**; Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (**B, **O**, H**)**; **Recency State Rank (L**, **M)**; Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MO**, MH**)**
12-Month Frequency	**Intermediate Drug-Specific Past Month Indicator for Stimulants (Y)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Hallucinogens (**PM, PY, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana (**PM, **PY**, **LF)**, **Tranquilizers (**PM, PY, **LF)**, and Pain Relievers (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, P3Y, **LF)**, **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, PY, P3Y, **LF)**; **Drug-Specific Lifetime Indicator for** Sedatives (Y), **Crack (Y)**, **Heroin (Y)**, and Cocaine (Y); **MSA (3 Levels) (R**, **SM)**; **Pipe Recency (PM**, **LF)**; Race/Hispanic Recode (4 Levels) (B, O, H); **Census Region (N**, M, S**)**; Gender (M)	Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, P3Y, **LF)**, Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	N/A	N/A

Table D.34 Stimulants: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, **P3Y**, **LF)**, **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, PY, **P3Y**, **LF)**; Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); MSA (3 Levels) (R, SM); **Pipe Recency (PM**, **LF)**; Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), **Tranquilizers (PM**, **PY**, **LF)**, and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigars (PM, PY, P3Y, LF), **Smokeless Tobacco (**PM, PY, P3Y, **LF)**, and Cigarettes (PM, PY, P3Y, LF); **Drug-Specific Age at First Use for Cigarettes**, Cigars, **Alcohol**, Inhalants, Marijuana, Hallucinogens, Pain Relievers, **Tranquilizers**, Daily Cigarette Use, and **Smokeless Tobacco**; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.35 Stimulants: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.35 Stimulants: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
30-Day Frequency	N/A	N/A

Table D.35 Stimulants: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.36 Stimulants: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.36 Stimulants: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**; Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), Sedatives (Y), **Cocaine (Y)**, **Crack (Y)**, Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, and Marijuana (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**; Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, **Inhalants (Y)**, Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), **Heroin (Y)**, and **Cigars (Y)**; Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); **Census Region (N**, **M**, **S)**; Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), **Inhalants (**PM, **PY**, **LF)**, Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, P3Y, **LF)**, **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**, and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); **Drug-Specific Lifetime Indicator for** Sedatives (Y), **Crack (Y)**, Heroin (Y), and Cocaine (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM); Age; Age Squared; Drug-Specific Lifetime Indicator for Heroin (Y)
30-Day Frequency	N/A	N/A

Table D.36 Stimulants: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.37 Sedatives: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, Marijuana (Y), Hallucinogens (Y), **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); **Lifetime State Rank (L, M)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Recency State Rank (L, M); **Census Region (N**, **M**, **S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (**PM, **PY**, **LF)**; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **MSA (3 Levels) (R**, **SM)**; Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.37 Sedatives: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana** (PM, PY, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Inhalants (Y), Hallucinogens (Y), **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels)** (B, **O**, H); **Recency State Rank (**L, **M)**; **Census Region (N**, **M**, **S)**; **Gender (M)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Age**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, H); **Recency State Rank (L**, **M)**; Census Region (N, M, S); Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Inhalants (**PM, PY, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; **MSA (3 Levels) (R**, SM**)**; Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); **Census Region (N**, M, S**)**; Gender (M)	**Intermediate Drug-Specific Past Month Indicator for Sedatives (Y)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Hallucinogens (**PM, PY, **LF)**, Inhalants (PM, PY, LF), **Marijuana (**PM, PY, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Age**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; **Drug-Specific Lifetime Indicator for** Cocaine (Y), **Heroin (Y)**, and **Crack (Y)**; **MSA (3 Levels) (R**, **SM)**; Pipe Recency (PM, LF); **Race/Hispanic Recode (4 Levels) (B**, **O**, H); **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
30-Day Frequency	N/A	N/A

Table D.37 Sedatives: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Stimulants** (PM, PY, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers** (PM, PY, **LF)**; **Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives** (PM, **PY)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars** (PM, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes** (PM, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; MSA (3 Levels) (R, SM); **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, O, **H)**; **Census Region (**N, M, **S)**; **Gender (M)**	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), **Inhalants (PM**, **PY**, **LF)**, **Marijuana (**PM, PY, **LF)**, Stimulants (PM, PY, LF), **Tranquilizers** (PM, **PY**, **LF)**, and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars** (PM, PY, P3Y, **LF)**, **Smokeless Tobacco** (PM, **PY**, **P3Y**, **LF)**, and **Cigarettes** (PM, PY, P3Y, **LF)**; **Drug-Specific Age at First Use for Cigarettes**, **Cigars**, Alcohol, **Inhalants**, **Marijuana**, Hallucinogens, Pain Relievers, **Tranquilizers**, Stimulants, **Daily Cigarette Use**, and **Smokeless Tobacco**; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), Daily Cigarette Use (Y), and Crack (Y); **MSA (3 Levels) (R**, **SM)**; Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.38 Sedatives: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana** (PM, **PY**, **LF)**; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); **Education Level (4 Levels)** (L, HS, **SC)**; Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), **Inhalants (Y)**, **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Cigars (Y)**; **Marital Status (3 Levels) (M, WM)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels)** (B, O, **H)**; Recency State Rank (L, M); **Census Region** (N, M, **S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB, MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM**, **PY**, **LF)**; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Marital Status (3 Levels)** (M, **WM)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**

Table D.38 Sedatives: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Marijuana (PM, PY, LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**; **Education Level (4 Levels)** (L, **HS**, **SC**); **Employment Status** (FT, **PT**, **UN**); **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **Marital Status (3 Levels) (M, WM)**; **MSA (3 Levels) (R, SM)**; **Race/Hispanic Recode (4 Levels)** (B, **O**, **H**); Recency State Rank (L, M); **Census Region** (N, M, **S**); **Gender (M)**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Marijuana (PM, PY, LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**; Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; Age Squared; **Education Level (4 Levels) (L, HS, SC)**; Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), **Inhalants (Y)**, **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **Marital Status (3 Levels) (M, WM)**; **MSA (3 Levels) (R, SM)**; **Race/Hispanic Recode (4 Levels) (B**, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**
12-Month Frequency	**Intermediate Drug-Specific Past Month Indicator for Sedatives (Y)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF)**, **Hallucinogens (PM, PY, LF)**, **Inhalants (PM, PY, LF)**, **Marijuana** (PM, **PY, LF)**, **Stimulants (PM, PY, LF)**, **Tranquilizers (PM, PY, LF)**, and **Pain Relievers (PM, PY, LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF)**, **Smokeless Tobacco (PM, PY, P3Y, LF)**, and **Cigarettes (PM, PY, P3Y, LF)**; **Drug-Specific Lifetime Indicator for Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; **MSA (3 Levels) (R, SM)**; **Pipe Recency (PM, LF)**; **Race/Hispanic Recode (4 Levels) (B, O, H)**; **Census Region** (N, M, **S)**; Gender (M)	Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF)**, **Hallucinogens** (PM, **PY**, **LF)**, **Inhalants (PM, PY, LF)**, Marijuana (PM, PY, LF), **Stimulants (PM, PY, LF)**, **Tranquilizers** (PM, PY, **LF)**, and Pain Relievers (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigars (PM, PY, P3Y, LF), **Smokeless Tobacco (PM, PY, P3Y, LF)**, and **Cigarettes** (PM, PY, **P3Y**, **LF)**; **Age**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; **Education Level (4 Levels) (L, HS, SC)**; Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; **Marital Status (3 Levels) (M, WM)**; MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); **Race/Hispanic Recode (4 Levels)** (B, O, **H)**; **Recency State Rank** (L, **M)**; **Census Region** (N, **M**, **S)**; Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**
30-Day Frequency	N/A	N/A

Table D.38 Sedatives: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens** (PM, **PY**, **LF**), **Inhalants (PM, PY, LF)**, **Marijuana** (PM, **PY**, **LF**), Stimulants (PM, PY, LF), **Tranquilizers (PM, PY, LF)**, and Pain Relievers (PM, PY, LF); **Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF)**, and **Cigarettes** (PM, **PY**, **P3Y**, **LF**); Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), and Crack (Y); **MSA (3 Levels) (R, SM)**; **Pipe Recency (PM, LF)**; **Race/Hispanic Recode (4 Levels) (B**, O, **H)**; **Census Region (**N, **M**, **S)**; Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), Daily Cigarette Use (Y), and Crack (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.39 Sedatives: People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); **Lifetime State Rank (L, M)**; Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels)** (B, **O**, H**)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; **Education Level (4 Levels) (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Marital Status (4 Levels) (M**, **W**, **DS)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.39 Sedatives: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**; **Age Category (A1**, **A2)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), **Cocaine (Y)**, **Crack (Y)**, Heroin (Y), and **Cigars (Y)**; **Marital Status (4 Levels) (M, W, DS)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Recency State Rank (L, M); **Census Region (N**, **M**, **S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana** (PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**; Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and **Cigars (Y)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (**B, **O**, H); Recency State Rank (L, M); **Census Region (N**, **M**, **S)**; Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Hallucinogens (PM**, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Stimulants** (PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes** (PM, **PY**, **P3Y**, **LF)**; **Age Category (A1**, **A2)**; Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), and Crack (Y); **MSA (3 Levels) (R**, **SM)**; **Pipe Recency** (PM, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Cocaine (Y); Marital Status (2 Levels) (M); Gender (M)
30-Day Frequency	N/A	N/A

Table D.39 Sedatives: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), and Crack (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), Daily Cigarette Use (Y), and Crack (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.40 Cocaine (and Crack): People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Cocaine: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; **Cigarette Lifetime Indicator (Y)**; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, Chewing Tobacco (Y), Snuff (Y), and **Pipes (Y)**; **Lifetime State Rank (L**, **M)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH) Crack: Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, O, H); Age Squared; Cigarette Lifetime Indicator (Y); **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), Sedatives (Y), **Chewing Tobacco (Y)**, **Snuff (Y)**, and **Pipes (Y)**; **Lifetime State Rank (L**, **M)**; MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; **Census Region (N**, M, S**)**; Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**

Table D.40 Cocaine (and Crack): People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), **Inhalants (Y)**, Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Recency State Rank (L**, **M)**; Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, Sedatives (Y), Crack (Y), Heroin (Y), and **Cigars (Y)**; MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Hallucinogens (**PM, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, Marijuana (PM, PY, LF), **Sedatives (PM**, **PY**, **LF)**, Stimulants (PM, PY, LF), **Tranquilizers (**PM, **PY**, **LF)**, and **Pain Relievers (**PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for Crack (Y)**, and Heroin (Y); **MSA (3 Levels) (**R, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, H**)**; Census Region (N, M, S); **Gender (M)**	Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, Hallucinogens (PM, PY, LF), **Inhalants (PM**, **PY**, **LF)**, **Marijuana (**PM, PY, **LF)**, **Sedatives (PM**, **PY**, **LF)**, Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); **MSA (3 Levels) (R**, **SM)**; Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**

Table D.40 Cocaine (and Crack): People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Hallucinogens (**PM, PY, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana (**PM, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for Crack (Y)**, and **Heroin (Y)**; **MSA (3 Levels) (R**, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (N**, **M**, **S)**; **Gender (M)**	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, P3Y, **LF)**; Age; Census Region (N, M, S)

Table D.40 Cocaine (and Crack): People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants** (PM, **PY, LF**), Marijuana (PM, PY, LF), **Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF)**, and **Pain Relievers (PM, PY, LF)**; **Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars** (PM, **PY, P3Y, LF**), **Smokeless Tobacco** (PM, **PY, P3Y, LF**), and **Cigarettes** (PM, PY, **P3Y, LF**); Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); **MSA (3 Levels) (R, SM)**; **Pipe Recency (PM, LF)**; **Race/Hispanic Recode (4 Levels)** (B, **O**, **H**); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; Drug-Specific 30-Day Frequency for Cocaine; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

Table D.41 Cocaine (and Crack): People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Cocaine: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH) Crack: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.41 Cocaine (and Crack): People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.41 Cocaine (and Crack): People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana** (PM, PY, **LF)**, **Sedatives (PM**, **PY**, **LF)**, Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, Smokeless Tobacco (PM, PY, P3Y, LF), and **Cigarettes** (PM, **PY**, **P3Y**, **LF)**; Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); **MSA (3 Levels) (R**, **SM)**; Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	**Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), **Inhalants (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and Cigarettes (PM, PY, P3Y, LF); Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; **Education Level (4 Levels)** (L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for** Crack (Y), and **Heroin (Y)**; Marital Status (3 Levels) (M, WM); **MSA (3 Levels)** (R, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels)** (B, O, **H)**; **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**

Table D.41 Cocaine (and Crack): People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; Drug-Specific 30-Day Frequency for Cocaine; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.42 Cocaine (and Crack): People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Cocaine: Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; Age Squared; **Cigarette Lifetime Indicator (Y)**; **Education Level (4 Levels) (L, HS, SC)**; Employment Status (FT, PT, UN); **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, Marijuana (Y), Hallucinogens (Y), **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); **Marital Status (4 Levels) (M, W, DS)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)** Crack: Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.42 Cocaine (and Crack): People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Marital Status (4 Levels) (M**, **W**, **DS)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); **Census Region (**N, **M**, **S)**; Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), **Inhalants (**PM, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Stimulants (**PM, **PY**, **LF)**, Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.42 Cocaine (and Crack): People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Hallucinogens (**PM, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana (**PM, PY, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (**PM, **PY**, **LF)**, and **Pain Relievers (**PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and **Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Age Category (A1**, A2); **Drug-Specific Lifetime Indicator for Crack (Y)**, and **Heroin (Y)**; **MSA (3 Levels)** (R, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (N**, **M**, **S)**; **Gender (M)**	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; Age; Drug-Specific Lifetime Indicator for Crack (Y); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.42 Cocaine (and Crack): People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; Drug-Specific 30-Day Frequency for Cocaine; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.43 Heroin: People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	**Age**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; **Cigarette Lifetime Indicator (Y)**; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Chewing Tobacco (Y)**, **Snuff (Y)**, and **Pipes (Y)**; **Lifetime State Rank (L**, **M)**; MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels)** (B, **O**, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels)** (MB, **MO**, MH)
Recency: Past Year vs. Not Past Year	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana** (PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, Sedatives (Y), **Cocaine (Y)**, **Crack (Y)**, and **Cigars (Y)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**; Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, **Inhalants (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, and **Cigars (Y)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; **Gender (M)**; **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**

Table D.43 Heroin: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Marijuana (PM, PY, LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, and **Cigars (Y)**; **MSA (3 Levels) (R, SM)**; **Race/Hispanic Recode (4 Levels) (B, O, H)**; **Recency State Rank (L, M)**; **Census Region (**N, **M, S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Marijuana** (PM, **PY, LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY, P3Y, LF)**; Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; **Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, **Inhalants (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), and **Cigars (Y)**; **MSA (3 Levels) (R, SM)**; **Race/Hispanic Recode (4 Levels) (B, O, H)**; **Recency State Rank (L, M)**; **Census Region (N, M, S)**; **Gender (M)**; **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**
12-Month Frequency	**Intermediate Drug-Specific Past Month Indicator for Heroin (Y)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, **Cocaine (PM, PY, LF)**, **Crack (PM, PY, LF)**, **Hallucinogens (PM, PY, LF)**, **Inhalants (PM, PY, LF)**, **Marijuana** (PM, **PY, LF)**, **Sedatives (PM, PY, LF)**, **Stimulants (PM, PY, LF)**, **Tranquilizers (PM, PY, LF)**, and **Pain Relievers (PM, PY, LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF)**, **Smokeless Tobacco (PM, PY, P3Y, LF)**, and **Cigarettes** (PM, **PY, P3Y, LF)**; **MSA (3 Levels) (R, SM)**; Pipe Recency (PM, LF); **Race/Hispanic Recode (4 Levels) (B, O, H)**; **Census Region (N, M, S)**; **Gender (M)**	**Intermediate Drug-Specific Past Month Indicator for Heroin (Y)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana** (PM, **PY, LF)**; Age; **Race/Hispanic Recode (4 Levels) (B**, O, H)

Table D.43 Heroin: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (PM, PY**, **LF)**, and **Pain Relievers (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (PM**, **PY**, **P3Y**, **LF)**; **MSA (3 Levels) (R**, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (N**, **M**, **S)**; **Gender (M)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Crack (PM), Sedatives (PM), and Marijuana (PM); Age

Table D.43 Heroin: People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Cocaine (PM, PY, LF), Crack (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants** (PM, **PY, LF**), **Marijuana** (PM, **PY, LF**), **Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF)**, and **Pain Relievers (PM, PY, LF)**; **Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco** (PM, **PY, P3Y, LF**), and **Cigarettes (PM, PY, P3Y, LF)**; **MSA (3 Levels) (R, SM)**; **Pipe Recency (PM, LF)**; **Race/Hispanic Recode (4 Levels) (B, O, H)**; **Census Region (N, M, S)**; **Gender (M)**	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin; Drug-Specific 30-Day Frequency for Heroin; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Cocaine (PM, PY, LF), Crack (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF)**, and **Pain Relievers (PM, PY, LF)**; **Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars** (PM, PY, **P3Y, LF**), **Smokeless Tobacco** (PM, **PY, P3Y, LF**), and **Cigarettes** (PM, **PY, P3Y, LF**); **Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana**, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, **Sedatives, Cocaine, Crack, Daily Cigarette Use**, and **Smokeless Tobacco**; **Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; **Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y)**; **MSA (3 Levels) (R, SM)**; **Pipe Recency** (PM, **LF**); **Race/Hispanic Recode (4 Levels) (B, O, H)**; **Recency State Rank (L, M)**; **Census Region (N, M, S)**; **Gender (M)**; **Age * Gender**; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.44 Heroin: People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; **Cigarette Lifetime Indicator (Y)**; **Education Level (4 Levels) (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Chewing Tobacco (Y)**, **Snuff (Y)**, and **Pipes (Y)**; Lifetime State Rank (L, M); **Marital Status (3 Levels) (M, WM)**; **MSA (3 Levels) (R, SM)**; Race/Hispanic Recode (4 Levels) (B, O, H); **Census Region (N, M, S)**; Gender (M); **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**
Recency: Past Year vs. Not Past Year	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF**), and **Marijuana** (PM, **PY**, **LF**); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF**); Education Level (4 Levels) (L, HS, SC); **Employment Status** (FT, **PT**, **UN**); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), and Cigars (Y); **Marital Status (3 Levels) (M, WM)**; MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, **H)**; **Recency State Rank (L, M)**; Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB, MO**, MH**)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and Marijuana (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF**); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.44 Heroin: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (**FT, **PT**, **UN)**; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), and Cigars (Y); **Marital Status (3 Levels) (M**, **WM)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, H); **Recency State Rank (L**, **M)**; **Census Region (**N, **M**, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, MH)	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and Marijuana (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**; Age; Age Cubed; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (**B, **O**, **H)**; **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Cocaine (PM, PY, LF), **Crack (**PM, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigars (PM, PY, P3Y, LF), **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, **PY**, **P3Y**, **LF)**; MSA (3 Levels) (R, SM); **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, H); Census Region (N, M, S); Gender (M)	Intermediate Drug-Specific Past Month Indicator for Heroin (Y); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Crack (PM, PY, LF), Tranquilizers (PM, PY, LF), and Marijuana (PM, PY, LF); Marital Status (3 Levels) (M, WM)

Table D.44 Heroin: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana** (PM, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars** (PM, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, **PY**, **P3Y**, **LF)**; **MSA (3 Levels) (R**, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (N**, **M**, **S)**; **Gender (M)**	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Crack (PM), and Tranquilizers (PM); **MSA (3 Levels) (R)**; Race/Hispanic Recode (2 Levels) (W)

Table D.44 Heroin: People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF**), **Cocaine** (PM, **PY**, **LF**), **Crack** (PM, **PY**, **LF**), **Hallucinogens** (PM, **PY**, **LF**), **Inhalants** (PM, **PY**, **LF**), **Marijuana** (PM, PY, **LF**), **Sedatives (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers (PM**, **PY**, **LF)**; Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM**, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes** (PM, PY, **P3Y**, **LF)**; MSA (3 Levels) (R, SM); **Pipe Recency (PM**, **LF)**; Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin; Drug-Specific 30-Day Frequency for Heroin; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Cocaine (PM, PY, LF), Crack (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, Daily Cigarette Use, and Smokeless Tobacco; Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.45 Heroin: People 26 Years or Older

Imputation Step	**Variables Included in Response Propensity Model**	**Variables Included in Drug Model**
Lifetime	See Table D.3.	Age; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); **Age Squared**; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); **Lifetime State Rank (L, M)**; Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); **Gender (M)**; **Age * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**
Recency: Past Year vs. Not Past Year	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF**), and Marijuana (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF**); Age Category (A1, A2); **Education Level (4 Levels) (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), Crack (Y), and Cigars (Y); **Marital Status (4 Levels) (M**, **W**, **DS)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels)** (B, **O**, **H)**; Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.45 Heroin: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), Crack (Y), and **Cigars (Y)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Recency State Rank (L**, **M)**; **Census Region (N**, **M**, **S)**; Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; **Education Level (4 Levels) (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), **Pain Relievers (Y)**, Tranquilizers (Y), Stimulants (Y), Sedatives (Y), **Cocaine (Y)**, Crack (Y), and **Cigars (Y)**; Marital Status (4 Levels) (M, W, DS); **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Recency State Rank (L**, **M)**; Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Inhalants (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, and **Pain Relievers (**PM, PY, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**, and **Cigarettes (**PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); **MSA (3 Levels) (R**, **SM)**; **Pipe Recency (PM**, **LF)**; **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (N**, **M**, **S)**; Gender (M)	**Intermediate Drug-Specific Past Month Indicator for Heroin (Y)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Cocaine (**PM, PY, **LF)**, and **Marijuana (**PM, PY, **LF)**; Marital Status (2 Levels) (M); Race/Hispanic Recode (2 Levels) (W)

Table D.45 Heroin: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, **Cocaine (PM, PY, LF)**, **Crack (PM, PY, LF)**, **Hallucinogens (PM, PY, LF)**, **Inhalants (PM, PY, LF)**, **Marijuana (PM, PY, LF)**, **Sedatives (PM, PY, LF)**, **Stimulants (PM, PY, LF)**, **Tranquilizers (PM, PY, LF)**, and **Pain Relievers (**PM, PY, **LF)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, **PY, P3Y, LF)**, **Smokeless Tobacco (PM, PY, P3Y, LF)**, and **Cigarettes (PM, PY, P3Y, LF)**; **Age Category (A1, A2)**; **MSA (3 Levels) (R, SM)**; **Pipe Recency (PM, LF)**; **Race/Hispanic Recode (4 Levels) (B, O,** H); **Census Region (**N, **M, S)**; Gender (M)	Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Cocaine (PM); Employment Status (FT); Gender (M)

Table D.45 Heroin: People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, **Cocaine (PM, PY, LF)**, **Crack (PM, PY, LF)**, **Hallucinogens** (PM, **PY**, **LF)**, **Inhalants (PM, PY, LF)**, Marijuana (PM, PY, LF), **Sedatives (PM, PY, LF)**, **Stimulants** (PM, **PY**, **LF)**, **Tranquilizers (PM, PY, LF)**, and **Pain Relievers** (PM, **PY**, **LF)**; **Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY)**; **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars** (PM, **PY**, **P3Y**, **LF)**, **Smokeless Tobacco (PM, PY, P3Y, LF)**, and **Cigarettes** (PM, **PY**, **P3Y**, **LF)**; Age Category (A1, A2); MSA (3 Levels) (R, SM); **Pipe Recency (PM, LF)**; **Race/Hispanic Recode (4 Levels)** (B, **O**, H); Census Region (N, M, S); Gender (M)	Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin; Drug-Specific 30-Day Frequency for Heroin; Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Cocaine (PM, PY, LF), Crack (PM, PY, LF), Hallucinogens (PM, PY, LF), Inhalants (PM, PY, LF), Marijuana (PM, PY, LF), Sedatives (PM, PY, LF), Stimulants (PM, PY, LF), Tranquilizers (PM, PY, LF), and Pain Relievers (PM, PY, LF); Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y, LF), Smokeless Tobacco (PM, PY, P3Y, LF), and Cigarettes (PM, PY, P3Y, LF); Drug-Specific Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, Daily Cigarette Use, and Smokeless Tobacco; Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Pipe Recency (PM, LF); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.46 Stimulants (Core plus Noncore): People 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; **Cigarette Lifetime Indicator (Y)**; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); **Lifetime State Rank (L, M)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.46 Stimulants (Core plus Noncore): People 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana** (PM, **PY**, **LF**); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, P3Y, **LF**); **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, **Inhalants (Y)**, Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), **Cocaine (Y)**, **Crack (Y)**, Heroin (Y), and **Cigars (Y)**; **MSA (3 Levels)** (R, **SM**); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels)** **(**B, **O**, H**)**; Age Squared; Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels)** (B, **O**, H); **Recency State Rank (L, M)**; Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels)** **(**MB, **MO**, MH**)**
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.47 Stimulants (Core plus Noncore): People 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.47 Stimulants (Core plus Noncore): People 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (3 Levels) (M, WM); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.48 Stimulants (Core plus Noncore): People 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Cigarette Lifetime Indicator (Y); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y); Lifetime State Rank (L, M); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Recency: Past Year vs. Not Past Year	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; Age Cubed; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); Age * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.48 Stimulants (Core plus Noncore): People 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); **Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF)**; Age Category (A1, A2); Education Level (4 Levels) (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), **Pain Relievers (Y)**, Tranquilizers (Y), **Sedatives (Y)**, Cocaine (Y), Crack (Y), Heroin (Y), and **Cigars (Y)**; Marital Status (4 Levels) (M, W, DS); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Recency State Rank (L, M); Census Region (N, M, S); Gender (M); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	**Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, and Marijuana (PM, PY, LF); Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Age; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and **Cigars (Y)**; Marital Status (4 Levels) (M, W, DS); **MSA (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; **Recency State Rank** (L, **M)**; Census Region (N, M, S); Gender (M); Age * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

D.5 Household Composition Variables

Exhibit D.3 Definitions of Levels for Categorical Variables

Age Squared * Race/Hispanic Recode (4 Levels)
W: Age Squared * Not Hispanic/Latino and White Only,[1] B: Age Squared * Not Hispanic/Latino and Black/African American Only, O: Age Squared * Not Hispanic/Latino and Other, H: Age Squared * Hispanic/Latino

Age * Race/Hispanic Recode (4 Levels)
W: Age * Not Hispanic/Latino and White Only,[1] B: Age * Not Hispanic/Latino and Black/African American Only, O: Age * Not Hispanic/Latino and Other, H: Age * Hispanic/Latino

Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]

Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]

Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]

Other Family Members In Household
N: No, Y: Yes[1]

Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]

Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]

Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]

MSA (3 Levels)
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]

Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only,[1] B: Not Hispanic/Latino and Black/African American Only, O: Not Hispanic/Latino and Other, H: Hispanic/Latino

Census Region
N: Northeast, M: Midwest, S: South, W: West[1]

Gender
M: Male, F: Female[1]

Gender * Race/Hispanic Recode (4 Levels)
MW: Male Not Hispanic/Latino and White Only,[1] MB: Male Not Hispanic/Latino and Black/African American Only, MO: Male Not Hispanic/Latino and Other, MH: Male Hispanic/Latino, FW: Female Not Hispanic/Latino and White Only,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FO: Female Not Hispanic/Latino and Other,[1] FH: Female Hispanic/Latino[1]

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.49 Household Composition: People 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Household Size (TOTPEOP)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Total People in Household (Screener)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Total People in Household (Screener)
Number of People Younger than 18 in Household (KID17)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Household Size; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Household Size; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Number of Eligible 12 to 17 in Household (Screener)
Number of People Older than 64 in Household (HH65)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, O, H); **Age * Race/Hispanic Recode (4 Levels) (B**, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Household Size; Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Household Size; Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.49 Household Composition: People 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Other Family Present in Household (FAMSKIP)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; Age Squared; **Percentage Hispanic/Latino in Segment (H, M)**; Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; **Percentage Owner Occupied in Segment (H, M)**; MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H**)**; Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH**)**
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Number of Respondent's Family Members in Household Younger than 18 including Foster Relationships (KIDFMLY)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.49 Household Composition: People 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H)**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); **Age Squared**; Percentage Hispanic/Latino in Segment (H, M); **Other Family Members In Household (N)**; Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Number of Respondent's Family Members in Household Younger than 18 excluding Foster Relationships (KIDFAMSZ)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Excludes Foster Relationships); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); **Number of People Younger than 18 Years Old in Household**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Excludes Foster Relationships); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.3 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.50 Household Composition: People 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Household Size (TOTPEOP)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, **PT**, **UN)**; Percentage Hispanic/Latino in Segment (H, M); **Marital Status (3 Levels) (M**, **WM)**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Total People in Household (Screener)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Total People in Household (Screener)
Number of People Younger than 18 in Household (KID17)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Number of Eligible 12 to 17 in Household (Screener)
Number of People Older than 64 in Household (HH65)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.50 Household Composition: People 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Other Family Present in Household (FAMSKIP)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Number of People Older than 64 Years Old in Household; **Household Size**; **Number of People Younger than 18 Years Old in Household**; **Marital Status (3 Levels) (M, WM)**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); **Number of People Older than 64 Years Old in Household**; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.50 Household Composition: People 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 including Foster Relationships (KIDFMLY)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); **Number of People Older than 64 Years Old in Household**; **Household Size**; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; **Marital Status (3 Levels) (**M, **WM)**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.50 Household Composition: People 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 excluding Foster Relationships (KIDFAMSZ)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Excludes Foster Relationships); **Number of Respondent's Family Members in Household (Includes Foster Relationships)**; Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Excludes Foster Relationships); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.3 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.51 Household Composition: People 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Household Size (TOTPEOP)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Total People in Household (Screener)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (**B, **O**, H); **Age * Race/Hispanic Recode (4 Levels) (**B, **O**, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Total People in Household (Screener)
Number of People Younger than 18 in Household (KID17)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; **Marital Status (4 Levels) (M**, **W**, **DS)**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Number of Eligible 12 to 17 in Household (Screener)
Number of People Older than 64 in Household (HH65)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); **Percentage Owner Occupied in Segment (**H, **M)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.51 Household Composition: People 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Other Family Present in Household (FAMSKIP)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Number of People Older than 64 Years Old in Household; **Household Size**; **Number of People Younger than 18 Years Old in Household**; **Marital Status (4 Levels) (M, W, DS)**; **Percentage Owner Occupied in Segment (H, M)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.51 Household Composition: People 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 including Foster Relationships (KIDFMLY)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); **Other Family Members In Household (N)**; **Number of Respondent's Family Members in Household (Includes Foster Relationships)**; Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H)**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.51 Household Composition: People 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 excluding Foster Relationships (KIDFAMSZ)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Percentage Hispanic/Latino in Segment (H, M)**; Other Family Members In Household (N); Number of Respondent's Family Members in Household (Excludes Foster Relationships); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; **Marital Status (4 Levels) (M, W, DS)**; **Percentage Owner Occupied in Segment (H, M)**; **MSA (3 Levels) (R, SM)**; Race/Hispanic Recode (4 Levels) (B, O, H); **Census Region (**N, M, **S)**; Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; **Education Level (4 Levels) (L, HS, SC)**; **Employment Status (FT, PT, UN)**; Percentage Hispanic/Latino in Segment (H, M); **Other Family Members In Household (N)**; **Number of Respondent's Family Members in Household (Excludes Foster Relationships)**; **Number of Respondent's Family Members in Household (Includes Foster Relationships)**; **Number of People Older than 64 Years Old in Household**; **Household Size**; **Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)**; **Number of People Younger than 18 Years Old in Household**; **Marital Status (4 Levels) (M, W, DS)**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.3 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.52 Household Composition: People 65 Years or Older

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Household Size (TOTPEOP)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, **O**, H**)**; Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**; Total People in Household (Screener)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (**B, **O**, H**)**; **Age * Race/Hispanic Recode (4 Levels) (**B, **O**, H**)**; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); **Marital Status (4 Levels) (M**, **W**, **DS)**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (**B, **O**, H**)**; Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MO**, MH**)**; Total People in Household (Screener)
Number of People Younger than 18 in Household (KID17)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH); Number of Eligible 12 to 17 in Household (Screener)
Number of People Older than 64 in Household (HH65)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Percentage Hispanic/Latino in Segment (H**, **M)**; Household Size; **Number of People Younger than 18 Years Old in Household**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Census Region (**N, M, **S)**; Gender (M); Age * Gender; Age Squared * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, **MH)**	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MO**, MH**)**

Table D.52 Household Composition: People 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Other Family Present in Household (FAMSKIP)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; **Education Level (4 Levels) (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Percentage Hispanic/Latino in Segment (H, M)**; Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; **Marital Status (4 Levels) (M, W, DS)**; **Percentage Owner Occupied in Segment (H, M)**; **MSA (3 Levels) (R, SM)**; Race/Hispanic Recode (4 Levels) (B, O, H); **Census Region (N, M, S)**; Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); **Number of People Older than 64 Years Old in Household**; **Household Size**; **Number of People Younger than 18 Years Old in Household**; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, O, H); **Age * Race/Hispanic Recode (4 Levels) (B**, O, H); Age Squared; **Education Level (4 Levels) (L, HS, SC)**; Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of People Older than 64 Years Old in Household; **Household Size**; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MO, MH)

Table D.52 Household Composition: People 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 including Foster Relationships (KIDFMLY)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B**, O, **H)**; Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MO**, MH**)**
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, **PT**, **UN)**; Percentage Hispanic/Latino in Segment (H, M); Other Family Members In Household (N); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; **Marital Status (4 Levels) (M**, **W**, **DS)**; Percentage Owner Occupied in Segment (H, M); **MSA (3 Levels) (**R, **SM)**; Race/Hispanic Recode (4 Levels) (B, O, H); **Census Region (**N, **M**, **S)**; Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.52 Household Composition: People 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 excluding Foster Relationships (KIDFAMSZ)	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age * Race/Hispanic Recode (4 Levels) (B, O, H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); **Employment Status (FT, PT, UN)**; **Percentage Hispanic/Latino in Segment (H, M)**; Other Family Members In Household (N); Number of Respondent's Family Members in Household (Excludes Foster Relationships); Number of Respondent's Family Members in Household (Includes Foster Relationships); Number of People Older than 64 Years Old in Household; Household Size; Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (B, O, H)**; Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)**	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); **Other Family Members In Household (N)**; **Number of Respondent's Family Members in Household (Excludes Foster Relationships)**; **Number of Respondent's Family Members in Household (Includes Foster Relationships)**; **Number of People Older than 64 Years Old in Household**; Household Size; **Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)**; **Number of People Younger than 18 Years Old in Household**; **Marital Status (4 Levels) (M, W, DS)**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.3 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

D.6 Income Variables

Exhibit D.4 Definitions of Levels for Categorical Variables

Age Squared * Race/Hispanic Recode (4 Levels)
W: Age Squared * Not Hispanic/Latino and White Only, B: Age Squared * Not Hispanic/Latino and Black/African American Only, O: Age Squared * Not Hispanic/Latino and Other,[1] H: Age Squared * Hispanic/Latino
Age Cubed * Race/Hispanic Recode (4 Levels)
W: Age Cubed * Not Hispanic/Latino and White Only, B: Age Cubed * Not Hispanic/Latino and Black/African American Only, O: Age Cubed * Not Hispanic/Latino and Other,[1] H: Age Cubed * Hispanic/Latino
Age * Race/Hispanic Recode (4 Levels)
W: Age * Not Hispanic/Latino and White Only, B: Age * Not Hispanic/Latino and Black/African American Only, O: Age * Not Hispanic/Latino and Other,[1] H: Age * Hispanic/Latino
Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Intermediate Family Received Food Stamps
Y: Yes, N: No[1]
Intermediate Family Received Public Assistance
Y: Yes, N: No[1]
Intermediate Family Received Social Security or Railroad Payments
Y: Yes, N: No[1]
Intermediate Family Received Supplemental Security Income
Y: Yes, N: No[1]
Intermediate Family Received Welfare/Job Placement/Childcare Services
Y: Yes, N: No[1]
Intermediate Family Received Income from a Job
Y: Yes, N: No[1]
Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]
Income State Rank
L: Low State Income Rank (lowest tertile), M: Middle State Income Rank (middle tertile), H: High State Income Rank (highest tertile)[1]
Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]
Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]
Percentage Non-Hispanic/Latino Black/African American in Segment
H: ≥ 50%, M: [10,50)%, L: < 10%[1]
Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]
MSA (3 Levels)
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]
Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only, B: Not Hispanic/Latino and Black/African American Only, O: Not Hispanic/Latino and Other,[1] H: Hispanic/Latino
Total Family Income > or < $20,000
GT: ≥ $20,000, LT: < $20,000[1]
Family Received Public Assistance
N: No, Y: Yes[1]
Family Received Social Security or Railroad Payments
N: No, Y: Yes[1]

Exhibit D.4 Definitions of Levels for Categorical Variables (continued)

Family Received Supplemental Security Income
N: No, Y: Yes[1]
Family Received Welfare/Job Placement/Childcare Services
N: No, Y: Yes[1]
Family Received Income from a Job
N: No, Y: Yes[1]
Family Received Food Stamps
N: No, Y: Yes[1]
Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
Gender
M: Male, F: Female[1]
Gender * Race/Hispanic Recode (4 Levels)
MW: Male Not Hispanic/Latino and White Only, MB: Male Not Hispanic/Latino and Black/African American Only, MO: Male Not Hispanic/Latino and Other,[1] MH: Male Hispanic/Latino, FW: Female Not Hispanic/Latino and White Only,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FO: Female Not Hispanic/Latino and Other,[1] FH: Female Hispanic/Latino[1]

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.53 Dichotomous Income Indicators in Response Propensity Models

Age Group	Variables Included in Response Propensity (Dichotomous Income Indicators)
12-17	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared**; Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; **Age Cubed * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
18-25	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; **Age Cubed * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
26-64	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; **Age Cubed * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
65+	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); **Number of People Younger than 18 Years Old in Household**; Marital Status (4 Levels) (M, W, DS); **Percentage Non-Hispanic/Latino Black/African American in Segment (**H, **M)**; **Number of Adults in Household**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; **Age Cubed * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.54 Dichotomous Income Indicators in Predictive Mean Models: People 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Supplemental Security	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Intermediate Family Received Social Security or Railroad Payments (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Wages	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Food Stamps	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.54 Dichotomous Income Indicators in Predictive Mean Models: People 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Welfare Services	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
# Welfare Months	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Intermediate Family Received Food Stamps (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.54 Dichotomous Income Indicators in Predictive Mean Models: People 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Intermediate Family Received Food Stamps (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.55 Dichotomous Income Indicators in Predictive Mean Models: People 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); **Number of People Older than 64 Years Old in Household (Continuous)**; Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Supplemental Security	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Social Security or Railroad Payments (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Wages	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Food Stamps	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.55 Dichotomous Income Indicators in Predictive Mean Models: People 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Welfare Services	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
# Welfare Months	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Food Stamps (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.55 Dichotomous Income Indicators in Predictive Mean Models: People 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Food Stamps (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.56 Dichotomous Income Indicators in Predictive Mean Models: People 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Supplemental Security	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Social Security or Railroad Payments (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Wages	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); **Employment Status (FT, PT, UN)**; Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); **Intermediate Family Received Supplemental Security Income (Y)**; Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Food Stamps	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.56 Dichotomous Income Indicators in Predictive Mean Models: People 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Welfare Services	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; **Education Level (4 Levels) (**L, HS, **SC)**; Employment Status (FT, PT, UN); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); **Number of People Older than 64 Years Old in Household (Continuous)**; Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Number of Adults in Household**; **Percentage Owner Occupied in Segment (H, M)**; **MSA (3 Levels) (R, SM)**; Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
# Welfare Months	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Food Stamps (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.56 Dichotomous Income Indicators in Predictive Mean Models: People 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Food Stamps (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.57 Dichotomous Income Indicators in Predictive Mean Models: People 65 Years or Older

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Supplemental Security	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Social Security or Railroad Payments (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Wages	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
Food Stamps	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.57 Dichotomous Income Indicators in Predictive Mean Models: People 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; Intermediate Family Received Social Security or Railroad Payments (Y); **Intermediate Family Received Supplemental Security Income (Y)**; Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); **Number of People Younger than 18 Years Old in Household**; **Marital Status (4 Levels) (**M, W, **DS)**; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Number of Adults in Household**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**
Welfare Services	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); **Intermediate Family Received Public Assistance (Y)**; Intermediate Family Received Social Security or Railroad Payments (Y); **Intermediate Family Received Supplemental Security Income (Y)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; **Percentage Owner Occupied in Segment (H**, **M)**; MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (**MW, MB, **MH)**
# Welfare Months	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (**FT, **PT**, **UN)**; **Intermediate Family Received Food Stamps (Y)**; Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); **Intermediate Family Received Supplemental Security Income (Y)**; Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); **Intermediate Family Received Income from a Job (Y)**; Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (H**, **M)**; Number of Adults in Household; **Percentage Owner Occupied in Segment (H**, **M)**; **MSA (3 Levels) (R**, **SM)**; Race/Hispanic Recode (4 Levels) (W, B, H); **Census Region (N**, **M**, **S)**; **Gender (M)**; **Age * Gender**; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**

Table D.57 Dichotomous Income Indicators in Predictive Mean Models: People 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate Family Received Food Stamps (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

Table D.58 Income Finer Categories in Response Propensity Models

Age Group	Variables Included in Response Propensity for Income Models (Finer Categorization)
12-17	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared**; Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Number of Adults in Household**; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) (W, B, H)**; Total Family Income > or < $20,000 (GT); Family Received Public Assistance (N); Family Received Social Security or Railroad Payments (N); Family Received Supplemental Security Income (N); Family Received Welfare/Job Placement/Childcare Services (N); Family Received Income from a Job (N); **Family Received Food Stamps (N)**; Census Region (N, M, S); Gender (M); Age * Gender; **Age Squared * Gender**; **Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)**
18-25	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Total Family Income > or < $20,000 (GT); Family Received Public Assistance (N); Family Received Social Security or Railroad Payments (N); Family Received Supplemental Security Income (N); Family Received Welfare/Job Placement/Childcare Services (N); Family Received Income from a Job (N); Family Received Food Stamps (N); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
26-64	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); **Age Cubed**; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Total Family Income > or < $20,000 (GT); Family Received Public Assistance (N); Family Received Social Security or Railroad Payments (N); Family Received Supplemental Security Income (N); Family Received Welfare/Job Placement/Childcare Services (N); Family Received Income from a Job (N); Family Received Food Stamps (N); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.58 Income Finer Categories in Response Propensity Models (continued)

Age Group	Variables Included in Response Propensity for Income Models (Finer Categorization)
65+	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Total Family Income > or < $20,000 (GT); Family Received Public Assistance (N); Family Received Social Security or Railroad Payments (N); Family Received Supplemental Security Income (N); Family Received Welfare/Job Placement/Childcare Services (N); Family Received Income from a Job (N); Family Received Food Stamps (N); Census Region (N, M, S); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.59 Income Finer Categories in Predictive Mean Models

Age Group	Variables Included in Income Models (Finer Categorization)
12-17	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Total Family Income > or < $20,000 (GT); Family Received Public Assistance (N); Family Received Social Security or Railroad Payments (N); Family Received Supplemental Security Income (N); Family Received Welfare/Job Placement/Childcare Services (N); Family Received Income from a Job (N); Family Received Food Stamps (N); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
18-25	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (3 Levels) (M, WM); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Total Family Income > or < $20,000 (GT); Family Received Public Assistance (N); Family Received Social Security or Railroad Payments (N); Family Received Supplemental Security Income (N); Family Received Welfare/Job Placement/Childcare Services (N); Family Received Income from a Job (N); Family Received Food Stamps (N); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)
26-64	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Total Family Income > or < $20,000 (GT); Family Received Public Assistance (N); Family Received Social Security or Railroad Payments (N); Family Received Supplemental Security Income (N); Family Received Welfare/Job Placement/Childcare Services (N); Family Received Income from a Job (N); Family Received Food Stamps (N); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

Table D.59 Income Finer Categories in Predictive Mean Models (continued)

Age Group	Variables Included in Income Models (Finer Categorization)
65+	Age; Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Income State Rank (L, M); Number of People Older than 64 Years Old in Household (Continuous); Number of People Younger than 18 Years Old in Household; Marital Status (4 Levels) (M, W, DS); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Number of Adults in Household; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (W, B, H); Total Family Income > or < $20,000 (GT); Family Received Public Assistance (N); Family Received Social Security or Railroad Payments (N); Family Received Supplemental Security Income (N); Family Received Welfare/Job Placement/Childcare Services (N); Family Received Income from a Job (N); Family Received Food Stamps (N); Census Region (N, M, S); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

D.7 Health Insurance Variables

Exhibit D.5 Definitions of Levels for Categorical Variables

Age Squared * Race/Hispanic Recode (4 Levels)
W: Age Squared * Not Hispanic/Latino and White Only,[1] B: Age Squared * Not Hispanic/Latino and Black/African American Only, O: Age Squared * Not Hispanic/Latino and Other, H: Age Squared * Hispanic/Latino

Age * Race/Hispanic Recode (4 Levels)
W: Age * Not Hispanic/Latino and White Only,[1] B: Age * Not Hispanic/Latino and Black/African American Only, O: Age * Not Hispanic/Latino and Other, H: Age * Hispanic/Latino

Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]

Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]

Family Participation in Government Assistance Programs
N: No,[1] Y: Yes

Household Size (2 Levels)
L: At Least Four People in Household, S: Fewer than Four People in Household[1]

Family Income Recode
L: Income Less than $20,000, M1: Income $20,000-$49,999, M2: Income $50,000-$74,999, H: Income $75,000 or More[1]

Intermediate MEDICAID/CHIP Coverage
Y: Yes, N: No[1]

Intermediate CHAMPUS Coverage
Y: Yes, N: No[1]

Intermediate MEDICARE Coverage
Y: Yes, N: No[1]

Other Family Members In Household
N: No, Y: Yes[1]

Marital Status (2 Levels)
M: Married, NM: Never Been Married[1]

Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]

Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]

MSA (3 Levels)
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]

Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only,[1] B: Not Hispanic/Latino and Black/African American Only, O: Not Hispanic/Latino and Other, H: Hispanic/Latino

Family Received Social Security or Railroad Payments
Y: Yes, N: No[1]

Family Received Income from a Job
Y: Yes, N: No[1]

Lifetime Military Service
N: No,[1] Y: Yes

Exhibit D.5 Definitions of Levels for Categorical Variables (continued)

Gender
M: Male, F: Female[1]

Gender * Race/Hispanic Recode (4 Levels)
MW: Male Not Hispanic/Latino and White Only,[1] MB: Male Not Hispanic/Latino and Black/African American Only, MO: Male Not Hispanic/Latino and Other, MH: Male Hispanic/Latino, FW: Female Not Hispanic/Latino and White Only,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FO: Female Not Hispanic/Latino and Other,[1] FH: Female Hispanic/Latino[1]

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.60 Health Insurance: Response Propensity Models

Age Group	Set of Variables Used to Determine Nonresponse	Variables Included in Response Propensity Model
12-17	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	**Age**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age Squared**; Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
	Other Health Insurance	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
18-25	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Income Recode (L, M1, M2); Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
	Other Health Insurance	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Income Recode (L, M1, M2); Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

Table D.60 Health Insurance: Response Propensity Models (continued)

Age Group	Set of Variables Used to Determine Nonresponse	Variables Included in Response Propensity Model
26-64	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	Age; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **O**, **H)**; Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Income Recode (L, M1, M2); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); **Age * Gender**; **Age Squared * Gender**; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
	Other Health Insurance[1]	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Income Recode (L, M1, M2); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
65+	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Family Income Recode (L); Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M)
	Other Health Insurance[1]	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Income Recode (L, M1, M2); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: An asterisk "*" represents an interaction between two variables.

[1] The 26-64 and 65+ age groups were included in the same response propensity model for other health insurance.

Table D.61 Health Insurance: Predictive Mean Models, People 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Household Size; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Medicare	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Intermediate MEDICAID/CHIP Coverage (Y); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
CHAMPUS	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Family Income Recode (L); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Private Health Insurance	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate CHAMPUS Coverage (Y); Intermediate MEDICARE Coverage (Y); Household Size; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Other Health Insurance	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Household Size; Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: An asterisk "*" represents an interaction between two variables.

Table D.62 Health Insurance: Predictive Mean Models, People 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Other Family Members In Household (N); Household Size; Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Medicare[1]	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate MEDICAID/CHIP Coverage (Y); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
CHAMPUS	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Income Recode (L); **Intermediate MEDICAID/CHIP Coverage (Y)**; **Intermediate MEDICARE Coverage (Y)**; Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Lifetime Military Service (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Private Health Insurance	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate CHAMPUS Coverage (Y); Intermediate MEDICARE Coverage (Y); Other Family Members In Household (N); Household Size; Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Other Health Insurance	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Other Family Members In Household (N); Household Size; Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: An asterisk "*" represents an interaction between two variables.

[1] The 18-25 and 26-64 age groups were included in the same predictive mean model for Medicare.

Table D.63 Health Insurance: Predictive Mean Models, People 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Other Family Members In Household (N); Household Size; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Medicare[1]	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Intermediate MEDICAID/CHIP Coverage (Y); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
CHAMPUS	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Income Recode (L); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Lifetime Military Service (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Private Health Insurance	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate CHAMPUS Coverage (Y); Intermediate MEDICARE Coverage (Y); Other Family Members In Household (N); Household Size; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)
Other Health Insurance[2]	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Other Family Members In Household (N); Household Size; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: An asterisk "*" represents an interaction between two variables.

[1] The 18-25 and 26-64 age groups were included in the same predictive mean model for Medicare.

[2] The 26-64 and 65+ age groups were included in the same predictive mean model for other health insurance.

Table D.64 Health Insurance: Predictive Mean Models, People 65 Years or Older

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Family Participation in Government Assistance Programs (Y); Household Size (2 Levels) (L); Family Income Recode (L); Other Family Members In Household (N); Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M)
Medicare	Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Intermediate MEDICAID/CHIP Coverage (Y); Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Gender (M)
CHAMPUS	Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Family Income Recode (L); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Lifetime Military Service (Y); Gender (M)
Private Health Insurance	Age; Age Squared; Education Level (4 Levels) (L, HS, SC); Family Participation in Government Assistance Programs (Y); Household Size (2 Levels) (L); Family Income Recode (L); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate CHAMPUS Coverage (Y); Intermediate MEDICARE Coverage (Y); Other Family Members In Household (N); Marital Status (2 Levels) (M); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M)
Other Health Insurance[1]	Age; Age Squared * Race/Hispanic Recode (4 Levels) (B, O, H); Age * Race/Hispanic Recode (4 Levels) (B, O, H); Age Squared; Education Level (4 Levels) (L, HS, SC); Family Participation in Government Assistance Programs (Y); Family Income Recode (L, M1, M2); Other Family Members In Household (N); Household Size; Marital Status (4 Levels) (M, W, DS); Percentage Owner Occupied in Segment (H, M); MSA (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) (B, O, H); Family Received Social Security or Railroad Payments (Y); Family Received Income from a Job (Y); Gender (M); Age * Gender; Age Squared * Gender; Gender * Race/Hispanic Recode (4 Levels) (MB, MO, MH)

NOTE: An asterisk "*" represents an interaction between two variables.

[1] The 26-64 and 65+ age groups were included in the same predictive mean model for other health insurance.

D.8 Pair Relationship Variables

Exhibit D.6 Definitions of Levels for Categorical Variables

Age Category (Older)
Y: 12-14, T1: 15-17, T2: 18-20, T3: 21-25, A1: 26-34, A2: 35-49, A3: 50+, UK: Unknown[1]
Age Category - Non-Pair
Y: 12-14, T1: 15-17, T2: 18-20, T3: 21-25, A1: 26-34, A2: 35-49, A3: 50+, UK: Unknown[1]
Percentage Black/African American in Segment
H: ≥ 40%, M: [10,40)%, L: < 10%[1]
Education Level (4 Levels) - Non-Pair
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status - Non-Pair
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]
Marital Status (2 Levels) - Non-Pair
MWD: Married, Widowed or Divorced, NM: Never Been Married[1]
Marital Status (Older) (2 Levels)
MWD: Married, Widowed or Divorced, NM: Never Been Married[1]
Education Level (Older) (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status (Older)
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Marital Status (Older) (4 Levels)
M: Married, W: Widowed, D: Divorced, NM: Never Been Married[1]
Gender (Older)
F: Female, M: Male[1]
Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10,50)%, L: < 10%[1]
Population Density (3 Levels)
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]
Race/Hispanic Recode (Older) (4 Levels)
H: Hispanic/Latino, O: Not Hispanic/Latino and Other, B: Not Hispanic/Latino and Black/African American Only, W: Not Hispanic/Latino and White Only[1]
Race/Hispanic Recode (4 Levels) - Non-Pair
W: Not Hispanic/Latino and White Only,[1] B: Not Hispanic/Latino and Black/African American Only, O: Not Hispanic/Latino and Other, H: Hispanic/Latino
Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
Gender - Non-Pair
M: Male,[1] F: Female
Education Level (Younger) (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status (Younger)
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Marital Status (Younger) (4 Levels)
M: Married, W: Widowed, D: Divorced, NM: Never Been Married[1]
Gender (Younger)
F: Female, M: Male[1]

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.65 Model Summaries (Pair Relationships)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
0 (12-14, 12-14)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); **Race/Hispanic Recode (Older) (4 Levels) (**H, O, **B)**; **Census Region (**N, M, **S)**; Gender (Younger) (F)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Gender (Older) (F); **Percentage Owner Occupied in Segment (H**, **M)**; Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); **Census Region (N**, M, S**)**; Gender (Younger) (F)	Number in Household Aged 0-11; Number in Household Aged 12-17; **Number in Household Aged 18-25**; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); **Percentage Hispanic/Latino in Segment (H**, **M)**; Gender (Older) (F); **Percentage Owner Occupied in Segment (H**, **M)**; Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); **Census Region (N**, M, S**)**; Gender (Younger) (F)
1 (12-14, 15-17)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.65 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
2 (12,14, 18-25)	**Age Category (Older) (T3)**; Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); **Employment Status (Older) (FT**, **PT**, **UN)**; **Marital Status (Older) (4 Levels) (M**, **W**, **D)**; Gender (Older) (F); **Percentage Owner Occupied in Segment (H, M)**; **Population Density (3 Levels) (R**, **SM)**; Race/Hispanic Recode (Older) (4 Levels) (H, O, B); **Census Region (N**, **M**, **S)**; Gender (Younger) (F)	**Age Category (Older) (T3)**; Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)
3 (15-17, 15-17)	Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); **Percentage Hispanic/Latino in Segment (H, M)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); **Race/Hispanic Recode (Older) (4 Levels) (H**, O, B**)**; **Census Region (N**, **M**, **S)**; Gender (Younger) (F)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.65 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
4 (15-17, 18-25)	Age Category (Older) (T3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.65 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model: Including Household Size	Variables Included in Predictive Mean Model: Not Including Household Size
5 (18-20, 18-25)	Age Category (Older) (T3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Younger) (FT, PT, UN); **Marital Status (Younger) (4 Levels) (**M, **W**, D**)**; Gender (Younger) (F)	Age Category (Older) (T3); Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); **Education Level (Older) (4 Levels) (**L, **HS**, **SC)**; Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); **Population Density (3 Levels) (R**, **SM)**; Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); **Education Level (Younger) (4 Levels) (**L, **HS**, **SC)**; Employment Status (Younger) (FT, PT, UN); **Marital Status (Younger) (4 Levels) (M**, **W**, **D)**; **Gender (Younger) (F)**	Age Category (Older) (T3); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); **Education Level (Older) (4 Levels) (L**, **HS**, **SC)**; **Employment Status (Older) (FT**, **PT**, **UN)**; **Marital Status (Older) (4 Levels) (M**, **W**, **D)**; **Gender (Older) (F)**; Percentage Owner Occupied in Segment (H, M); **Population Density (3 Levels) (R**, **SM)**; Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); **Education Level (Younger) (4 Levels) (L**, **HS**, **SC)**; **Employment Status (Younger) (FT**, **PT**, **UN)**; **Marital Status (Younger) (4 Levels) (M**, **W**, **D)**; **Gender (Younger) (F)**

Table D.65 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
6 (21-25, 21-25)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Younger) (FT, PT, UN); **Marital Status (Younger) (4 Levels) (**M, **W**, D**)**; Gender (Younger) (F)	Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Younger) (FT, PT, UN); **Marital Status (Younger) (4 Levels) (**M, **W**, D**)**; Gender (Younger) (F)	**Number in Household Aged 0-11**; Number in Household Aged 12-17; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (**M, **W**, D**)**; Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Younger) (FT, PT, UN); **Marital Status (Younger) (4 Levels) (**M, **W**, D**)**; Gender (Younger) (F)

Table D.65 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
7 (12-14, 26+)	**Percentage Black/African American in Segment (H, M)**; **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); **Population Density (3 Levels) (R, SM)**; Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)
8 (15-17, 26+)	Age Category (Older) (A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.65 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
9 (18-20, 26+)	Age Category (Older) (A2, A3); Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Marital Status (Older) (4 Levels) (M, W, D)**; **Gender (Older) (F)**; **Percentage Owner Occupied in Segment (H, M)**; **Population Density (3 Levels) (R, SM)**; Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Education Level (Younger) (4 Levels) (L, HS, SC); **Employment Status (Younger) (**FT, PT, **UN)**; **Marital Status (Younger) (4 Levels) (**M, **W**, D); **Gender (Younger) (F)**	Age Category (Older) (A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); **Education Level (Younger) (4 Levels) (L, HS, SC)**; Employment Status (Younger) (FT, PT, UN); **Marital Status (Younger) (4 Levels) (**M, **W**, **D)**; Gender (Younger) (F)	Age Category (Older) (A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); **Education Level (Younger) (4 Levels) (L, HS, SC)**; Employment Status (Younger) (FT, PT, UN); **Marital Status (Younger) (4 Levels) (**M, **W**, **D)**; Gender (Younger) (F)

Table D.65 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model: Including Household Size	Variables Included in Predictive Mean Model: Not Including Household Size
10 (21+, 26+)	Age Category (Older) (A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Younger) (FT, PT, UN); Marital Status (Younger) (4 Levels) (M, W, D); Gender (Younger) (F)	Age Category (Older) (A2, A3); Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Younger) (FT, PT, UN); Marital Status (Younger) (4 Levels) (M, W, D); Gender (Younger) (F)	Age Category (Older) (A2, A3); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Marital Status (Older) (4 Levels) (M, W, D); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Younger) (FT, PT, UN); Marital Status (Younger) (4 Levels) (M, W, D); Gender (Younger) (F)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.6 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table D.66 Model Summaries (Multiplicities)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Parent-Child (12-20) Child Focus	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	**Age Category (Older) (T2**, **T3**, A1, A2**)**; Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	**Age Category (Older) (T2**, **T3**, A1, **A2)**; Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); **Marital Status (Older) (2 Levels) (MWD)**; **Education Level (Older) (4 Levels) (**L, HS, **SC)**; Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.66 Model Summaries (Multiplicities) (continued)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Parent-Child (12-20) Parent Focus	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); **Percentage Hispanic/Latino in Segment** (H, **M)**; Marital Status (Older) (2 Levels) (MWD); **Education Level (Older) (4 Levels) (**L, **HS**, **SC)**; Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); **Population Density (3 Levels) (**R, **SM)**; Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	**Age Category (Older) (T2**, T3, A1, **A2)**; Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); **Education Level (Older) (4 Levels) (**L, HS, **SC)**; Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	**Age Category (Older) (T2**, T3, A1, A2**)**; **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)
Sibling (12-14) Sibling (15-17) Older Sibling Focus	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.66 Model Summaries (Multiplicities) (continued)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Sibling (12-14) Sibling (15-17) Younger Sibling Focus	**Percentage Black/African American in Segment (H, M)**; **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); **Percentage Hispanic/Latino in Segment (**H, **M)**; Gender (Older) (F); **Percentage Owner Occupied in Segment (H, M)**; **Population Density (3 Levels) (R**, **SM)**; **Race/Hispanic Recode (Older) (4 Levels) (**H, O, **B)**; **Census Region (**N, **M, S)**; Gender (Younger) (F)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)
Sibling (12-17) Sibling (18-25) Older Sibling Focus	Age Category (Older) (T2); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.66 Model Summaries (Multiplicities) (continued)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Sibling (12-17) Sibling (18-25) Younger Sibling Focus	**Age Category (Older) (T2)**; Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); **Education Level (Older) (4 Levels) (L, HS, SC)**; Employment Status (Older) (FT, PT, UN); **Gender (Older) (F)**; Percentage Owner Occupied in Segment (H, M); **Population Density (3 Levels)** (R, **SM)**; Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T2); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T2); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.6 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table D.67 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-Child (12-20) Child Focus, Both Pair Members Younger than 18	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)
Parent-Child (12-20) Child Focus, at Least One Pair Member Older than 18	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.67 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Parent-Child (12-20) Parent Focus, Both Pair Members Younger than 18	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)
Parent-Child (12-20) Parent Focus, at Least One Pair Member Older than 18	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Number in Household Aged 0-11; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.67 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Both Pair Members Younger than 18	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); **Percentage Owner Occupied in Segment (**H, **M)**; Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)
Sibling (12-14) Sibling (15-17), Older Sibling Focus, at Least One Pair Member Older than 18	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.67 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Sibling (12-17) Sibling (18-25), Older Sibling Focus, Both Pair Members Younger than 18	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); **Race/Hispanic Recode (Older) (4 Levels) (**H, O, **B)**; Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)
Sibling (12-17) Sibling (18-25), Older Sibling Focus, at Least One Pair Member Older than 18	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.67 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse, Both Pair Members Younger than 18	Age Category (Older) (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	**Age Category (Older) (Y)**; **Percentage Black/African American in Segment (H**, **M)**; **Household Size, Reconciled Across Pairs (Continuous)**; **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); **Percentage Owner Occupied in Segment (H**, **M)**; Population Density (3 Levels) (R, SM); **Race/Hispanic Recode (Older) (4 Levels) (**H, O, **B)**; **Census Region (**N, **M**, **S)**; Gender (Younger) (F)	**Age Category (Older) (Y)**; Number in Household Aged 0-11; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; **Percentage Black/African American in Segment (H**, **M)**; Percentage Hispanic/Latino in Segment (H, M); Gender (Older) (F); **Percentage Owner Occupied in Segment (H**, **M)**; Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); **Census Region (**N, **M**, S); Gender (Younger) (F)
Spouse-Spouse, at Least One Pair Member Older than 18	Age Category (Older) (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	**Age Category (Older) (**T3, A1, **A2**, **A3)**; Percentage Black/African American in Segment (H, M); **Household Size, Reconciled Across Pairs (Continuous)**; **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); **Marital Status (Older) (2 Levels) (MWD)**; Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); **Gender (Older) (F)**; Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); **Race/Hispanic Recode (Older) (4 Levels) (**H, O, **B)**; Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T3, A1, A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); **Marital Status (Older) (2 Levels) (MWD)**; Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.67 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse with Children, Both Pair Members Younger than 18	Age Category (Older) (T1, T2, T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T1, T2, T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T1, T2, T3, A1, A2, A3); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

Table D.67 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse with Children, at Least One Pair Member Older than 18	Age Category (Older) (T1, T2, T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T1, T2, T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)	Age Category (Older) (T1, T2, T3, A1, A2, A3); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Gender (Older) (F); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (Older) (4 Levels) (H, O, B); Census Region (N, M, S); Gender (Younger) (F)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.6 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table D.68 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-Child (12-20) Child Focus, Younger than 18	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)
Parent-Child (12-20) Child Focus, Older than 18	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T3, A1, A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)

Table D.68 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-Child (12-20) Parent Focus, Younger than 18	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)
Parent-Child (12-20) Parent Focus, Older than 18	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	**Age Category - Non-Pair (**T3, **A1**, **A2**, **A3)**; Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); **Marital Status (2 Levels) - Non-Pair (MWD)**; Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	**Age Category - Non-Pair (T3**, **A1**, **A2**, **A3)**; Number in Household Aged 0-11; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)

Table D.68 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Younger than 18	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Older than 18	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T3, A1, A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)

Table D.68 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-17) Sibling (18-25), Older Sibling Focus, Younger than 18	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)
Sibling (12-17) Sibling (18-25), Older Sibling Focus, Older than 18	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T3, A1, A2, A3); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+; Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)

Table D.68 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse, Younger than 18	Age Category - Non-Pair (Y); Percentage Black/African American in Segment (H, M); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); **Percentage Black/African American in Segment (H**, **M)**; **Household Size, Reconciled Across Pairs (Continuous)**; **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) - Non-Pair (B**, O, H**)**; Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (Y); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; **Percentage Black/African American in Segment (H**, **M)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); **Race/Hispanic Recode (4 Levels) - Non-Pair (B**, O, H**)**; Census Region (N, M, S); Gender - Non-Pair (F)
Spouse-Spouse, Older than 18	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); **Household Size, Reconciled Across Pairs (Continuous)**; **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T3, A1, A2, A3); **Number in Household Aged 0-11**; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); **Marital Status (2 Levels) - Non-Pair (MWD)**; Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)

Table D.68 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model: Including Household Size	Variables Included in Predictive Mean Model: Not Including Household Size
Spouse-Spouse with Children	Age Category - Non-Pair (T1, T2, T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Household Size, Reconciled Across Pairs (Continuous); **Household Size, Reconciled Across Pairs (5 Levels)**; Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T1, T2, T3, A1, A2, A3); Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); **Household Size, Reconciled Across Pairs (Continuous)**; Household Size, Reconciled Across Pairs (5 Levels); Percentage Hispanic/Latino in Segment (H, M); Marital Status (2 Levels) - Non-Pair (MWD); Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)	Age Category - Non-Pair (T1, T2, T3, A1, A2, A3); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**; Percentage Black/African American in Segment (H, M); Education Level (4 Levels) - Non-Pair (L, HS, SC); Employment Status - Non-Pair (FT, PT, UN); Percentage Hispanic/Latino in Segment (H, M); **Marital Status (2 Levels) - Non-Pair (MWD)**; Percentage Owner Occupied in Segment (H, M); Population Density (3 Levels) (R, SM); Race/Hispanic Recode (4 Levels) - Non-Pair (B, O, H); Census Region (N, M, S); Gender - Non-Pair (F)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.6 for definitions of levels for categorical variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Appendix E: Hot-Deck Procedure Summaries

This page intentionally left blank

Appendix E: Hot-Deck Procedure Summaries

E.1 Introduction

For the majority of variables that had missing values imputed in the 2014 National Survey on Drug Use and Health (NSDUH), the imputation method used was predictive mean neighborhood (PMN). This appendix summarizes the predictive mean vectors and the constraints applied during the PMN hot-deck step.[1] It is organized by groups of variables requiring imputation: demographics, household composition (roster), lifetime use of drugs, recency and frequency of drug use, age at first drug use, income, health insurance, and pair. There are three types of tables associated with each variable or set of variables imputed using the PMN method: (1) Logical Constraints, (2) Likeness Constraints, and (3) Constraints and Portion of the Predictive Mean Vector.

For variables that do not apply any logical constraints in the PMN process, the "Logical Constraints" table is not applicable. The "Constraints and Portion of the Predictive Mean Vector" table specifies the following for each missingness pattern:

1. the number of item nonrespondents exhibiting the pattern ("Total Number of Cases");
2. the set of logical constraints applied to the potential donors ("Logical Constraints");
3. the elements of the predictive mean vector ("Predictive Mean Vector") used to calculate the Mahalanobis distance from recipient to potential donor, as well as to restrict the donor set via the delta constraint; and
4. the set of likeness constraints utilized in each try and the number of item nonrespondents who found donors on each try by different age groups ("Likeness Constraints: Number of Cases").[2]

In the tables that follow, the phrase "Donor's predicted means each must be within x percent of recipient's predicted means" appears in each of the multivariate imputation tables, and the phrase "Donor's predicted mean must be within x percent of recipient's predicted mean" appears in each of the univariate imputation tables. In either case, it represents one of the likeness constraints, known as the "delta constraint," and also defines the neighborhood. When this constraint is loosened, the neighborhood of potential donors is abandoned, and the candidate with the predicted mean closest to the recipient's (subject to the constraints that are still on the pool of donors) is chosen as the donor.

Although statistical imputation of the drug use or income variables could not have proceeded separately within each state because of insufficient pools of donors, the PMN procedure does incorporate information about the state of residence of each respondent. For the drug use variables, in the hot-deck step of PMN, respondents were separated into three state

[1] See Chapter 3 for details on PMN imputation.
[2] If a cell contains the text "No Cases" within "Likeness Constraints: Number of Cases," this means that there were no cases requiring imputation in that category.

usage-level categories for each drug, depending on the response variable of interest. Respondents from states with high usage of a given drug were placed in one category, respondents from medium-usage states were placed in another category, and respondents from low-usage states were placed in a third category. For the income variables, the states were separated into three income groups, depending upon the proportion of families within those states with incomes greater than or equal to $20,000. As with the drug use variables, respondents from high-income states (by this measure) were placed in one category, respondents from medium-income states were placed in another category, and respondents from low-income states were placed in a third category. In the tables that follow, this variable is identified as the "state rank" for the drug use and income variables. It was used as a likeness constraint, where the set of eligible donors for each recipient was restricted so that donors and recipients were both from states with the same state rank.

E.2 Demographics

Tables E.1 through E.11 present information on the imputation procedures for the core demographic variables: marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level. Tables E.12 through E.19 present information for the noncore demographic variables: employment status, indicator of birth in the United States, and immigrant age of entry into the United States. In several instances, variable names are used without description for the purposes of brevity (see Chapter 4 for details). The segment-level variable, SEGID (Segment ID), was used only in the likeness constraints for demographic imputation.

As described in the sample design report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2015d) within each state, state sampling (SS) regions were formed, which were further partitioned into clusters of adjacent blocks called "segments." The segment ID number was a two-letter state abbreviation followed by a two-digit SS region and a two-digit segment identifier, which uniquely identifies each segment. Although the segment identifier was not used as a covariate because of the large number of levels, it was used as a constraint in the hot-deck step of the PMN procedure for race, Hispanicity, education, and employment status, as noted in Chapters 4 and 5.

E.2.1 Marital Status Variables

Table E.1 Likeness Constraints for Marital Status

Constraint #	Likeness Constraint
LikC1	Donor's age must be within 3 years of recipient's age
LikC2	Donor's predicted means must be within 5 percent of recipient's predicted means

Table E.2 Constraints and Portion of the Predictive Mean Vector for Marital Status

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	13	None	1. M1 2. M2 3. M3	No Cases	1,2: 3 1: 1	1,2: 9

[1] The predictive mean vector components are defined by the following:
1. M1 = P(Married)
2. M2 = P(Widowed)
3. M3 = P(Divorced or separated)

E.2.2 Race Variables

Table E.3 Logical Constraints for Race

Constraint #	Logical Constraint
LogC1	Donor must be Asian, in part or in full
LogC2	Donor must be more than one race
LogC3	Donor must not be white only
LogC4	Donor must be white only or white and American Indian/Alaska Native only
LogC5	Donor must not be American Indian/Alaska Native, in part or in full

Table E.4 Likeness Constraints for Race

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC3	If recipient was Hispanic/Latino nonspecific, donor must be of Hispanic/Latino origin
LikC4	If recipient selected one or more Hispanic/Latino categories, including Mexican, Puerto Rican, Central or South American, Cuban, Dominican, and Spaniard, then donor's Hispanic/Latino group value must be equal to one of the Hispanic/Latino groups mentioned by recipient
LikC5	Donor must be Mexican (Hispanic/Latino or non-Hispanic/Latino)
LikC6	Donor must be Cuban (Hispanic/Latino or non-Hispanic/Latino)
LikC7	Donor must be Central or South American (Hispanic/Latino or non-Hispanic/Latino)
LikC8	Donor must be Dominican (Hispanic/Latino or non-Hispanic/Latino)
LikC9	Donor must be Spanish (Hispanic/Latino or non-Hispanic/Latino)
LikC10	Donor must be Puerto Rican (Hispanic/Latino or non-Hispanic/Latino)

Table E.5 Constraints and Portion of the Predictive Mean Vector for Race

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	2,363	None	1. R1 2. R2 3. R3 4. R4	1-4: 147 2-4: 373 3,4: 158	1-4: 135 2-4: 291 3,4: 210	1-4: 85 2-4: 557 3,4: 407
2	Known to be Asian	5	1	1. R1 2. R2 3. R3 4. R4	No Cases	1-4: 0 2-4: 2	1-4: 0 2-4: 1 3,4: 2
3	Known to be multiple race, but no other information	19	2	1. R1 2. R2 3. R3 4. R4	1-4: 0 2-4: 0 3,4: 3	1-4: 0 2-4: 1 3,4: 8	1-4: 1 2-4: 0 3,4: 6
4	Known to be nonwhite, but no other information	9	3	1. R2/(1-R1) 2. R3/(1-R1) 3. R4/(1-R1)	1-4: 0 2-4: 1 3,4: 1	1-4: 0 2-4: 0 3,4: 1	1-4: 0 2-4: 4 3,4: 2
5	Known to be white or both white and American Indian/Alaska Native	2	4	1. R1/(1-R2-R3-R4)	1-4: 0 2-4: 1	No Cases	1-4: 0 2-4: 1
6	Known not to be American Indian/ Alaska Native, in part or in full	1	5	1. R1/(1-R3) 2. R2/(1-R3) 3. R4/(1-R3)	1-4: 0 2-4: 1	No Cases	No Cases
7	Known to be non-Hispanic Mexican	3	None	1. R1 2. R2 3. R3 4. R4	1-5: 0 2-5: 2	No Cases	1-5: 0 2-5: 1
8	Known to be non-Hispanic Cuban	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases

Table E.5 Constraints and Portion of the Predictive Mean Vector for Race (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
9	Known to be non-Hispanic Central or South American	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases
10	Known to be non-Hispanic Dominican	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases
11	Known to be non-Hispanic Spanish	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases
12	Known to be non-Hispanic Puerto Rican	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases

[1] The predictive mean vector components are defined by the following:
1. R1 = P(White Only)
2. R2 = P(Black Only)
3. R3 = P(American Indian/Alaska Native Only)
4. R4 = P(Asian Only)

E.2.3 Hispanic/Latino Origin Variables

Table E.6 Likeness Constraints for Hispanic/Latino Origin

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.7 Constraints and Portion of the Predictive Mean Vector for Hispanic/Latino Origin

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	165	None	1. H1	1,2: 76 2: 64	1,2: 2 2: 7	1,2: 5 2: 11

[1] The predictive mean vector components are defined by the following:
1. H1 = P(Hispanic/Latino origin)

E.2.4 Hispanic/Latino Group Variables

Table E.8 Likeness Constraints for Hispanic/Latino Group

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	If recipient had 8 ≤ EDHOGRP ≤ 21, donor's values for IRRACWH, IRRACEBK, IRRACENA, IRRACEAS, IRRACENH, and IRRACEPI must indicate a subset of the racial categories mentioned by recipient

Table E.9 Constraints and Portion of the Predictive Mean Vector for Hispanic/Latino Group

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases[2]
1	Completely missing	59	None	1. H1 2. H2 3. H3	1-3: 8 2,3: 41 3: 10

[1] The predictive mean vector components are defined by the following:
1. H1 = P(Mexican)
2. H2 = P(Puerto Rican)
3. H3 = P(Central or South American)

[2] The hot-deck program for the Hispanic/Latino group is not separated into age groups.

E.2.5 Education Variables

Table E.10 Likeness Constraints for Education Level

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Age of donor = age of recipient

Table E.11 Constraints and Portion of the Predictive Mean Vector for Education Level

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector, by Age Group[1]		Likeness Constraints: Number of Cases, by Age Group		
				12-17	18+	12-17	18-25	26+
1	Missing	9	None	1. E1 2. E2 3. E3 4. E4	1. E1 2. E2 3. E3	1-3: 0 2,3: 3	1-3: 0 2,3: 0 3: 1	1,2: 1 2: 4

[1] The predictive mean vector components are defined by the following:
12-17 Age Group
1. E1 = P(Less than elementary school)
2. E2 = P(Elementary school)
3. E3 = P(Middle school)
4. E4 = P(Some high school)

18+ Age Group
1. E1 = P(Less than high school)
2. E2 = P(High school)
3. E3 = P(Some college)

E.2.6 Employment Variables

Table E.12 Logical Constraints for Employment Status

Constraint #	Logical Constraint
LogC1	Donor must be employed

Table E.13 Likeness Constraints for Employment Status

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Donor's age must be within 5 years of recipient's age

Table E.14 Constraints and Portion of the Predictive Mean Vector for Employment Status

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group 12-17	18-25	26+
1	Completely missing	34	None	1. E1 2. E2 3. E3	1-3: 1 2,3: 1	1-3: 0 2,3: 5	1-3: 2 2,3: 21 3: 4
2	Known to be employed; part-time vs. full-time status unknown	23	1	1. E1/ (E1+E2)	1-3: 0 2,3: 5	1-3: 0 2,3: 6	1-3: 4 2,3: 8

[1] The predictive mean vector components are defined by the following:
1. E1 = P(Employed full time)
2. E2 = P(Employed part time)
3. E3 = P(Unemployed)

E.2.7 Immigrant Variables

Table E.15 Likeness Constraints for Indicator of Birth in the United States

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.16 Constraints and Portion of the Predictive Mean Vector for Indicator of Birth in the United States

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group 12-17	18-25	26+
1	Completely missing	37	None	1. B1	1,2: 4 2: 5	1,2: 1 2: 4	1,2: 8 2: 15

[1] The predictive mean vector components are defined by the following:
1. B1 = P(Born in the United States)

Table E.17 Logical Constraints for Age of Entry in the United States

Constraint #	Likeness Constraint
LogC1	Donor's age of entry must be less than recipient's current age
LogC2	Difference between recipient's current age and donor's age of entry must be less than or equal to 1 if recipient lived in the United States less than a year, or difference must be greater than 1 if recipient lived in the United States more than a year

Table E.18 Likeness Constraints for Age of Entry in the United States

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.19 Constraints and Portion of the Predictive Mean Vector for Age of Entry in the United States

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases[2]
1	Completely missing	23	1,2	1. PrAgeEntry	1,2: 4 2: 19

[1] The predictive mean vector components are defined by the following:
1. PrAgeEntry = Predicted age of entry

[2] The hot-deck program for immigrant age of entry is not separated into age groups.

E.3 Drug Variables

Tables E.20 through E.75 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures to variables for lifetime drug use, recency and frequency of use, and age at first use.

E.3.1 Lifetime Drug Use

There were a large number of missingness patterns for lifetime drug use. The usable case rule (Section 2.2.3) required a certain minimum response pattern, but other than that any combination of the lifetime drug variables may have been missing for a given respondent record.

The probabilities associated with the 14 lifetime drug variables listed in Table E.20 formed the full predictive mean vector. Only the probabilities associated with the lifetime drug variables for which values were missing were used in the predictive mean vector for each item nonrespondent. The predicted mean for cigarette lifetime use was never calculated because no respondents were missing the lifetime variable for cigarette use.[3]

Note that if only a child lifetime drug variable was missing, the predicted mean associated with the parent drug was used. No predicted means were calculated for child drugs. For example, if the only missing lifetime drug use variable was the one for methamphetamine, the predictive mean "vector" would be a scalar: the predicted mean associated with lifetime use of any stimulant.[4]

[3] The lifetime variable for cigarette use was never missing because the usable case rule required a nonmissing response to the cigarette gate question.

[4] Crack is an exception to this rule. Crack is considered a child drug of cocaine, but it is unique among child drugs in that it is covered by a separate module in the questionnaire. Those denying cocaine lifetime use are automatically assigned a negative response for crack lifetime use, and those with a missing response for cocaine lifetime use are automatically assigned a missing response for crack lifetime use.

Table E.20 Elements of Full Predictive Mean Vector for Lifetime Drug Use

Lifetime Drug Use	Predicted Mean
Heroin Lifetime	P(Lifetime User)
Crack Lifetime	P(Lifetime User | Lifetime User of Cocaine), if cocaine lifetime use is known P(Lifetime User of Cocaine)* P(Lifetime User | Lifetime User of Cocaine), if cocaine lifetime use is unknown
Cocaine Lifetime	P(Lifetime User)
Sedatives Lifetime	P(Lifetime User)
Stimulants/Methamphetamine Lifetime	P(Lifetime User)
Tranquilizers Lifetime	P(Lifetime User)
Pain Relievers/OxyContin Lifetime	P(Lifetime User)
Hallucinogens/LSD/PCP/Ecstasy Lifetime	P(Lifetime User)
Marijuana Lifetime	P(Lifetime User)
Inhalants Lifetime	P(Lifetime User)
Alcohol Lifetime	P(Lifetime User)
Pipes Lifetime	P(Lifetime User)
Snuff/Chewing Tobacco Lifetime	P(Lifetime User)
Cigars Lifetime	P(Lifetime User)

Table E.21 Logical Constraints for Lifetime

Constraint #	Logical Constraint
LogC1	Donor must be a lifetime user of pain relievers if recipient is known to have used pain relievers but missing both indicators for OxyContin and "other" pain reliever

Table E.22 Likeness Constraints for Lifetime

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Lifetime use of donor = lifetime use of recipient, for each nonmissing lifetime indicator
LikC4	If recipient was missing the lifetime indicator(s) for any member of a family of drugs,[1] donor's lifetime indicator(s) agreed with recipient's nonmissing lifetime indicator(s) within that family

[1] The smokeless tobacco family includes chewing tobacco and snuff. The hallucinogens family includes LSD, PCP, Ecstasy, and other hallucinogens. The pain relievers family includes OxyContin and other pain relievers. The stimulants family includes methamphetamine and other stimulants. The cocaine family includes crack.

Table E.23 Constraints and Portion of the Predictive Mean Vector for Lifetime

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing one or more of the lifetime drug use variables	833	1	1. LF for gate question associated with each missing variable	1-3: 370 1,2,4: 66 1,4: 37	1-3: 64 1,2,4: 45 1,4: 22	1-3: 118 1,2,4: 64 1,4: 47

[1] The predictive mean vector components are defined by the following:
1. LF = P(Lifetime user of x), where x is the drug associated with the missing gate question

Table E.24 Likeness Constraints for Cigarette Ever Daily Used

Constraint #	Likeness Constraint
LikC1	Age of donor = age of recipient
LikC2	State rank of donor = state rank of recipient
LikC3	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC4	If the recipient was a past year cigarette user, the donor must also be a past year cigarette user
LikC5	If the recipient was a past 3-year but not past year or lifetime but not past 3-years cigarette user, the donor's cigarette recency must match the recipient's cigarette recency
LikC6	If the recipient was a past 3-year but not past year cigarette user, the donor cannot have used cigarettes in the past year
LikC7	If the recipient was a lifetime but not past 3-year cigarette user, the donor cannot have used cigarettes in the past 3 years
LikC8	Age of donor must be greater than or equal to age of recipient

Table E.25 Constraints and Portion of the Predictive Mean Vector for Cigarette Ever Daily Used

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing CIGDLYMO	25	None	1. CIG	1-5: 5 1,3-5: 1	1-5: 7	1-5: 7 1,3-5: 1 1,4,5: 3 1,6,7: 0 6-8: 1

[1] The predictive mean vector components are defined by the following:
1. CIG = P(Daily use at some point in lifetime | respondent was a lifetime user but not a current daily user)

E.3.2 Drug Recency and Frequency

Many tables in this section abbreviate certain words. "Recency" is an abbreviation for "recency of use," "frequency" or "Freq." is an abbreviation for "frequency of use," and "30-day binge drink" or "DR5DAY" is an abbreviation for the "number of days in the past 30 days when the respondent consumed five or more alcoholic drinks."

An empty cell within "Missingness Pattern" indicates that all information was available. A entry of "Missing" in a cell indicates that all information was missing. Other entries for missingness patterns provide the available information, indicating that this information was partially missing. For example, an entry of "Past year" indicates that the individual used the drug of interest in the past year, but the specific recency of past month versus past year not past month must be imputed. However, if the entry is shown in parentheses, then the specific recency is known and no imputation is required. For example, an entry of "(Past year)" indicates that the individual used the drug of interest in the past year and the specific recency of past month versus past year not past month is already known and does not require imputation.

Table E.26 Logical Constraints for Cigarette Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must have used cigarettes within the past 3 years (a cigarette recency category of 1, 2, or 3)
LogC2	Donor cannot be a past month cigarette user (cigarette recency ≠ 1)
LogC3	Donor cannot be a past year cigarette user (cigarette recency ≠ 1 or 2)
LogC4	Donor must be a past year cigarette user (cigarette recency = 1 or 2)
LogC5	Donor must be a past month cigarette user (cigarette recency = 1)
LogC6	If recipient was never a daily user of cigarettes (CIGDLYMO = 2), donor's 30-day cigarette frequency cannot equal 30
LogC7	If recipient's age at first use equals his or her current age, donor's 30-day cigarette frequency (1) cannot be greater than the number of days between recipient's interview date and his or her date of first cigarette use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and his or her birthday (inclusive)
LogC8	Donor must be a past year but not past month cigarette user or a past 3 years but not past year cigarette user (cigarette recency = 2 or 3)

Table E.27 Likeness Constraints for Cigarette Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.28 Constraints and Portion of the Predictive Mean Vector for Cigarette Recency and Frequency

#	Missingness Pattern[1]	Total Number of Cases	Logical Constraints	Predictive Mean Vector[2]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Past year, 30-day frequency missing	12	4,6	1. R1/(R1+R2) 2. (R1*D)/(R1+R2) 3. R1*(1-D)*PM/(R1+R2)	1,2: 3 1: 7	1,2: 2	No Cases
2	Recency and 30-day frequency missing	10	6	1. R1 2. R2 3. R3 4. R1*D 5. R1*(1-D)*PM	1,2: 0 1: 1	1,2: 0 1: 3	1,2: 0 1: 6
3	Past month, 30-day frequency missing	10	5-7	1. D 2. PM	1,2: 2 1: 2	1,2: 3	1,2: 3
4	Not past year	135	3,6	1. R3/(R3+R4)	1,2: 58	1,2: 71	1,2: 6
5	Not past month	58	2,6	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1,2: 34 1: 3	1,2: 14	1,2: 6 1: 1
6	Past year but not past month, or past 3 years but not past year	87	6,8	1. R2/(R2+R3)	1,2: 60	1,2: 23 1: 1	1,2: 3
7	Past 3 years, 30-day frequency missing	1	1,6	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. (R1*D)/(R1+R2+R3) 4. R1*(1-D)*PM/(R1+R2+R3)	1,2: 0 1: 1	No Cases	No Cases

[1] The response to CIGDLYMO, the edited response to the "ever daily used" question, technically could be used to subdivide each of the first three missingness patterns into two: one for respondents with CIGDLYMO = 2, and the other for respondents with CIGDLYMO ≠ 2. This was not done, because the benefit derived from this change would likely be insignificant. Respondents with CIGDLYMO = 2 technically have zero probability of being a daily user, so the predictive mean vectors could be simplified by setting D = 0. For example, the predictive mean vector for respondents in Missingness Pattern 2 with CIGDLYMO = 2 might look like this: (1) R1, (2) R2, (3) R3, (4) R1*PM.

[2] The predictive mean vector components are defined by the following:
1. R1 = P(Past month cigarette use | lifetime cigarette use)
2. R2 = P(Past year but not past month cigarette use | lifetime cigarette use)
3. R3 = P(Past 3 years but not past year cigarette use | lifetime cigarette use)
4. R4 = P(Lifetime but not past 3 years cigarette use | lifetime cigarette use)
5. D = P(Daily cigarette use | past month cigarette use)
6. PM = P(Cigarette use on a given day in the past month | past month use)

Table E.29 Logical Constraints for Cigar Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must have used cigars within the past 3 years (a cigar recency category of 1, 2, or 3)
LogC2	Donor cannot be a past month cigar user (cigar recency ≠ 1)
LogC3	Donor must be a past year cigar user (cigar recency = 1 or 2)
LogC4	Donor must be a past month cigar user (cigar recency = 1)
LogC5	Donor cannot be a past month or past year cigar user (cigar recency ≠ 1 or 2)
LogC6	If recipient's age at first use equals his or her current age, donor's 30-day cigar frequency (1) cannot be greater than the number of days between recipient's interview date and his or her date of first cigar use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and his or her birthday (inclusive)
LogC7	Donor must be a past year but not past month cigar user or a past 3 years but not past year cigar user (cigar recency = 2 or 3)

Table E.30 Likeness Constraints for Cigar Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.31 Constraints and Portion of the Predictive Mean Vector for Cigar Recency and Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Past year, 30-day frequency missing	9	3	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	1,2: 4 1: 3	1,2: 2	No Cases
2	Recency and 30-day frequency missing	12	None	1. R1 2. R2 3. R3 4. R1*PM	1,2: 0 1: 5	1,2: 0 1: 3	1,2: 0 1: 4
3	Past month, 30-day frequency missing	3	4,6	1. PM	1,2: 1 1: 1	No Cases	1,2: 1
4	Not past year	111	5	1. R3/(R3+R4)	1,2: 37 1: 3	1,2: 51	1,2: 20
5	Not past month	48	2	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1,2: 18 1: 7	1,2: 16	1,2: 6 1: 1
6	Past year but not past month, or past 3 years but not past year	87	7	1. R2/(R2+R3)	1,2: 40 1: 1	1,2: 39 1: 2	1,2: 5
7	Past 3 years, 30-day frequency missing	0	1	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. (R1*PM)/(R1+R2+R3)	No Cases	No Cases	No Cases

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month cigar use | lifetime cigar use)
2. R2 = P(Past year but not past month cigar use | lifetime cigar use)
3. R3 = P(Past 3 years but not past year cigar use | lifetime cigar use)
4. R4 = P(Lifetime but not past 3 years cigar use | lifetime cigar use)
5. PM = P(Cigar use on a given day in the past month | past month cigar use)

Table E.32 Logical Constraints for Smokeless Tobacco Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must have used chewing tobacco within the past 3 years (a chew recency category of 1, 2, or 3)
LogC2	Donor must have used snuff within the past 3 years (a chew recency category of 1, 2, or 3)
LogC3	Donor must be a lifetime user of chewing tobacco
LogC4	Donor must be a lifetime user of snuff
LogC5	If recipient's age at first chewing tobacco use equals his or her current age, donor's 30-day chewing tobacco frequency (1) cannot be greater than the number of days between recipient's interview date and his or her date of first chewing tobacco use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and his or her birthday (inclusive)
LogC6	If recipient's age at first snuff use equals his or her current age, donor's 30-day snuff frequency (1) cannot be greater than the number of days between recipient's interview date and his or her date of first snuff use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and his or her birthday (inclusive)
LogC7	Donor must be a past month chewing tobacco user (chew recency = 1)
LogC8	Donor must be a past month snuff user (snuff recency = 1)
LogC9	Donor's snuff recency must equal recipient's snuff recency
LogC10	Donor's chewing tobacco recency must equal recipient's chewing tobacco recency
LogC11	Donor must have used chewing tobacco within the past year (snuff recency = 1 or 2)
LogC12	Donor must have used snuff within the past year (chew recency = 1 or 2)
LogC13	Donor must be a past 3 years but not past year or lifetime but not past 3 years chewing tobacco user (chew recency = 3 or 4)
LogC14	Donor must be a past 3 years but not past year or lifetime but not past 3 years snuff user (snuff recency = 3 or 4)
LogC15	Donor must be a past year but not past month, past 3 years but not past year, or lifetime but not past 3 years chewing tobacco user (chew recency = 2, 3, or 4)
LogC16	Donor must be a past year but not past month, past 3 years but not past year, or lifetime but not past 3 years snuff user (snuff recency = 2, 3, or 4)
LogC17	Donor must be a past year but not past month or a past 3 years but not past year chewing tobacco user (chew recency = 2 or 3)
LogC18	Donor must be a past year but not past month or a past 3 years but not past year snuff user (snuff recency = 2 or 3)

Table E.33 Likeness Constraints for Smokeless Tobacco Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Donor's recencies for chewing tobacco and snuff agree with recipient's recencies for chewing tobacco and snuff (when nonmissing)
LikC4	Donor's lifetime use statuses for chewing tobacco and snuff agree with recipient's lifetime use statuses for chewing tobacco and snuff (when nonmissing)

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
1	(Past month)	(Past month)	Missing	Missing	0	5-8	1. DC 2. PMC 3. DS 4. PMS	No Cases	No Cases	No Cases
2	(Past month)		Missing		2	5,7	1. DC 2. PMC	No Cases	No Cases	1-3: 1 1,2,4: 0 1,4: 1
3		(Past month)		Missing	3	6,8	1. DS 2. PMS	No Cases	1-3: 0 1,2,4: 0 1,4: 2	1-3: 0 1,2,4: 0 1,4: 1
4		Missing		Missing	4	4,6	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS	1-3: 0 1,2,4: 0 1,4: 1	1-3: 0 1,2,4: 0 1,4: 1	1-3: 0 1,2,4: 0 1,4: 2
5	(Past month)	Missing	Missing	Missing	0	4-7	1. R1 2. R2 3. R3 4. DC 5. PMC 6. RS1*DS 7. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases
6	Missing		Missing		7	3,5	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC	1-3: 0 1,2,4: 0 1,4: 4	1-3: 0 1,2,4: 0 1,4: 1	1-3: 0 1,2,4: 0 1,4: 2

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern: Chew Recency	Missingness Pattern: Snuff Recency	Missingness Pattern: Chew 30-Day Freq.	Missingness Pattern: Snuff 30-Day Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
7	Missing	(Past month)	Missing	Missing	1	3,5,6,8	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. DS 7. PMS	No Cases	No Cases	1-3: 0 1,2,4: 0 1,4: 1
8		Past year		Missing	8	6,12	1. R1/(R1+R2) 2. RS1*DS/ (RS1+RS2) 3. RS1*(1-DS)*PMS/ (RS1+RS2)	1-3: 0 1,2,4: 0 1,4: 7	No Cases	1-3: 0 1,2,4: 0 1,4: 1
9	Past year		Missing		4	5,11	1. R1/(R1+R2) 2. RC1*DC/ (RC1+RC2) 3. RC1*(1-DC)*PMC/ (RC1+RC2)	1-3: 0 1,2,4: 0 1,4: 2	1-3: 0 1,2,4: 0 1,4: 2	No Cases
10	Missing	Missing	Missing	Missing	3	3-6	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. RS1*DS 7. RS1*(1-DS)*PMS	1-3: 0 1,2,4: 0 1,4: 1	1-3: 0 1,2,4: 0 1,4: 2	No Cases
11	Not past year				38	13	1. R3/(R3+R4)	1-3: 7 1,2,4: 1 1,4: 9	1-3: 16	1-3: 5
12		Not past year			54	14	1. R3/(R3+R4)	1-3: 10 1,2,4: 1 1,4: 5	1-3: 29 1,2,4: 0 1,4: 2	1-3: 6 1,2,4: 0 1,4: 1

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
13	Not past year	Not past year			8	13,14	1. R3/(R3+R4)	1-3: 1 1,2,4: 0 1,4: 1	1-3: 3 1,2,4: 0 1,4: 2	1-3: 1
14	Not past month				17	15	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 1 1,2,4: 0 1,4: 5	1-3: 2 1,2,4: 3 1,4: 2	1-3: 2 1,2,4: 0 1,4: 2
15		Not past month			34	16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 6 1,2,4: 0 1,4: 13	1-3: 4 1,2,4: 1 1,4: 4	1-3: 2 1,2,4: 0 1,4: 4
16	Not past month	Not past month			5	15,16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 1 1,2,4: 0 1,4: 1	1-3: 2	1-3: 0 1,2,4: 0 1,4: 1
17	Not past month	(Past month)		Missing	0	6,8,15	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DS 4. PMS	No Cases	No Cases	No Cases
18	(Past month)	Not past month	Missing		0	5,7,16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DC 4. PMC	No Cases	No Cases	No Cases
19	Not past month	Missing		Missing	0	4,6,15	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
20	Missing	Not past month	Missing		0	3,5,16	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC	No Cases	No Cases	No Cases
21	Not past month	Not past year			0	14,15	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases
22	Not past year	Not past month			1	13,16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 0 1,2,4: 0 1,4: 1	No Cases	No Cases
23	Not past year	Missing		Missing	0	4,6,13	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases
24	Missing	Not past year	Missing		0	3,5,14	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC	No Cases	No Cases	No Cases
25	Past year	Past year	Missing	Missing	0	5,6,11,12	1. R1/(R1+R2) 2. RC1*DC/(RC1+RC2) 3. RC1*(1-DC)*PMC/(RC1+RC2) 4. RS1*DS/(RS1+RS2) 5. RS1*(1-DS)*PMS/(RS1+RS2)	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
26		Past 3 years		Missing	0	2,6	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RS1*DS/ (RS1+RS2+RS3) 4. RS1*(1-DS)*PMS/ (RS1+RS2+RS3)	No Cases	No Cases	No Cases
27	Past 3 years		Missing		1	1,5	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/ (RC1+RC2+RC3) 4. RC1*(1-DC)* PMC/(RC1+RC2+ RC3)	1-3: 0 1,2,4: 0 1,4: 1	No Cases	No Cases
28	Past year but not past month, or past 3 years but not past year				21	17	1. R2/(R2+R3)	1-3: 9 1,2,4: 0 1,4: 1	1-3: 9 1,2,4: 1	1-3: 1
29		Past year but not past month, or past 3 years but not past year			38	18	1. R2/(R2+R3)	1-3: 20 1,2,4: 0 1,4: 2	1-3: 13 1,2,4: 0 1,4: 1	1-3: 2
30	Past year but not past month, or past 3 years but not past year	Past year but not past month, or past 3 years but not past year			12	17,18	1. R2/(R2+R3)	1-3: 8	1-3: 4	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
31	Past year but not past month, or past 3 years but not past year	(Past month)		Missing	0	6,8,17	1. R2/(R2+R3) 2. DS 3. PMS	No Cases	No Cases	No Cases
32	(Past month)	Past year but not past month, or past 3 years but not past year	Missing		1	5,7,18	1. R2/(R2+R3) 2. DC 3. PMC	1-3: 0 1,2,4: 0 1,4: 1	No Cases	No Cases
33	Past year but not past month, or past 3 years but not past year	Missing		Missing	0	4,6,17	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases
34	Missing	Past year but not past month, or past 3 years but not past year	Missing		0	3,5,18	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC	No Cases	No Cases	No Cases
35	Past year but not past month, or past 3 years but not past year	Not past year			0	14,17	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
36	Not past year	Past year but not past month, or past 3 years but not past year			0	13,18	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases
37	Past 3 years	Past 3 years	Missing	Missing	0	1,2,5,6	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/ (RC1+RC2+RC3) 4. RC1*(1-DC)* PMC/(RC1+RC2+ RC3) 5. RS1*DS/ (RS1+RS2+RS3) 6. RS1*(1-DS)* PMS/(RS1+RS2+ RS3)	No Cases	No Cases	No Cases
38	Not past month	Past year but not past month, or past 3 years but not past year			0	15,18	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases
39	Missing	Past 3 years	Missing	Missing	0	2,3,5,6	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/ (RC1+RC2+RC3) 4. RC1*(1-DC)* PMC/(RC1+RC2+ RC3) 5. DS 6. PMS	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
40	Past 3 years	(Past month)	Missing	Missing	0	1,5,6,8	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/(RC1+RC2+RC3) 4. RC1*(1-DC)*PMC/(RC1+RC2+RC3) 5. DS 6. PMS	No Cases	No Cases	No Cases
41	(Past month)	Not past year	Missing		0	5,7,14	1. R3/(R3+R4) 2. DC 3. PMC	No Cases	No Cases	No Cases
42	Past 3 years	Past year but not past month, or past 3 years but not past year	Missing		0	1,5,18	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3.RC1*DC/(RC1+RC2+RC3) 4. RC1*(1-DC)*PMC/(RC1+RC2+RC3)	No Cases	No Cases	No Cases
43	Past year but not past month, or past 3 years but not past year	Past 3 years		Missing	0	2,6,17	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3.RS1*DS/(RS1+RS2+RS3) 4. RS1*(1-DS)*PMS/(RS1+RS2+RS3)	No Cases	No Cases	No Cases
44	Past year but not past month, or past 3 years but not past year	Not past month			0	16,17	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
45	Past Year	Past year but not past month, or past 3 years but not past year	Missing		0	5,11,18	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/(RC1+RC2) 4.RC1*(1-DC)*PMC/(RC1+RC2)	No Cases	No Cases	No Cases

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month smokeless tobacco use | lifetime smokeless tobacco use)
2. R2 = P(Past year but not past month smokeless tobacco use | lifetime smokeless tobacco use)
3. R3 = P(Past 3 years but not past year smokeless tobacco use | lifetime smokeless tobacco use)
4. R4 = P(Lifetime but not past 3 years smokeless tobacco use | lifetime smokeless tobacco use)
5. RC1 = P(Past month chewing tobacco use | lifetime chewing tobacco use)
6. RC2 = P(Past year but not past month chewing tobacco use | lifetime chewing tobacco use)
7. RC3 = P(Past 3 years but not past year chewing tobacco use | lifetime chewing tobacco use)
8. RS1 = P(Past month snuff use | lifetime snuff use)
9. RS2 = P(Past year but not past month snuff use | lifetime snuff use)
10. RS3 = P(Past 3 years but not past year snuff use | lifetime snuff use)
11. DC = P(Daily chewing tobacco use | past month chewing tobacco use)
12. DS = P(Daily snuff use | past month snuff use)
13. PMC = P(Chewing tobacco use on a given day in the past month | past month use of chewing tobacco)
14. PMS = P(Snuff use on a given day in the past month | past month use of snuff)

Table E.35 Likeness Constraints for Pipe Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.36 Constraints and Portion of the Predictive Mean Vector for Pipe Recency and Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	4	None	1. R1	1,2: 2	1,2: 1	1,2: 2

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month pipe use | lifetime pipe use)

Table E.37 Logical Constraints for Various Drugs Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be less than or equal to recipient's maximum possible past year drug frequency of use Recipient's maximum possible drug frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used drug, as determined by the month of first drug use (2) if the maximum period recipient could have used drug is greater than 30, but recipient is a past month drug user with a nonmissing 30-day drug frequency, the past year drug frequency must be less than or equal to the maximum period (minus the number of days recipient did not use drug in the past month) (3) if recipient is not a past month drug user, the past year drug frequency must be less than or equal to the maximum period (minus 30)
LogC2	Donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than or equal to recipient's minimum possible past year drug frequency of use Recipient's minimum possible drug frequency of use in the past year is limited by the following factors: (1) if recipient is a past month drug user, it must be at least as much as the 30-day drug frequency (2) if recipient is not a past month drug user but is a past year drug user, it must be at least 1

Table E.37 Logical Constraints for Various Drugs Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC3	Donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than or equal to recipient's 30-day drug use
LogC4	Donor's 30-day drug use must be less than or equal to recipient's maximum number of days could have used drug in past 30 days
LogC5	Donor's 30-day drug use must be greater than or equal to recipient's minimum number of days could have used drug in past 30 days
LogC6	Donor's 30-day drug use must be greater than or equal to recipient's DR5DAY (# days had 5+ drinks in past 30 days)
LogC7	Donor's 30-day drug use must be greater than or equal to donor's proportion of past year drug use * recipient's maximum number of days could have used drug in past year [minus 335]
LogC8	Donor must be a past month drug user (drug recency = 1)
LogC9	If recipient's age at first drug use equals his or her current age, donor's 30-day drug frequency (1) cannot be greater than recipient's days between his or her interview date and date of first drug use (inclusive) and (2) cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC10	If recipient's age at first drug use equals his or her current age, (1) donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year cannot be greater than recipient's days between his or her interview date and date of first drug use (inclusive) and (2) donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC11	Recipient's estimated 30-day drug frequency is not given/legitimately skipped (estimated drug frequency ≠ 1-6)
LogC12	If recipient's age at first drug use equals his or her current age, (1) donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year cannot be greater than recipient's days between his or her interview date and date of first drug use (minus 30) and (2) donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year cannot be greater than recipient's days between the interview date and birthday (minus 30)
LogC13	Donor must be a past year but not past month drug user (drug recency = 2)
LogC14	Donor's DR5DAY value must be less than or equal to recipient's 30-day drug frequency
LogC15	If recipient's age at first drug use equals his or her current age, (1) donor's DR5DAY must be less than recipient's days between his or her interview date and date of first drug use (inclusive) and (2) donor's DR5DAY must be less than recipient's days between his or her interview date and birthday (inclusive)
LogC16	Donor must be a past month or past year but not past month drug user (drug recency = 1 or 2)
LogC17	Donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than or equal to donor's 30-day drug frequency
LogC18	If recipient's month/year of first drug use data indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than recipient's 30-day drug frequency

Table E.37 Logical Constraints for Various Drugs Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC19	If recipient's month/year of first drug use data indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than recipient's DR5DAY value
LogC20	If recipient's month/year of first drug use data indicate that he or she must have used drug at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month drug user, then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than donor's 30-day drug frequency
LogC21	If recipient's month/year of first drug use data indicate that he or she must have used drug at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than 1
LogC22	If recipient's month/year of first drug use data indicate that he or she must have used drug at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month drug user, then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than 1

Table E.38 Likeness Constraints for Various Drugs Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Donor's drug recency must match recipient's drug recency (when nonmissing)[1]

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.39 Constraints and Portion of the Predictive Mean Vector for Alcohol Recency and Frequency

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Mo. Freq.	30-Day Freq.	30-Day Binge Drink				12-17	18-25	26+
1	(Past month)	Missing	Missing		21	1,2,4-11,17,19	1. PM 2. PY	1-3: 2 1,3: 3	1-3: 2 1,3: 2	1-3: 10 1,3: 2
2	(Past month)		Missing		206	4-6,8,9,11	1. PM	1-3: 20	1-3: 74 1,3: 1	1-3: 110 1,3: 1
3	(Past month)	Missing			160	1-3,8,10,18	1. PY	1-3: 35 1,3: 1	1-3: 51	1-3: 73
4	(Past year but not past month)	Missing			134	1,2,12,13	1. PY	1-3: 60 1,3: 1	1-3: 30	1-3: 43
5	(Past month)			Missing	494	8,14,15	1. PMB	1-3: 43 1,3: 3	1-3: 177	1-3: 271
6	(Past month)		Missing	Missing	14	4,5,8,9,11	1. PM 2. PMB	1-3: 3 1,3: 3	1-3: 5	1-3: 3
7	(Past month)	Missing		Missing	80	1-3,8,10,14, 15,18	1. PY 2. PMB	1-3: 12 1,3: 9	1-3: 27 1,3: 1	1-3: 28 1,3: 3
8	(Past month)	Missing	Missing	Missing	26	1,2,4,5,7-11, 17,20	1. PM 2. PY 3. PMB	1-3: 1 1,3: 6	1-3: 4 1,3: 4	1-3: 5 1,3: 6
9	Past year		Missing	Missing	295	4,5,9,11,16	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. R1*PMB/(R1+R2)	1-3: 30 1,3: 26	1-3: 47 1,3: 42	1-3: 89 1,3: 61
10	Past year	Missing	Missing	Missing	80	1,2,4,5,7,9-12, 16,17,20	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY 4. R1*PMB/(R1+R2)	1-3: 1 1,3: 17	1-3: 0 1,3: 21	1-3: 1 1,3: 40

Table E.39 Constraints and Portion of the Predictive Mean Vector for Alcohol Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Mo. Freq.	30-Day Freq.	30-Day Binge Drink				12-17	18-25	26+
11	Lifetime (known)	Missing	Missing	Missing	298	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY 5. R1*PMB	1-3: 46 1,3: 158	1-3: 4 1,3: 68	1-3: 1 1,3: 21

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month alcohol use | lifetime alcohol use)
2. R2 = P(Past year but not past month alcohol use | lifetime alcohol use)
3. PM = P(Alcohol use on a given day in the past month | past month alcohol use)
4. PY = P(Alcohol use on a given day in the past year | past year alcohol use)
5. PMB = P(Past month binge alcohol use | past month alcohol use)

Table E.40 Constraints and Portion of the Predictive Mean Vector for Inhalants Recency and Frequency

#	Missingness Pattern: Recency	Missingness Pattern: 12-Month Freq.	Missingness Pattern: 30-Day Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
1	(Past month)	Missing	Missing	2	1,2,4,5,8-11, 17,20	1. PM 2. PY	1-3: 0 1,3: 1	1-3: 0 1,3: 1	No Cases
2	(Past month)		Missing	2	4,5,8,9,11	1. PM	1-3: 0 1,3: 2	No Cases	No Cases
3	(Past month)	Missing		10	1-3,8,10,18	1. PY	1-3: 7 1,3: 2	No Cases	1-3: 1
4	(Past year but not past month)	Missing		10	1,2,12,13	1. PY	1-3: 7 1,3: 2	No Cases	1-3: 0 1,3: 1
5	Past year		Missing	11	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	1-3: 0 1,3: 6	1-3: 0 1,3: 2	1-3: 0 1,3: 3
6	Past year	Missing	Missing	1	1,2,4,5,7,9-12, 16,17,20	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	1-3: 0 1,3: 1	No Cases	No Cases
7	Missing	Missing	Missing	156	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 9 1,3: 100	1-3: 0 1,3: 32	1-3: 0 1,3: 15

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month inhalants use | lifetime inhalants use)
2. R2 = P(Past year but not past month inhalants use | lifetime inhalants use)
3. PM = P(Inhalants use on a given day in the past month | past month inhalants use)
4. PY = P(Inhalants use on a given day in the past year | past year inhalants use)

Table E.41 Constraints and Portion of the Predictive Mean Vector for Marijuana Recency and Frequency

#	Missingness Pattern			Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Month Freq.	30-Day Freq.				12-17	18-25	26+
1	(Past month)	Missing	Missing	10	1,2,4,5,8-11, 17,20	1. PM 2. PY	1-3: 2 1,3: 1	1-3: 3	1-3: 1 1,3: 3
2	(Past month)		Missing	5	4,5,8,9,11	1. PM	1-3: 1	1-3: 4	No Cases
3	(Past month)	Missing		85	1,2,3,8,10,18	1. PY	1-3: 37 1,3: 1	1-3: 27	1-3: 20
4	(Past year but not past month)	Missing		45	1,2,12,13	1. PY	1-3: 21	1-3: 13	1-3: 11
5	Past year		Missing	83	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-3: 19 1,3: 10	1-3: 26 1,3: 4	1-3: 18 1,3: 6
6	Past year	Missing	Missing	106	1,2,4,5,7,9- 12,16,17,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-3: 0 1,3: 23	1-3: 8 1,3: 44	1-3: 6 1,3: 25
7	Missing	Missing	Missing	220	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 5 1,3: 126	1-3: 1 1,3: 57	1-3: 0 1,3: 31

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month marijuana use | lifetime marijuana use)
2. R2 = P(Past year but not past month marijuana use | lifetime marijuana use)
3. PM = P(Marijuana use on a given day in the past month | past month marijuana use)
4. PY = P(Marijuana use on a given day in the past year | past year marijuana use)

Table E.42 Constraints and Portion of the Predictive Mean Vector for Heroin Recency and Frequency

#	Missingness Pattern: Recency	Missingness Pattern: 12-Month Freq.	Missingness Pattern: 30-Day Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
1	(Past month)	Missing	Missing	0	1,2,4,5,8-11, 17,20	1. PM 2. PY	No Cases	No Cases	No Cases
2	(Past month)		Missing	0	4,5,8,9,11	1. PM	No Cases	No Cases	No Cases
3	(Past month)	Missing		1	1-3,8,10,18	1. PY	1-3: 0 1,3: 1	No Cases	No Cases
4	(Past year but not past month)	Missing		1	1,2,12,13	1. PY	No Cases	No Cases	1-3: 1
5	Past year		Missing	4	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	1-3: 0 1,3: 4
6	Past year	Missing	Missing	21	1,2,4,5,7,9-12, 16,17,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-3: 0 1,3: 1	1-3: 0 1,3: 12	1-3: 0 1,3: 8
7	Missing	Missing	Missing	12	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 0 1,3: 4	1-3: 0 1,3: 4	1-3: 0 1,3: 4

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month heroin use | lifetime heroin use)
2. R2 = P(Past year but not past month heroin use | lifetime heroin use)
3. PM = P(Heroin use on a given day in the past month | past month heroin use)
4. PY = P(Heroin use on a given day in the past year | past year heroin use)

Table E.43 Constraints and Portion of the Predictive Mean Vector for Tranquilizers Recency and Frequency

#	Missingness Pattern		Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Month Freq.				12-17	18-25	26+
1	(Past month)	Missing	8	1,2,10,21	1. PY	1-3: 2 1,3: 1	1-3: 1	1-3: 2 1,3: 2
2	(Past year but not past month)	Missing	13	1,2,10	1. PY	1-3: 3 1,3: 1	1-3: 5	1-3: 4
3	Past year		1	16	1. R1/(R1+R2)	No Cases	1-3: 1	No Cases
4	Past year	Missing	13	1,2,10,12,22	1. R1/(R1+R2) 2. PY	1-3: 1 1,3: 1	1-3: 1 1,3: 3	1-3: 3 1,3: 4
5	Missing	Missing	108	1,2,10,12	1. R1 2. R2 3. (R1+R2)*PY	1-3: 2 1,3: 36	1-3: 3 1,3: 35	1-3: 1 1,3: 31

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month tranquilizers use | lifetime tranquilizers use)
2. R2 = P(Past year but not past month tranquilizers use | lifetime tranquilizers use)
3. PY = P(Tranquilizers use on a given day in the past year | past year tranquilizers use)

Table E.44 Constraints and Portion of the Predictive Mean Vector for Sedatives Recency and Frequency

#	Missingness Pattern: Recency	Missingness Pattern: 12-Month Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
1	(Past month)	Missing	1	1,2,10,21	1. PY	1-3: 0 1,3: 1	No Cases	No Cases
2	(Past year but not past month)	Missing	5	1,2,12	1. PY	1-3: 1 1,3: 1	1-3: 1 1,3: 1	1-3: 0 1,3: 1
3	Past year		0	16	1. R1/(R1+R2)	No Cases	No Cases	No Cases
4	Past year	Missing	4	1,2,10,12,22	1. R1/(R1+R2) 2. PY	1-3: 0 1,3: 1	1-3: 0 1,3: 1	1-3: 0 1,3: 2
5	Missing	Missing	34	1,2,10,12	1. R1 2. R2 3. (R1+R2)*PY	1-3: 0 1,3: 14	1-3: 0 1,3: 9	1-3: 0 1,3: 11

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month sedatives use | lifetime sedatives use)
2. R2 = P(Past year but not past month sedatives use | lifetime sedatives use)
3. PY = P(Sedatives use on a given day in the past year | past year sedatives use)

Table E.45 Logical Constraints for Cocaine Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must be a past month cocaine user (cocaine recency = 1)
LogC2	Donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to recipient's maximum possible past year cocaine frequency of use Recipient's maximum possible cocaine frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used cocaine, as determined by the month of first cocaine use (2) if the maximum period recipient could have used cocaine is greater than 30, but recipient is a past month cocaine user with a nonmissing 30-day cocaine frequency, the past year cocaine frequency must be less than or equal to the maximum period (the number of days recipient did not use cocaine in the past month) (3) if recipient is not a past month cocaine user, the past year cocaine frequency must be less than or equal to the maximum period (minus 30)
LogC3	Donor's proportion of past year cocaine use * recipient's minimum number of days could have used cocaine in the past year must be greater than or equal to recipient's minimum possible past year cocaine frequency of use Recipient's minimum possible cocaine frequency of use in the past year is limited by the following factors: (1) if recipient is a past month cocaine user, it must be at least as much as the 30-day cocaine frequency (2) if recipient is not a past month cocaine user but is a past year cocaine user, it must be at least 1
LogC4	If recipient's age at first cocaine use equals his or her current age, and (1) if donor is a past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive) (2) if donor is a past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and date of first drug use (inclusive) (3) if donor is a past year but not past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (minus 30) (4) if donor is a past year but not past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and date of first drug use (minus 30)
LogC5	If recipient's age at first cocaine use equals his or her current age, (1) donor's 30-day cocaine use must be less than the number of days between recipient's interview date and birthday (inclusive) (2) donor's 30-day cocaine use must be less than the number of days between recipient's interview date and date of first drug use (inclusive)

Table E.45 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC6	Donor's 30-day cocaine use must be less than recipient's maximum number of days could have used cocaine in past 30 days
LogC7	Donor's 30-day cocaine use must be greater than recipient's minimum number of days could have used cocaine in past 30 days
LogC8	Donor's 30-day cocaine use must be greater than or equal to donor's proportion of past year drug use * recipient's maximum number of days could have used drug in past year (minus 335)
LogC9	Recipient's estimated cocaine 30-day frequency is not given/legitimately skipped (estimated cocaine frequency ≠ 1-6)
LogC10	Donor must be a past month or past year but not past month cocaine user (cocaine recency = 1 or 2)
LogC11	Donor's proportion of past year crack use * recipient's minimum number of days could have used crack in the past year must be greater than or equal to recipient's minimum possible past year crack frequency of use Recipient's minimum possible crack frequency of use in the past year is limited by the following factors: (1) if recipient is a past month crack user, it must be at least as much as the 30-day crack frequency (2) if recipient is not a past month crack user but is a past year crack user, it must be at least 1
LogC12	Donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be less than or equal to recipient's maximum possible past year crack frequency of use Recipient's maximum possible crack frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used crack, as determined by the month of first crack use (2) if the maximum period recipient could have used crack is greater than 30, but recipient is a past month crack user with a nonmissing 30-day crack frequency, the past year crack frequency must be less than or equal to the maximum period (the number of days recipient did not use in the past month) (3) if recipient is not a past month crack user, the past year crack frequency must be less than or equal to the maximum period (minus 30)

Table E.45 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC13	If recipient's age at first crack use equals his or her current age, and (1) if donor is a past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive) (2) if donor is a past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and date of first drug use (inclusive) (3) if donor is a past year but not past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (minus 30) (4) if donor is a past year but not past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and date of first drug use (minus 30)
LogC14	Donor's 30-day crack use must be less than number of days between recipient's interview date and birthday (inclusive)
LogC15	Donor's 30-day crack use must be less than recipient's maximum number of days could have used crack in past 30 days
LogC16	Donor's 30-day crack use must be greater than or equal to donor's proportion of past year drug use * recipient's maximum number of days could have used drug in past year (minus 335)
LogC17	Donor's 30-day crack use must be greater than recipient's minimum number of days could have used crack in past 30 days
LogC18	Recipient's estimated 30-day crack frequency is not given/legitimately skipped (estimated crack frequency ≠ 1-6)
LogC19	Donor must be a past month crack user (crack recency = 1)
LogC20	Donor must be a past month or past year not past month crack user (crack recency = 1, 2)
LogC21	Donor must be a crack user (crack recency ≠ 91)
LogC22	If recipient's month/year of first use data for cocaine indicate that he or she must have used cocaine at least once in the interval (1 year before interview, 30 days before interview), and if recipient is a past month user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be greater than recipient's 30-day cocaine frequency
LogC23	If recipient's month/year of first use data for cocaine indicate that he or she must have used cocaine at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be greater than donor's 30-day cocaine frequency
LogC24	If recipient's month/year of first use data for crack indicate that he or she must have used crack at least once in the interval (1 year before interview, 30 days before interview), and if recipient is a past month user, then donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be greater than recipient's 30-day crack frequency

Table E.45 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC25	If recipient's month/year of first use data for crack indicate that he or she must have used crack at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used in the past year must be greater than donor's 30-day crack frequency
LogC26	Donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be less than or equal to donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year
LogC27	If recipient is a past month user of cocaine or cocaine recency is missing, if donor is a past month user of cocaine, and if donor is a past year not past month user of crack, then donor's proportion of days used crack in the last year * recipient's maximum period used crack in the past year cannot be larger than recipient's 12-month frequency minus donor's 30-day frequency

Table E.46 Likeness Constraints for Cocaine Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Donor's cocaine recency must match recipient's drug recency (when nonmissing)[1]
LikC4	Donor's crack recency agrees with recipient's crack recency (when nonmissing)

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.47 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.				12-17	18-25	26+
1	(Past year)		Missing				23	2-4,10,22	1. PY	1-4: 0 1-3: 0 1,3: 2	1-4: 6 1-3: 1	1-4: 6 1-3: 4 1,3: 4
2	(Past year)	(Past year)		Missing			1	11-13,20,24	1. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
3	(Past year)	(Past year)	Missing	Missing			0	2-4,10-13, 20,22,24,26	1. PY	No Cases	No Cases	No Cases
4	(Past month)				Missing		12	1,5-7,9	1. PM	1-4: 0 1-3: 0 1,3: 0 3: 2	1-4: 0 1-3: 0 1,3: 1 3: 1	1-4: 0 1-3: 2 1,3: 6
5	(Past month)	(Past month)				Missing	0	14-16,18,19	1. PM	No Cases	No Cases	No Cases
6	(Past month)	(Past month)			Missing	Missing	0	1,5-7,9,14-16, 18,19	1. PM	No Cases	No Cases	No Cases
7	(Past month)		Missing		Missing		5	1-9,23	1. PM 2. PY	No Cases	1-4: 0 1-3: 0 1,3: 1	1-4: 0 1-3: 0 1,3: 4
8	(Past month)	(Past year)		Missing	Missing		1	1,5-7,9,11-13, 20,24	1. PM 2. PY	1-4: 0 1-3: 0 1,3: 0 3: 0 None: 0 None*: 1	No Cases	No Cases
9	(Past month)	(Past year)	Missing	Missing	Missing		1	1-9,11-13, 20,23,24,26	1. PM 2. PY	No Cases	1-4: 0 1-3: 0 1,3: 0 3: 1	No Cases
10	(Past month)	(Past month)	Missing			Missing	0	2-4,14-16, 18,19,22	1. PM 2. PY	No Cases	No Cases	No Cases
11	(Past month)	(Past month)		Missing		Missing	0	11-19,25	1. PM 2. PY	No Cases	No Cases	No Cases

Table E.47 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.				12-17	18-25	26+
12	(Past month)	(Past month)	Missing	Missing		Missing	0	2-4,11-19, 22,25,26	1. PM 2. PY	No Cases	No Cases	No Cases
13	(Past month)	(Past month)	Missing		Missing	Missing	0	1-9,14-16, 18,19,23	1. PM 2. PY	No Cases	No Cases	No Cases
14	(Past month)	(Past month)		Missing	Missing	Missing	0	1,5-7,9,11-19, 25	1. PM 2. PY	No Cases	No Cases	No Cases
15	(Past month)	(Past month)	Missing	Missing	Missing	Missing	0	1-9,11-19, 23,25,26	1. PM 2. PY	No Cases	No Cases	No Cases
16	(Past month)	Past year				Missing	0	14-16,18,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
17	(Past month)	Past year			Missing	Missing	0	1,5-7,9,14-16, 18,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
18	(Past month)	Past year		Missing		Missing	1	11-18,20,25	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-4: 0 1-3: 0 1,3: 1	No Cases
19	(Past month)	Past year		Missing	Missing	Missing	0	1,5-7,9,11-18, 20,25,27	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
20	(Past month)	Missing		Missing		Missing	0	11-18,21	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
21	(Past month)	Missing		Missing	Missing	Missing	0	1,5-7,9,11-18, 21,27	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
22	(Past month)	Missing	Missing	Missing	Missing	Missing	0	1-9,11-18, 21,23,26	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases

Table E.47 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.				12-17	18-25	26+
23	Missing		Missing		Missing		57	2-9	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 6	1-4: 0 1-3: 0 1,3: 34	1-4: 0 1-3: 0 1,3: 17
24	Missing	Missing	Missing	Missing	Missing	Missing	15	2-9,11-18, 21,26	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 2	1-4: 0 1-3: 0 1,3: 7	1-4: 0 1-3: 0 1,3: 6
25	Past year				Missing		9	5-7,9,10	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-4: 0 1-3: 0 1,3: 2	1-4: 0 1-3: 0 1,3: 3	1-4: 0 1-3: 0 1,3: 4
26	Past year	Past year			Missing	Missing	0	5-7,9,10, 14-16,18,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
27	Past year		Missing		Missing		4	2-10,23	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-4: 0 1-3: 0 1,3: 2	1-4: 0 1-3: 0 1,3: 1	1-4: 0 1-3: 0 1,3: 1
28	Past year	Past year	Missing		Missing	Missing	0	2-10,14-16, 18,20,23	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
29	Past year	Past year		Missing	Missing	Missing	1	5-7,9-18, 20,25,27	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-4: 0 1-3: 0 1,3: 1	No Cases
30	Past year	Past year	Missing	Missing	Missing	Missing	1	2-18,20,23,25, 26	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
31	Past year	Missing		Missing	Missing	Missing	0	5-7,9-18,21,27	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases

Table E.47 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern: Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
32	Past year	Missing	Missing	Missing	Missing	Missing	0	2-18,21,23,26	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	No Cases
33	(Past month)	Missing	Missing	Missing		Missing	1	2-4,10-18, 21,22,26	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1

*A donor could not be found for these cases. Imputed values were randomly assigned within the bounds created during the editing process.

NOTE: Cocaine users included crack users and cocaine users who were not crack users.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month cocaine use | lifetime cocaine use)
2. R2 = P(Past year but not past month cocaine use | lifetime cocaine use)
3. PM = P(Cocaine use on a given day in the past month | past month use of cocaine)
4. PY = P(Cocaine use on a given day in the past year | past year use of cocaine)

Table E.48 Logical Constraints for Hallucinogens Recency and Frequency

<table>
<tr><th>Constraint #</th><th>Logical Constraint</th></tr>
<tr><td>LogC1</td><td>Donor's proportion of past year hallucinogen use * recipient's maximum number of days could have used hallucinogens in the past year must be less than or equal to recipient's maximum possible past year hallucinogen frequency of use

Recipient's maximum possible hallucinogen frequency of use in the past year is limited by the following factors:
(1) it must be less than or equal to the maximum period recipient could have used hallucinogens, as determined by the month of first hallucinogen use
(2) if the maximum period recipient could have used hallucinogens is greater than 30, but recipient is a past month hallucinogen user with a nonmissing 30-day hallucinogen frequency, the past year hallucinogen frequency must be less than or equal to the maximum period (the number of days recipient did not use hallucinogens in the past month)
(3) if recipient is not a past month hallucinogen user, the past year hallucinogen frequency must be less than or equal to the maximum period (minus 30)</td></tr>
<tr><td>LogC2</td><td>Donor's proportion of past year hallucinogen use * recipient's minimum number of days could have used hallucinogens in the past year must be greater than or equal to recipient's minimum possible past year hallucinogen frequency of use

Recipient's minimum possible hallucinogen frequency of use in the past year is limited by the following factors:
(1) if recipient is a past month hallucinogen user, it must be at least as much as the 30-day hallucinogen frequency
(2) if recipient is not a past month hallucinogen user but is a past year hallucinogen user, it must be at least 1</td></tr>
<tr><td>LogC3</td><td>If recipient's age at first hallucinogen use equals his or her current age, and
(1) if donor is a past month hallucinogen user, then donor's proportion of past year hallucinogen use * recipient's maximum number of days could have used hallucinogens in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
(2) if donor is a past month hallucinogen user, then donor's proportion of past year hallucinogen use * recipient's maximum number of days could have used hallucinogens in the past year must be less than or equal to the number of days between recipient's interview date and date of first drug use (inclusive)
(3) if donor is a past year but not past month hallucinogen user, then donor's proportion of past year hallucinogen use * recipient's maximum number of days could have used hallucinogens in the past year must be less than or equal to the number of days between recipient's interview date and birthday (minus 30)
(4) if donor is a past year but not past month hallucinogen user, then donor's proportion of past year hallucinogen use * recipient's maximum number of days could have used hallucinogens in the past year must be less than or equal to the number of days between recipient's interview date and date of first drug use (minus 30)</td></tr>
<tr><td>LogC4</td><td>If recipient's age at first hallucinogen use equals his or her current age,
(1) donor's 30-day hallucinogen use must be less than the number of days between recipient's interview date and birthday (inclusive)
(2) donor's 30-day hallucinogen use must be less than the number of days between recipient's interview date and date of first drug use (inclusive)</td></tr>
</table>

Table E.48 Logical Constraints for Hallucinogens Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC5	Donor's 30-day hallucinogen use must be less than recipient's maximum number of days could have used hallucinogens in past 30 days
LogC6	Donor's 30-day hallucinogen use must be greater than recipient's minimum number of days could have used hallucinogens in past 30 days
LogC7	Donor must be an LSD user (LSD recency ≠ 91)
LogC8	Donor must be a PCP user (PCP recency ≠ 91)
LogC9	Donor must be an Ecstasy user (Ecstasy recency ≠ 91)
LogC10	Donor must be a past month hallucinogens user (hallucinogen recency = 1)
LogC11	Donor must be a hallucinogen past year but not past month or past month hallucinogen user (hallucinogen recency = 1 or 2)
LogC12	Donor must be an LSD past year but not past month or past month LSD user (LSD recency = 1 or 2)
LogC13	Donor must be a PCP past year but not past month or past month PCP user (PCP recency = 1 or 2)
LogC14	Donor must be an Ecstasy past year but not past month or past month Ecstasy user (Ecstasy recency = 1 or 2)
LogC15	Donor cannot have used hallucinogens more recently than recipient
LogC16	If recipient is a past month hallucinogens user and recipient's month/year of first use data for hallucinogens indicate that he or she must have used hallucinogens at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year hallucinogens use * recipient's maximum number of days could have used hallucinogens in the past year must be greater than recipient's 30-day hallucinogen frequency
LogC17	If recipient's month/year of first use data for hallucinogens indicate that he or she must have used hallucinogens at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month hallucinogens user, then donor's proportion of past year use * recipient's maximum number of days could have used hallucinogens in the past year must be greater than donor's 30-day hallucinogen frequency
LogC18	If recipient has never used hallucinogens other than LSD, PCP, and Ecstasy, then donor must not have recency values that would cause recipient to have imputation-revised recency for overall hallucinogens less than the minimum of the imputation-revised recencies for LSD, PCP, and Ecstasy

Table E.49 Likeness Constraints for Hallucinogens Recency and Frequency

Constraint #	Likeness Constraint
LikC1	Donor's parent drug recency must match recipient's drug recency (when nonmissing)[1]
LikC2	Donor's indicator of lifetime use of other hallucinogens = Recipient's indicator of lifetime use of other hallucinogens
LikC3	State rank of donor = state rank of recipient
LikC4	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC5	Donor's recencies for LSD, PCP, and Ecstasy agree with recipient's recencies for LSD, PCP, and Ecstasy (when nonmissing)
LikC6	If recipient is a lifetime but not past year user or has never used LSD, PCP, or Ecstasy, then donor must also be a lifetime but not past year user or have never used LSD, PCP, or Ecstasy

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
1		Missing					2	7,15	1. R1 2. R2	1-5: 0 1-4,6: 0 1-4: 1	1-5: 0 1-4,6: 0 1-4: 1	No Cases
2			Missing				1	8,15	1. R1 2. R2	1-5: 0 1-4,6: 0 1-4: 1	No Cases	No Cases
3		Missing	Missing				0	7,8,15,18	1. R1 2. R2	No Cases	No Cases	No Cases
4				Missing			3	9,15	1. R1 2. R2	1-5: 0 1-4,6: 0 1-4: 1	1-5: 0 1-4,6: 1 1-4: 1	No Cases
5		Missing		Missing			0	7,9,15,18	1. R1 2. R2	No Cases	No Cases	No Cases
6			Missing	Missing			0	8,9,15,18	1. R1 2. R2	No Cases	No Cases	No Cases
7		Missing	Missing	Missing			0	7-9,15,18	1. R1 2. R2	No Cases	No Cases	No Cases
8	(Past year)				Missing		13	1-3,11,16	1. PY	1-5: 1 1-4,6: 1 1-4: 2	1-5: 2 1-4,6: 4 1-4: 1	1-5: 0 1-4,6: 0 1-4: 2
9	(Past year)	Missing			Missing		0	1-3,7,11,16	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
10	(Past year)		Missing		Missing		0	1-3,8,11,16	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
11	(Past year)	Missing	Missing		Missing		0	1-3,7,8,11,16, 18	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
12	(Past year)			Missing	Missing		0	1-3,9,11,16	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
13	(Past year)	Missing		Missing	Missing		0	1-3,7,9,11,16, 18	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
14	(Past year)		Missing	Missing	Missing		0	1-3,8,9,11,16, 18	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
15	(Past year)	Missing	Missing	Missing	Missing		0	1-3,7-9,11,16, 18	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
16	(Past month)					Missing	5	4-6,10,19	1. PM	1-5: 0 1-4,6: 2	1-5: 1 1-4,6: 2	No Cases
17	(Past month)	Missing				Missing	0	4-7,10,19	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
18	(Past month)		Missing			Missing	0	4-6,8,10,19	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
19	(Past month)	Missing	Missing			Missing	0	4-8,10,18,19	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
20	(Past month)			Missing		Missing	0	4-6,9,10,19	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
21	(Past month)	Missing		Missing		Missing	0	4-7,9,10,18, 19	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
22	(Past month)		Missing	Missing		Missing	0	4-6,8-10, 18,19	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
23	(Past month)	Missing	Missing	Missing		Missing	0	4-10,18,19	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
24	(Past month)				Missing	Missing	0	1-6,10,17,19	1. PM 2. PY	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
25	(Past month)	Missing			Missing	Missing	0	1-7,10,17,19	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
26	(Past month)		Missing		Missing	Missing	0	1-6,8,10,17, 19	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
27	(Past month)	Missing	Missing		Missing	Missing	0	1-8,10,17-19	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
28	(Past month)			Missing	Missing	Missing	0	1-6,9,10,17, 19	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
29	(Past month)	Missing		Missing	Missing	Missing	0	1-7,9,10, 17-19	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
30	(Past month)		Missing	Missing	Missing	Missing	0	1-6,8-10, 17-19	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
31	(Past month)	Missing	Missing	Missing	Missing	Missing	0	1-10,17-19	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
32	(Past month)	Past year					1	12	1. R1/(R1+R2)	No Cases	1-5: 0 1-4,6: 0 1-4: 1	No Cases
33	(Past month)		Past year				0	13	1. R1/(R1+R2)	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

	Missingness Pattern									Likeness Constraints: Number of Cases, by Age Group		
#	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	12-17	18-25	26+
34	(Past month)	Past year	Past year				0	12,13,18	1. R1/(R1+R2)	No Cases	No Cases	No Cases
35	(Past month)			Past year			0	14	1. R1/(R1+R2)	No Cases	No Cases	No Cases
36	(Past month)	Past year		Past year			0	12,14,18	1. R1/(R1+R2)	No Cases	No Cases	No Cases
37	(Past month)		Past year	Past year			0	13,14,18	1. R1/(R1+R2)	No Cases	No Cases	No Cases
38	(Past month)	Past year	Past year	Past year			0	12-14,18	1. R1/(R1+R2)	No Cases	No Cases	No Cases
39	(Past month)	Past year			Missing		0	1-3,11,12,16	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
40	(Past month)		Past year		Missing		0	1-3,11,13,16	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
41	(Past month)	Past year	Past year		Missing		0	1-3,11-13, 16,18	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
42	(Past month)			Past year	Missing		0	1-3,11,14,16	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
43	(Past month)	Past year		Past year	Missing		0	1-3,11,12,14, 16,18	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
44	(Past month)		Past year	Past year	Missing		0	1-3,11,13,14, 16,18	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
45	(Past month)	Past year	Past year	Past year	Missing		0	1-3,11-14, 16,18	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
46	(Past month)	Past year				Missing	0	4-7,10,12,19	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
47	(Past month)		Past year			Missing	0	4-6,8,10,13, 19	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
48	(Past month)	Past year	Past year			Missing	0	4-6,10,12,13, 18,19	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
49	(Past month)			Past year		Missing	0	4-6,9,10,14, 19	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
50	(Past month)	Past year		Past year		Missing	0	4-6,10,12,14, 18,19	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
51	(Past month)		Past year	Past year		Missing	0	4-6,10,13,14, 18,19	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
52	(Past month)	Past year	Past year	Past year		Missing	0	4-6,10,12-14, 18,19	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
53	(Past month)	Past year			Missing	Missing	0	1-6,10,12,17, 19	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
54	(Past month)		Past year		Missing	Missing	0	1-6,10,13,17, 19	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
55	(Past month)	Past year	Past year		Missing	Missing	0	1-6,10,12,13, 17-19	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
56	(Past month)			Past year	Missing	Missing	0	1-6,10,14,17, 19	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
57	(Past month)	Past year		Past year	Missing	Missing	0	1-6,10,12,14, 17-19	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
58	(Past month)		Past year	Past year	Missing	Missing	0	1-6,10,13,14, 17-19	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
59	(Past month)	Past year	Past year	Past year	Missing	Missing	0	1-6,10,12-14, 17-19	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
60	Missing	(Not past year)	(Not past year)	(Not past year)	Missing	Missing	43	1-6,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4,6: 0 1-4: 17	1-5: 0 1-4,6: 0 1-4: 16	1-5: 0 1-4,6: 0 1-4: 10

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
61	Missing	Missing	(Not past year)	(Not past year)	Missing	Missing	51	1-7,18,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4,6: 0 1-4: 6	1-5: 0 1-4,6: 0 1-4: 21	1-5: 0 1-4,6: 0 1-4: 24
62	Missing	(Not past year)	Missing	(Not past year)	Missing	Missing	12	1-6,8,18,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4,6: 0 1-4: 2	1-5: 0 1-4,6: 0 1-4: 3	1-5: 0 1-4,6: 0 1-4: 7
63	Missing	Missing	Missing	(Not past year)	Missing	Missing	3	1-8,18,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	1-5: 0 1-4,6: 0 1-4: 2	1-5: 0 1-4,6: 0 1-4: 1
64	Missing	(Not past year)	(Not past year)	Missing	Missing	Missing	38	1-6,9,18,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4,6: 0 1-4: 7	1-5: 0 1-4,6: 0 1-4: 21	1-5: 0 1-4,6: 0 1-4: 10
65	Missing	Missing	(Not past year)	Missing	Missing	Missing	6	1-7,9,18,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4,6: 0 1-4: 1	1-5: 0 1-4,6: 0 1-4: 4	1-5: 0 1-4,6: 0 1-4: 1
66	Missing	(Not past year)	Missing	Missing	Missing	Missing	0	1-6,8,9,18,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
67	Missing	Missing	Missing	Missing	Missing	Missing	3	1-9,18,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4,6: 0 1-4: 2	No Cases	1-5: 0 1-4,6: 0 1-4: 1
68	Past year	(Not past month)	(Not past month)	(Not past month)		Missing	10	4-6,11,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4,6: 0 1-4: 4	1-5: 1 1-4,6: 1 1-4: 4	No Cases
69	Past year	Missing	(Not past month)	(Not past month)		Missing	7	4-7,11,18,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4,6: 0 1-4: 1	1-5: 0 1-4,6: 0 1-4: 5	1-5: 0 1-4,6: 0 1-4: 1

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
70	Past year	(Not past month)	Missing	(Not past month)		Missing	1	4-6,8,11,18, 19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-5: 0 1-4,6: 0 1-4: 1	No Cases
71	Past year	Missing	Missing	(Not past month)		Missing	0	4-8,11,18,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
72	Past year	(Not past month)	(Not past month)	Missing		Missing	4	4-6,9,11,18, 19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4,6: 0 1-4: 1	1-5: 1 1-4,6: 0 1-4: 2	No Cases
73	Past year	Missing	(Not past month)	Missing		Missing	0	4-7,9,11,18, 19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
74	Past year	(Not past month)	Missing	Missing		Missing	0	4-6,8,9,11,18, 19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
75	Past year	Missing	Missing	Missing		Missing	1	4-9,11,18,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4,6: 0 1-4: 1	No Cases	No Cases
76	Past year	(Not past month)	(Not past month)	(Not past month)	Missing	Missing	5	1-6,11,17,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4,6: 0 1-4: 2	1-5: 0 1-4,6: 0 1-4: 2	1-5: 0 1-4,6: 0 1-4: 1
77	Past year	Missing	(Not past month)	(Not past month)	Missing	Missing	0	1-7,11,17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
78	Past year	(Not past month)	Missing	(Not past month)	Missing	Missing	0	1-6,8,11, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
79	Past year	Missing	Missing	(Not past month)	Missing	Missing	0	1-8,11,17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
80	Past year	(Not past month)	(Not past month)	Missing	Missing	Missing	1	1-6,9,11, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4,6: 0 1-4: 1	No Cases	No Cases
81	Past year	Missing	(Not past month)	Missing	Missing	Missing	1	1-7,9,11, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4,6: 0 1-4: 1	No Cases	No Cases
82	Past year	(Not past month)	Missing	Missing	Missing	Missing	0	1-6,8,9,11, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
83	Past year	Missing	Missing	Missing	Missing	Missing	0	1-9,11,17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
84	Past year	Past year	(Not past month)	(Not past month)		Missing	3	4-6,11,12,18, 19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4,6: 0 1-4: 2	1-5: 0 1-4,6: 0 1-4: 1	No Cases
85	Past year	(Not past month)	Past year	(Not past month)		Missing	1	4-6,11,13,18, 19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4,6: 0 1-4: 1	No Cases	No Cases
86	Past year	Past year	Past year	(Not past month)		Missing	0	4-6,11-13, 18,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
87	Past year	(Not past month)	(Not past month)	Past year		Missing	3	4-6,11,14,18, 19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-5: 0 1-4,6: 0 1-4: 3	No Cases
88	Past year	Past year	(Not past month)	Past year		Missing	2	4-6,11,12,14, 18,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4,6: 0 1-4: 1	1-5: 0 1-4,6: 0 1-4: 1	No Cases
89	Past year	(Not past month)	Past year	Past year		Missing	0	4-6,11,13,14, 18,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
90	Past year	Past year	Past year	Past year		Missing	0	4-6,11-14, 18,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
91	Past year	Past year	(Not past month)	(Not past month)	Missing	Missing	4	1-6,11,12, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4,6: 0 1-4: 3	1-5: 0 1-4,6: 0 1-4: 1	No Cases
92	Past year	(Not past month)	Past year	(Not past month)	Missing	Missing	1	1-6,11,13, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4,6: 0 1-4: 1	No Cases	No Cases
93	Past year	Past year	Past year	(Not past month)	Missing	Missing	0	1-6,11-13, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
94	Past year	(Not past month)	(Not past month)	Past year	Missing	Missing	1	1-6,11,14, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-5: 0 1-4,6: 0 1-4: 1	No Cases
95	Past year	Past year	(Not past month)	Past year	Missing	Missing	0	1-6,11,12,14, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
96	Past year	(Not past month)	Past year	Past year	Missing	Missing	0	1-6,11,13,14, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
97	Past year	Past year	Past year	Past year	Missing	Missing	0	1-6,11-14, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
98	Past year	Missing	Missing	Past year		Missing	0	4-8,11,14,18, 19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallucinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.				12-17	18-25	26+
99	Past year	(Not past month)	Past year	Missing	Missing	Missing	0	1-6,9,11,13, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
100	Past year	Missing	Past year	(Not past month)	Missing	Missing	0	1-7,11,13, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
101	Past year	Missing	Past year	Missing	Missing	Missing	0	1-7,9,11,13, 17-19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases

NOTE: Hallucinogen users included users of LSD, users of PCP, and users of Ecstasy.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month hallucinogens use | lifetime hallucinogens use)
2. R2 = P(Past year but not past month hallucinogens use | lifetime hallucinogens use)
3. PM = P(Hallucinogens use on a given day in the past month | past month hallucinogens use)
4. PY = P(Hallucinogens use on a given day in the past year | past year hallucinogens use)

Table E.51 Logical Constraints for Stimulants and Pain Relievers Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year must be less than or equal to recipient's maximum possible past year parent drug frequency of use Recipient's maximum possible parent drug frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used parent drug, as determined by the month of first parent drug use (2) if recipient is not a past month parent drug user, the past year parent drug frequency must be less than or equal to the maximum period (minus 30)
LogC2	Donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year must be greater than or equal to recipient's minimum possible past year parent drug frequency of use (For these drugs, the minimum possible past year parent drug frequency of use is always 1.)
LogC3	Recipient's proportion of past year parent drug use * maximum number of days could have used parent drug in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC4	Donor must be a past month parent drug user (parent drug recency = 1)
LogC5	If recipient's age at first parent drug use equals his or her current age, (1) donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year cannot be greater than recipient's days between his or her interview date and date of first parent drug use (inclusive) and (2) donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC6	Donor must be a past year but not past month parent drug user (parent drug recency = 2)
LogC7	If recipient's age at first parent drug use equals his or her current age, (1) donor's proportion of past year parent drug use* recipient's maximum number of days could have used parent drug in the past year cannot be greater than recipient's days between his or her interview date and date of first parent drug use (minus 30) and (2) donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (minus 30)
LogC8	Donor must be a past month or past year but not past month parent drug user (parent drug recency = 1 or 2)
LogC9	Donor's parent drug recency must equal recipient's parent drug recency, or donor's parent drug recency must equal recipient's parent drug recency (minus 10)
LogC10	Donor must be a past month, past year but not past month, or lifetime but not past year child drug user (child drug recency = 1, 2, or 3)
LogC11	If the number of days between recipient's interview and birthday (inclusive) is between 0 and 30, child drug recency must not equal 2 or 3
LogC12	If the number of days between recipient's interview and birthday (inclusive) is between 0 and 365, child drug recency must not equal 3

Table E.51 Logical Constraints for Stimulants and Pain Relievers Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC13	If recipient's age at first parent drug use equals his or her current age, or recipient's age at first child drug use equals his or her current age, or recipient's number of days between his or her interview date and date at first child drug use is less than 30, then donor's parent drug recency must not equal 3
LogC14	Donor must be a past month or past year but not past month child drug user (child drug recency = 1 or 2)
LogC15	Donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year must be less than or equal to recipient's maximum possible past year child drug frequency of use Recipient's maximum possible child drug frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used child drug, as determined by the month of first child drug use (2) if recipient is not a past month child drug user, the past year child drug frequency must be less than or equal to the maximum period (minus 30)
LogC16	Donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year must be greater than or equal to recipient's minimum possible past year child drug frequency of use (For these drugs, the minimum possible past year child drug frequency of use is always 1.)
LogC17	Recipient's proportion of past year child drug use * maximum number of days could have used child drug in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC18	If recipient's age at first child drug use equals his or her current age, (1) donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year cannot be greater than recipient's days between his or her interview date and date of first child drug use (inclusive) and (2) donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC19	If recipient's age at first child drug use equals his or her current age, (1) donor's proportion of past year child drug use* recipient's maximum number of days could have used child drug in the past year cannot be greater than recipient's days between his or her interview date and date of first child drug use (minus 30) and (2) donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (minus 30)
LogC20	Donor must be a past month child drug user (child drug recency = 1)
LogC21	Donor must be a past year but not past month child drug user (child drug recency = 2)
LogC22	Donor must be a past month, past year but not past month, or lifetime but not past year parent drug user (parent drug recency = 1, 2, or 3)
LogC23	If recipient's month/year of first use data for the parent drug indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year must be greater than 1

Table E.51 Logical Constraints for Stimulants and Pain Relievers Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC24	If recipient's month/year of first use data for the child drug indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year must be greater than 1
LogC25	If recipient's month/year of first use data for the parent drug indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month parent drug user, then donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year must be greater than 1
LogC26	If recipient's month/year of first use data for the child drug indicate that he or she must have used child drug at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month child drug user, then donor's proportion of past year use * recipient's maximum number of days could have used in the past year must be greater than 1
LogC27	If recipient is not a lifetime user of any type of the parent drug except for the child drug, then donor must not have used parent drug more recently than recipient has used child drug
LogC28	If recipient is not a lifetime user of any type of the parent drug except for the child drug, then donor must not have used parent drug more recently than donor has used child drug
LogC29	Donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year cannot be greater than donor's proportion of past year parent drug use * recipient's maximum number of days could have used child drug in the past year

Table E.52 Likeness Constraints for Stimulants and Pain Relievers Recency and Frequency

Constraint #	Likeness Constraint
LikC1	Donor's drug recency must match recipient's drug recency (when nonmissing)[1]
LikC2	State rank of donor = state rank of recipient
LikC3	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC4	Donor's methamphetamine recency agrees with recipient's methamphetamine recency (when nonmissing)
LikC5	Donor's OxyContin recency agrees with recipient's OxyContin recency (when nonmissing)

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.53 Constraints and Portion of the Predictive Mean Vector for Stimulants Recency and Frequency

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Stimulant Recency	Meth. Recency	Stimulant 12-Mo. Freq.	Meth. 12-Mo. Freq.				12-17	18-25	26+
1	(Past month)		Missing		12	1-5,23	1. PY	1-4: 2 1-3: 1 1,2: 3	1-4: 1 1-3: 1 1,2: 2	1-4: 2
2	(Past year but not past month)		Missing		6	1-3,6,7	1. PY	1-4: 2 1-3: 0 1,2: 1	1-4: 0 1-3: 1	1-4: 1 1-3: 0 1,2: 1
3	Past year				0	8,27	1. R1/(R1+R2)	No Cases	No Cases	No Cases
4	Past year		Missing		12	1-3,5,7,8,25,27	1. R1/(R1+R2) 2. PY	1-4: 0 1-3: 0 1,2: 1	1-4: 2 1-3: 0 1,2: 4	1-4: 1 1-3: 0 1,2: 4
5	Missing		Missing		47	1,2,5,7,22,25,27	1. R1 2. R2 3. (R1+R2)*PY	1-4: 0 1-3: 0 1,2: 21	1-4: 1 1-3: 0 1,2: 16	1-4: 0 1-3: 0 1,2: 9
6	(Past month)	(Past month)		Missing	1	4,15-18,20,24	1. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,2: 1
7	(Past year not missing)	(Past year not past month)		Missing	2	15-17,19,21	1. PY	1-4: 0 1-3: 0 1,2: 1	No Cases	1-4: 1
8	(Past month)	Past year			0	14	1. R1/(R1+R2)	No Cases	No Cases	No Cases
9	(Past month)	Past year	Missing		0	1-3,5,7,14,23	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
10	(Past month)	Past year		Missing	1	14-19,26	1. R1/(R1+R2) 2. PY	No Cases	1-4: 0 1-3: 0 1,2: 1	No Cases
11	(Past month)	Past year	Missing	Missing	0	1-3,5,7,14-19, 23,26,29	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
12	(Past year not missing)	Missing		Missing	2	10,15,16,18,19, 22,29	1. R1 2. R2 3. (R1+R2)*PY	1-4: 0 1-3: 0 1,2: 1	No Cases	1-4: 0 1-3: 0 1,2: 1

Table E.53 Constraints and Portion of the Predictive Mean Vector for Stimulants Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Stimulant Recency	Meth. Recency	Stimulant 12-Mo. Freq.	Meth. 12-Mo. Freq.				12-17	18-25	26+
13	(Past month)	(Past month)	Missing	Missing	1	1-5,15-18,20, 23,24,29	1. PY	1-4: 0 1-3: 0 1,2: 1	No Cases	No Cases
14	(Past month)	(Past year not past month)	Missing	Missing	0	1-5,15-17,19, 21,23,29	1. PY	No Cases	No Cases	No Cases
15	(Past year not past month)	(Past year not past month)	Missing	Missing	1	1-3,6,7,15-17, 19,21,29	1. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,2: 1
16	Past year	Past year			1	8,14,28	1. R1/(R1+R2)	1-4: 0 1-3: 0 1,2: 1	No Cases	No Cases
17	Past year	Past year	Missing		0	1-35,7,8,14, 25,28	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
18	Past year	Past year		Missing	2	8,14-19,19, 26,28	1. R1/(R1+R2) 2. PY	1-4: 0 1-3: 0 1,2: 1	1-4: 0 1-3: 0 1,2: 1	No Cases
19	Past year	Past year	Missing	Missing	3	1-3,5,7,8,14-19, 25,26,28,29	1. R1/(R1+R2) 2. PY	1-4: 0 1-3: 0 1,2: 1	No Cases	1-4: 0 1-3: 0 1,2: 2
20	Past year	Missing		Missing	4	8,10,15,16,18, 19,28,29	1. R1/(R1+R2) 2. PY	1-4: 0 1-3: 0 1,2: 2	1-4: 0 1-3: 0 1,2: 2	No Cases
21	Past year	Missing	Missing	Missing	0	1-3,5,7,8,10,12, 15,16,18,19,25, 28,29	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
22	(Past month)	Missing		Missing	0	4,10,15,16,18, 19,29	1. R1 2. R2 3. (R1+R2)*PY	No Cases	No Cases	No Cases
23	(Past month)	Missing	Missing	Missing	0	1-5,10,15,16,18, 19,23,29	1. R1 2. R2 3. (R1+R2)*PY	No Cases	No Cases	No Cases
24	Missing	Missing	Missing	Missing	32	1,2,5,7,10,15, 16,18,19,22,28, 29	1. R1 2. R2 3. (R1+R2)*PY	1-4: 0 1-3: 0 1,2: 12	1-4: 0 1-3: 0 1,2: 9	1-4: 0 1-3: 0 1,2: 11

Table E.53 Constraints and Portion of the Predictive Mean Vector for Stimulants Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Stimulant Recency	Meth. Recency	Stimulant 12-Mo. Freq.	Meth. 12-Mo. Freq.				12-17	18-25	26+
25	Past year	(Past year not past month)	Missing	Missing	0	1-3,5,7,8,15-19, 25,27,29	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
26	Past year	Lifetime	Missing	Missing	0	1,2,5,7,10,15, 16,18,19,22,29	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases

Meth. = methamphetamine.

NOTE: Users of stimulants included users of methamphetamine.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month stimulants use | lifetime stimulants use)
2. R2 = P(Past year but not past month stimulants use | lifetime stimulants use)
3. PY = P(Stimulants use on a given day in the past year | past year stimulants use)

Table E.54 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Pain Reliever Recency	Oxy. Recency	Pain Reliever 12-Mo. Freq.	Oxy. 12-Mo. Freq.				12-17	18-25	26+
1	(Past month)		Missing		44	1-5,23	1. PY	1-3,5: 14 1-3: 3 1,2: 1	1-3,5: 4	1-3,5: 19 1-3: 0 1,2: 3
2	(Past year but not past month)		Missing		56	1-3,6,7	1. PY	1-3,5: 24 1-3: 1	1-3,5: 10	1-3,5: 20 1-3: 1
3	Past year				1	7,8	1. R1/(R1+R2)	No Cases	1-3,5: 1	No Cases
4	Past year		Missing		15	1-3,5,7,8,25	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 3	1-3,5: 2 1-3: 1 1,2: 2	1-3,5: 1 1-3: 2 1,2: 4
5	Missing		Missing		266	1-3,5,7,25	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 14 1-3: 2 1,2: 102	1-3,5: 7 1-3: 1 1,2: 64	1-3,5: 17 1-3: 3 1,2: 56
6	(Past month)	(Past month)		Missing	0	4,15-20,24	1. PY	No Cases	No Cases	No Cases
7	(Past year not missing)	(Past year not past month)		Missing	3	8,14-20	1. PY	No Cases	1-3,5: 1 1-3: 0 1,2: 1	1-3,5: 1
8	(Past year not missing)	Past year			0	7,8	1. R1/(R1+R2)	No Cases	No Cases	No Cases
9	(Past year not missing)	Past year	Missing		0	1-3,7,8,23	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
10	(Past year not missing)	Past year		Missing	5	7,8,15-17,26	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 3	1-3,5: 1	1-3,5: 1
11	(Past year not missing)	Past year	Missing	Missing	0	1-3,7,8,15-17, 23,28	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
12	(Past year not missing)	Missing		Missing	8	7,8,15-17	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 0 1-3: 0 1,2: 3	1-3,5: 0 1-3: 0 1,2: 3	1-3,5: 0 1-3: 0 1,2: 2

Table E.54 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency (continued)

#	Missingness Pattern: Pain Reliever Recency	Missingness Pattern: Oxy. Recency	Missingness Pattern: Pain Reliever 12-Mo. Freq.	Missingness Pattern: Oxy. 12-Mo. Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
13	(Past month)	(Past month)	Missing	Missing	2	1-4,7,15-17,20, 23,24	1. PY	1-3,5: 0 1-3: 0 1,2: 1	1-3,5: 1	No Cases
14	(Past month)	(Past year not past month)	Missing	Missing	0	1-4,7,15-17,21, 23	1. PY	No Cases	No Cases	No Cases
15	(Past year not past month)	(Past year not past month)	Missing	Missing	3	1-3,6,7,15-17,21	1. PY	1-3,5: 0 1-3: 0 1,2: 1	1-3,5: 1	1-3,5: 0 1-3: 0 1,2: 1
16	Past year	Past year			0	7,8,14,19,28	1. R1/(R1+R2)	No Cases	No Cases	No Cases
17	Past year	Past year	Missing		0	1-3,7,8,14,19, 25,28	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
18	Past year	Past year		Missing	3	7,8,14-17,19, 26,28	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 2	1-3,5: 0 1-3: 0 1,2: 1	No Cases
19	Past year	Past year	Missing	Missing	8	1-3,7,8,14-17, 19,25,26,28	1. R1/(R1+R2) 2. PY	1-3,5: 1 1-3: 0 1,2: 1	1-3,5: 1 1-3: 0 1,2: 1	1-3,5: 1 1-3: 0 1,2: 3
20	Past year	Missing		Missing	11	7,8,10,15-17, 19,28	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 2	1-3,5: 1 1-3: 0 1,2: 3	1-3,5: 0 1-3: 0 1,2: 5
21	Past year	Missing	Missing	Missing	1	1-3,7,8,10, 15-17,19,25,28	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 1	No Cases	No Cases
22	(Past month)	Missing		Missing	0	4,7,10,15-17,19	1. R1 2. R2 3. (R1+R2)*PY	No Cases	No Cases	No Cases
23	(Past month)	Missing	Missing	Missing	0	1-4,7,10,15-17, 19,23	1. R1 2. R2 3. (R1+R2)*PY	No Cases	No Cases	No Cases
24	Missing	Missing	Missing	Missing	37	1-3,7,10,15-17, 19,22,28	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 1 1-3: 0 1,2: 9	1-3,5: 0 1-3: 0 1,2: 11	1-3,5: 1 1-3: 0 1,2: 15

Table E.54 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency (continued)

#	Missingness Pattern: Pain Reliever Recency	Missingness Pattern: Oxy. Recency	Missingness Pattern: Pain Reliever 12-Mo. Freq.	Missingness Pattern: Oxy. 12-Mo. Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
25	Past year	(Past year not past month)	Missing	Missing	0	1-3,7,8,14-17, 19,25	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
26	Past year	Lifetime	Missing	Missing	3	1-3,6-8,10, 15-17,19,25	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 1	1-3,5: 0 1-3: 0 1,2: 2	No Cases

Oxy. = OxyContin.

NOTE: Users of pain relievers included users of OxyContin.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month pain relievers use | lifetime pain relievers use)
2. R2 = P(Past year but not past month pain relievers use | lifetime pain relievers use)
3. PY = P(Pain relievers use on a given day in the past year | past year pain relievers use)

Table E.55 Logical Constraints for Core-Plus-Noncore Stimulants Recency

Constraint #	Logical Constraint
LogC8	Donor must be a past month or past year but not past month stimulant user (stimulant recency = 1 or 2)
LogC10	Donor must be a past month, past year but not past month, or lifetime but not past year methamphetamine user (methamphetamine recency = 1, 2, or 3)
LogC14	Donor must be a past month or past year but not past month methamphetamine user (methamphetamine recency = 1 or 2)
LogC22	Donor must be a past month, past year but not past month, or lifetime but not past year stimulant user (stimulant recency = 1, 2, or 3)
LogC27	If recipient is not a lifetime user of any type of stimulant except for methamphetamine, then donor must not have used stimulants more recently than recipient has used methamphetamines
LogC28	If recipient is not a lifetime user of any type of stimulants except for methamphetamines, then donor must not have used stimulants more recently than donor has used methamphetamines

Table E.56 Likeness Constraints for Core-Plus-Noncore Stimulants Recency

Constraint #	Likeness Constraint
LikC1	Donor's stimulant recency must match recipient's stimulant recency (when nonmissing)[1]
LikC2	State rank of donor = state rank of recipient
LikC3	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC4	Donor's methamphetamine recency agrees with recipient's methamphetamine recency (when nonmissing)

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.57 Constraints and Portion of the Predictive Mean Vector for Core-Plus-Noncore Stimulants Recency

#	Missingness Pattern: Stimulant Recency	Missingness Pattern: Meth. Recency	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
1	Past year		13	8,27	1. R1/(R1+R2)	1-4: 0 1-3: 2	1-4: 5 1-3: 1	1-4: 5
2	Missing		48	22,27	1. R1 2. R2	1-4: 4 1-3: 2 1,2: 14	1-4: 15 1-3: 0 1,2: 2	1-4: 7 1-3: 1 1,2: 3
3	(Past month)	Past year	1	14	1. R1/(R1+R2)	No Cases	1-4: 1	No Cases
4	(Past year not missing)	Missing	2	10	1. R1 2. R2	1-4: 0 1-3: 0 1,2: 1	No Cases	1-4: 0 1-3: 0 1,2: 1
5	Past year	Past year	7	8,14,28	1. R1/(R1+R2)	1-4: 0 1-3: 0 1,2: 3	1-4: 1	1-4: 3
6	Past year	Missing	4	8,10,28	1. R1/(R1+R2)	1-4: 1 1-3: 0 1,2: 1	1-4: 2	No Cases
7	Missing	Missing	49	10,22,28	1. R1 2. R2	1-4: 1 1-3: 0 1,2: 12	1-4: 3 1-3: 0 1,2: 7	1-4: 11 1-3: 0 1,2: 15

Meth. = methamphetamine.

NOTE: Users of stimulants included users of methamphetamine.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month stimulants use | lifetime stimulants use)
2. R2 = P(Past year but not past month stimulants use | lifetime stimulants use)
3. PY = P(Stimulants use on a given day in the past year | past year stimulants use)

E.3.3 Age at First Use

Table E.58 Logical Constraints for Age at First Use, Univariate Assignment

Constraint #	Logical Constraint
LogC1	Donor's age at first drug use must be less than or equal to recipient's age
LogC2	If recipient was not a past year drug user, then donor cannot have age at first drug use equal to recipient's age
LogC3	If recipient was not a past 3 years drug user, then donor cannot have age at first drug use equal to recipient's age, recipient's age minus 1, or recipient's age minus 2
LogC4	If recipient's 12-month drug frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first drug use equal to recipient's age
LogC5	If recipient's 30-day drug frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first drug use equal to recipient's age
LogC6	If recipient's most recent birthday was within the past 30 days, and recipient's drug recency was past year but not past month, then donor cannot have age at first drug use equal to recipient's age
LogC7	If recipient's drug recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's 12-month drug frequency was less than 30, then donor cannot have age at first drug use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first drug use must be earlier than recipient's most recent birthday.)
LogC8	Donor's age at first cigarette use must be less than or equal to recipient's age at first daily cigarette use (if existing)
LogC9	Donor's age at first daily cigarette use must be greater than or equal to recipient's imputation-revised age at first cigarette use

Table E.59 Likeness Constraints for Age at First Use, Univariate Assignment

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	If recipient was a past year drug user, then donor was a past year drug user; if recipient was a past 3 years but not past year drug user, then donor was the same; if recipient was a lifetime but not past 3 years drug user, then donor was the same
LikC4	Age of donor = age of recipient
LikC5	Age of donor must be greater than or equal to age of recipient
LikC6	If recipient was a past year drug user, then donor was a past year drug user; if recipient was not a past year drug user, then donor was the same
LikC7	If recipient was a past 3 years but not past year drug user, then donor was not a past year drug user; if recipient was a lifetime but not past 3 years drug user, then donor was the same
LikC8	If recipient was not a past year drug user, then donor was the same

Table E.60 Constraints and Portion of the Predictive Mean Vector for Cigarette Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
1	Missing	374	1-3,5,6,8	1. PrAFU	1-4: 117 2-4: 1	1-4: 116	1-4: 133 2-4: 4 3,4: 3

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigarette age at first use

Table E.61 Constraints and Portion of the Predictive Mean Vector for Cigarette Age at First Daily Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
1	Missing	150	1-3,5,6,9	1. PrCigAFU	1-4: 6	1-4: 18 2-4: 2	1-4: 109 2-4: 9 3,4: 3 4,7: 1 5,7: 2

[1] The predictive mean vector components are defined by the following:
1. PrCigAFU = Predicted cigarette age at first daily use

Table E.62 Constraints and Portion of the Predictive Mean Vector for Cigar Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
1	Missing	443	1-3,5,6	1. PrAFU	1-4: 80 2-4: 2 3,4: 0 4,7: 0 5,7: 1	1-4: 109 2-4: 1	1-4: 219 2-4: 19 3,4: 11 4,7: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigar age at first use

Table E.63 Constraints and Portion of the Predictive Mean Vector for Alcohol Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
1	Missing	649	1,2,4,6,7	1. PrAFU	1,2,4,6: 186	1,2,4,6: 114 2,4,6: 0 4,6: 1	1,2,4,6: 329 2,4,6: 12 4,6: 7

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted alcohol age at first use

Table E.64 Constraints and Portion of the Predictive Mean Vector for Inhalants Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group 12-17	18-25	26+
1	Missing	250	1,2,4,6,7	1. PrAFU	1,2,4,6: 151 2,4,6: 0 4,6: 1	1,2,4,6: 29 2,4,6: 4	1,2,4,6: 53 2,4,6: 4 4,6: 7 4,8: 0 5,8: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted inhalants age at first use

Table E.65 Constraints and Portion of the Predictive Mean Vector for Marijuana Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group 12-17	18-25	26+
1	Missing	242	1,2,4,6,7	1. PrAFU	1,2,4,6: 99	1,2,4,6: 51	1,2,4,6: 84 2,4,6: 6 4,6: 2

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted marijuana age at first use

Table E.66 Constraints and Portion of the Predictive Mean Vector for Tranquilizers Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group 12-17	18-25	26+
1	Missing	193	1,2,4,6,7	1. PrAFU	1,2,4,6: 38 2,4,6: 4 4,6: 3	1,2,4,6: 48	1,2,4,6: 70 2,4,6: 17 4,6: 10 4,8: 2 5,8: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted tranquilizers age at first use

Table E.67 Constraints and Portion of the Predictive Mean Vector for Sedatives Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group 12-17	18-25	26+
1	Missing	71	1,2,4,6,7	1. PrAFU	1,2,4,6: 7 2,4,6: 8 4,6: 6	1,2,4,6: 13 2,4,6: 3	1,2,4,6: 14 2,4,6: 10 4,6: 9 4,8: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted sedatives age at first use

Table E.68 Constraints and Portion of the Predictive Mean Vector for Heroin Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	13	1,2,4,6,7	1. PrAFU	1,2,4,6: 2 2,4,6: 0 4,6: 0 4,8: 0 5,8: 2	1,2,4,6: 2	1,2,4,6: 3 2,4,6: 2 4,6: 2

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted heroin age at first use

Table E.69 Logical Constraints for Age at First Use, Multivariate Assignment

Constraint #	Logical Constraint
LogC1	For each parent or child drug for which recipient was missing age at first use, donor's age at first use must be less than or equal to recipient's age
LogC2	For each parent or child drug for which recipient was missing age at first use, if recipient was not a past year user, then donor cannot have age at first use equal to recipient's age
LogC3	For each child drug for which recipient was missing age at first use, if recipient was not a past 3 years user, then donor cannot have age at first use equal to recipient's age, recipient's age minus 1, or recipient's age minus 2
LogC4	If recipient was missing the age at first use for the parent drug, and if recipient's parent drug 12-month frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have parent drug age at first use equal to recipient's age
LogC5	For each child drug for which recipient was missing age at first use, if recipient's 12-month frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first use equal to recipient's age
LogC6	For each child drug for which recipient was missing age at first use, if recipient's 30-day frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first use equal to recipient's age
LogC7	For each parent or child drug for which recipient was missing age at first use, if recipient's most recent birthday was within the past 30 days, and recipient's recency was past year but not past month, then donor cannot have age at first use equal to recipient's age
LogC8	If recipient was missing the age at first use for the parent drug, and recipient's parent drug recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's parent drug 12-month frequency was less than 30, then donor cannot have parent drug age at first use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first use must be earlier than recipient's most recent birthday.)
LogC9	For each child drug for which recipient was missing age at first use, if recipient's recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's 12-month frequency was less than 30, then donor cannot have age at first use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first use must be earlier than recipient's most recent birthday.)

Table E.69 Logical Constraints for Age at First Use, Multivariate Assignment (continued)

Constraint #	Logical Constraint
LogC10	For each child drug for which recipient was missing age at first use, if recipient's recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and the parent drug 12-month frequency was less than 30, then donor cannot have age at first use equal to recipient's age. (This constraint technically eliminates some potential donors who should be eligible, but it would be difficult to write a similar constraint that eliminates exactly the right donors from the neighborhood. For example, consider an 18-year-old respondent with missing LSDAGE whose birthday was 40 days ago, has IRLSDRC = 2, and has IRHALFY = 15. There is no question in the questionnaire for LSD 12-month frequency, so unless respondent did not use any hallucinogens other than LSD, the LSD 12-month frequency is unknown. The "conservative" constraint is to eliminate anyone from the neighborhood with IRLSDAGE = 18.)
LogC11	For each child drug for which recipient was missing age at first use, donor must have age at first use greater than or equal to recipient's parent drug age at first use (if existing)
LogC12	If recipient was missing the age at first use for the parent drug, donor must have parent drug age at first use less than or equal to the recipient's child drug(s) age(s) at first use (if existing)
LogC13	For each child drug for which recipient was missing age at first use, donor must be a lifetime user
LogC14	If recipient was not a lifetime user of "other" pain relievers/stimulants, then donor must have age at first use for the parent drug equal to age at first use for the child drug
LogC15	If recipient was not a lifetime user of "other" hallucinogens, then donor must not have values that would cause the minimum of the child drug(s) age(s) at first use to be greater than the overall hallucinogens age at first use for recipient. For example, if donor was not a lifetime user of "other" hallucinogens and has HALLAGE = 17, LSDAGE = missing, PCPAGE = 19, and ECSAGE = missing, then donor must have LSDAGE = 17 and/or ECSAGE = 17.

Table E.70 Likeness Constraints for Age at First Use, Multivariate Assignment

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Age of donor = age of recipient
LikC4	Age of donor must be greater than or equal to age of recipient
LikC5	For the parent drug, if recipient was a past year user, then donor was a past year user
LikC6	For the parent drug, if recipient was a lifetime but not past year user, then donor was a lifetime but not past year user
LikC7	For each child drug, if recipient was a past year user, then donor was a past year user
LikC8	For each child drug, if recipient was a lifetime but not past year user, then donor was not a past year user
LikC9	For each child drug for which recipient was missing age at first use, if recipient was a past year user, then donor was a past year user; if recipient was a past 3 years but not past year user, then donor was the same; if recipient was a lifetime but not past 3 years user, then donor was the same

Table E.70 Likeness Constraints for Age at First Use, Multivariate Assignment (continued)

Constraint #	Likeness Constraint
LikC10	For each child drug for which recipient was missing age at first use, if recipient was a past 3 years but not past year child drug user, then donor was not a past year child drug user; if recipient was a lifetime but not past 3 years child drug user, then donor was the same
LikC11	Donor agrees with recipient with respect to lifetime use for child drug(s)
LikC12	Donor was at least as old as recipient, but no more than 20 years older than recipient
LikC13	Donor was no more than 20 years older than recipient

Table E.71 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	265	1-3,6,7,13	1. PrSltAFU	1-3,9: 64 2-3,9: 7 3,9: 4	1-3,9: 90 2-3,9: 1	1-3,9: 77 2-3,9: 12 3,9: 9 3,10: 0 4,10: 1

[1] The predictive mean vector components are defined by the following:
1. PrSltAFU = Predicted smokeless tobacco age at first use

Table E.72 Constraints and Portion of the Predictive Mean Vector for Cocaine Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	66	1,2,4,5,7-9,11,12,13	1. PrCocAFU	1,2,3,5-8,11: 4 2,3,5-8,11: 0 2,3,5-8: 0 3,5-8: 1 3,6,8: 0 6,8,12: 1	1,2,3,5-8,11: 19 2,3,5-8,11: 2 2,3,5-8: 0 3,5-8: 2	1,2,3,5-8,11: 33 2,3,5-8,11: 3 2,3,5-8: 0 3,5-8: 1

[1] The predictive mean vector components are defined by the following:
1. PrCocAFU = Predicted cocaine age at first use

Table E.73 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	213	1,2,4,7,8, 10-13,15	1. PrHalAFU	1-3,5-8,11: 21 2,3,5-8,11: 3 2,3,5-8: 2 3,5-8: 3 3,6,8: 1 6,8,12: 1	1-3,5-8,11: 43 2,3,5-8,11: 7 2,3,5-8: 3 3,5-8: 1	1-3,5-8,11: 67 2,3,5-8,11: 20 2,3,5-8: 9 3,5-8: 23 3,6,8: 1 6,8,12: 7 6,8,13: 0 13: 0 13*: 1

*A donor could not be found for these cases. Imputed values were randomly assigned within the bounds created during the editing process.

[1] The predictive mean vector components are defined by the following:
1. PrHalAFU = Predicted hallucinogens age at first use

Table E.74 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	629	1,2,4,5, 7-9,11-14	1. PrAnlAFU	1-3,5-8,11: 178 2,3,5-8,11: 5 2,3,5-8: 0 3,5-8: 1	1-3,5-8,11: 138 2,3,5-8,11: 2 2,3,5-8: 0 3,5-8: 1	1-3,5-8,11: 221 2,3,5-8,11: 36 2,3,5-8: 4 3,5-8: 31 3,6,8: 3 6,8,12: 8 6,8,13: 1

[1] The predictive mean vector components are defined by the following:
1. PrAnlAFU = Predicted pain relievers age at first use

Table E.75 Constraints and Portion of the Predictive Mean Vector for Stimulants Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	134	1,2,4,5, 7-9,11-14	1. PrStmAFU	1-3,5-8,11: 19 2,3,5-8,11: 5 2,3,5-8: 0 3,5-8: 10 3,6,8: 0 6,8,12: 1	1-3,5-8,11: 32 2,3,5-8,11: 2 2,3,5-8: 0 3,5-8: 1	1-3,5-8,11: 42 2,3,5-8,11: 12 2,3,5-8: 2 3,5-8: 6 3,6,8: 1 6,8,12: 1

[1] The predictive mean vector components are defined by the following:
1. PrStmAFU = Predicted stimulants age at first use

E.4 Household Composition (Roster) Variables

Tables E.76 through E.97 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the eight household composition (roster) variables: number of rostered individuals, number of children younger than 18, number of individuals aged 65 or older, indicator of whether the respondent has family members in household, number of respondent's family members in the household (both including and excluding foster relationships), and number of respondent's family members younger than 18 in the household (both including and excluding foster relationships).

Table E.76 Likeness Constraints for Number of Rostered People

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means

Table E.77 Constraints and Portion of the Predictive Mean Vector for Number of Rostered People

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	71	None	1. C1	1: 11	1: 15	1: 34	1: 11

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.78 Logical Constraints for Number of Children Younger than 18

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE and nonmissing ages in the roster

Table E.79 Likeness Constraints for Number of Children Younger than 18

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.80 Constraints and Portion of the Predictive Mean Vector for Number of Children Younger than 18

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	266	1	1. C1	1,2: 64 2: 1	1,2: 81	1,2: 106 2: 1	1,2: 13

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.81 Logical Constraints for Number of People Aged 65 or Older

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE, IRKID17, and nonmissing ages in the roster

Table E.82 Likeness Constraints for Number of People Aged 65 or Older

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.83 Constraints and Portion of the Predictive Mean Vector for Number of People Aged 65 or Older

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	351	1	1. C1	1,2: 140 1: 1	1,2: 106 1: 1	1,2: 94 1: 1	1,2: 8

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.84 Likeness Constraints for Indicator of Whether the Respondent Has Family Members in Household

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	Donor's predicted means must be within 20 percent of recipient's predicted means
LikC3	IRKID17 of donor = IRKID17 of recipient
LikC4	If recipient was married, then donor was married; otherwise, if recipient was not currently married, then donor was not currently married

Table E.85 Constraints and Portion of the Predictive Mean Vector for Indicator of Whether the Respondent Has Family Members in Household

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	83	None	1. F1	1-4: 11	1-4: 21	1-4: 45 2-4: 1	1-4: 5

[1] The predictive mean vector components are defined by the following:
1. F1 = P(No other family members in the household)

Table E.86 Logical Constraints for Number of Respondent's Family Members in Household (Including Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE and nonmissing ages in the roster

Table E.87 Likeness Constraints for Number of Respondent's Family Members in Household (Including Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.88 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household (Including Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	107	1	1. C1	1-3: 23	1-3: 25	1-3: 53	1-3: 6

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.89 Logical Constraints for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRFMLYSZ and nonmissing ages in the roster

Table E.90 Likeness Constraints for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient
LikC4	IRFMLYSZ of donor = IRFMLYSZ of recipient

Table E.91 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	126	1	1. C1	1-4: 40 2-4: 0 3,4: 1	1-4: 26 2-4: 0 3,4: 3	1-4: 52 2-4: 0 3,4: 0 4: 1	1-4: 3

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.92 Logical Constraints for Number of Respondent's Family Members in Household (Excluding Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRFMLYSZ and nonmissing ages in the roster

Table E.93 Likeness Constraints for Number of Respondent's Family Members in Household (Excluding Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient
LikC4	IRFMLYSZ of donor = IRFMLYSZ of recipient

Table E.94 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household (Excluding Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	101	1	1. C1	1-4: 22	1-4: 22	1-4: 51	1-4: 6

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.95 Logical Constraints for Number of Respondent's Family Members in Household Younger than 18 (Excluding Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRFAMSZE and nonmissing ages in the roster

Table E.96 Likeness Constraints for Number of Respondent's Family Members in Household Younger than 18 (Excluding Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient
LikC4	IRFMLYSZ of donor = IRFMLYSZ of recipient
LikC5	IRFAMSZE of donor = IRFAMSZE of recipient
LikC6	IRKDFMLY of donor = IRKDFMLY of recipient

Table E.97 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household Younger than 18 (Excluding Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	113	1	1. C1	1-6: 35 2-6: 0 3-6: 0 4-6: 0 5,6: 1	1-6: 23 2-6: 0 3-6: 1	1-6: 49 2-6: 1	1-6: 3

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

E.5 Income Variables

Tables E.98 through E.103 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the income variables.

E.5.1 Binary Variable Phase

Five of the binary income variables were directly related to a respondent's socioeconomic status. Hence, if a recipient required imputation for one or more of these five variables (i.e., welfare payments, welfare services, food stamps, binary income, and months on welfare), but had information on at least one of these variables, the donors were restricted so that donors and recipients had the same values for these nonmissing variables. These five variables are referred to as "welfare-correlated variables."

There were a large number of missingness patterns for the source-of-income variables because they are imputed simultaneously in a set. A respondent could be missing any combination of the seven source-of-income variables. Because the constraints and predictive mean vectors can be described easily, only one row in Table E.100 is used to summarize the many multiple variable missingness patterns.

Table E.98 Logical Constraints for Source of Income

Constraint #	Logical Constraint
LogC1	IRFAMSKP of donor = IRFAMSKP of recipient[1]
LogC2	Recipient is missing months on welfare. If recipient is also known to have received either welfare payments or welfare services, then donor must also have received welfare payments or welfare services. (This prevents donor from giving a skip code to a recipient who should have a value from 1 to 12.)

[1] This is only a logical constraint when family binary total income is missing.

Table E.99 Likeness Constraints for Source of Income

Constraint #	Likeness Constraint
LikC1	Age of donor = age of recipient
LikC2	If any of the welfare-correlated edited binary income variables (welfare payments, welfare services, food stamps, binary total income, and months on welfare) were missing, then donor must match recipient on all nonmissing welfare-correlated edited binary income variables
LikC3	Donor's predicted means must be within 5 percent of recipient's predicted means for all missing family variables
LikC4	If recipient is missing months on welfare, then donor must match recipient with respect to personal welfare payments (if nonmissing) and welfare services (if nonmissing)
LikC5	If recipient is missing social security, then donor must match recipient with respect to whether there are adults aged 65 or older in the household
LikC6	If recipient is missing welfare payments and/or welfare services, then donor must match recipient with respect to whether there are children younger than 18 in the household
LikC7	If recipient is missing wages, then donor must match recipient with respect to whether there are adults aged 18 to 64 in the household
LikC8	Age of donor must be within 5 years of age of recipient
LikC9	If recipient is missing binary income at the family level but not at the personal level, then donor must match recipient with respect to binary income at the personal level
LikC10	If recipient is not missing binary income at the personal level, then donor must match recipient with respect to binary income at the personal level

Table E.100 Constraints and Portion of the Predictive Mean Vector for Source of Income

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Missing welfare months, receiving family payment and/or family service	136	1,2	1. WMS, and probabilities associated with other missing elements	1-7: 13 1,2,4-7: 30	1-7: 8 1,2,4-7: 38 2,4-8: 5	1-7: 5 1,2,4-7: 22 2,4-8: 11	1-7: 1 1,2,4-7: 1 2,4-8: 0 2,4-7: 2
2	Missing welfare months, not receiving welfare payments, missing welfare services	91	1,2	1. SVC*WMS, SVC, and probabilities associated with other missing elements	1-7: 13 1,2,4-7: 22	1-7: 5 1,2,4-7: 24	1-7: 3 1,2,4-7: 15	1-7: 2 1,2,4-7: 6 2,4-8: 1
3	Missing welfare months, missing welfare payments, not receiving welfare services	190	1,2	1. PMT*WMS, PMT, and probabilities associated with other missing elements	1-7: 30 1,2,4-7: 28 2,4-8: 1	1-7: 6 1,2,4-7: 61 2,4-8: 1	1-7: 9 1,2,4-7: 43	1-7: 3 1,2,4-7: 8
4	Missing welfare months, missing welfare payments, missing welfare services	237	1,2	1. [1-(1-PMT)(1-SVC)]*WMS, PMT, SVC, and probabilities associated with other missing elements	1-7: 16 1,2,4-7: 85	1-7: 2 1,2,4-7: 60	1-7: 2 1,2,4-7: 61	1-7: 1 1,2,4-7: 10

Table E.100 Constraints and Portion of the Predictive Mean Vector for Source of Income (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
5	All other missingness patterns with number of months on welfare nonmissing[2]	3,357	1	1. Probabilities associated with missing elements	1-7: 681 1,2,5-7: 62 2,5-8: 0 2,5-7: 0 5-7,9: 1	1-7: 1,028 1,2,5-7: 159	1-7: 820 1,2,5-7: 256 2,5-8: 9 2,5-7: 1	1-7: 212 1,2,5-7: 127 2,5-8: 1

[1] The predictive mean vector components are defined by the following:
1. PMT = P(Family in household received income from welfare payments)
2. SVC = P(Family in household received income from other welfare services)
3. WMS = P(Family in household received any welfare on a given month in the past year | family received any welfare in the past year)

[2] There were many other missingness patterns for source of income because a respondent could be missing any combination of the 10 source-of-income variables. Because the constraints and predictive mean vectors can be described easily, all of these other missingness patterns are represented by a single row in the table.

E.5.2 Specific Category Phase

Table E.101 Logical Constraints for Finer Categories of Income

Constraint #	Logical Constraint
LogC1	IRFAMSKP of donor = IRFAMSKP of recipient
LogC2	Family income (FINC2) of donor must be greater than personal income (PINC2) of recipient
LogC3	Personal income (PINC2) of donor must be less than family income (FINC2) of recipient

Table E.102 Likeness Constraints for Finer Categories of Income

Constraint #	Likeness Constraint
LikC1	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC2	PINC2 of donor = PINC2 of recipient
LikC3	FINC2 of donor = FINC2 of recipient
LikC4	IRPINC1 of donor = IRPINC1 of recipient
LikC5	IRFINC1 of donor = IRFINC1 of recipient

Table E.103 Constraints and Portion of the Predictive Mean Vector for Finer Categories of Income

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Missing personal income only	724	13	1. Xβ	1, 3-5: 77	1, 3-5: 166	1, 3-5: 286 3-5: 1	1, 3-5: 194
2	Missing family income only	4,639	1,2	1. Xβ	1,2,4,5: 1,213 2,4,5: 3	1,2,4,5: 1,757 2,4,5: 1	1,2,4,5: 1,429 2,4,5: 1	1,2,4,5: 230 2,4,5: 5
3	Completely missing	1,521	1	1. Xβ	1,4,5: 193	1,4,5: 262	1,4,5: 776	1,4,5: 290

[1] For more details, see Section 8.3.4.

E.6 Health Insurance Variables

Tables E.104 through E.107 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the health insurance variables. Prior to 2013 two methods were presented for health insurance variables, the "constituent variables" method and the "old" method. See Chapter 10 for an explanation of the "constituent variables" and why the "old" method is no longer presented.

In several instances, variable names are used without description for the purposes of brevity (see Chapter 10 for details). For the health insurance imputations, matches between donors and recipients were attempted on the nonmissing values of the variables CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. These variables are the edited indicators of whether the respondent received health insurance from Medicaid/state health insurance programs for children, Medicare, CHAMPUS, or private health insurance, respectively. These were the base variables used in the creation of the imputation-revised variables (IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT). In addition to the edited health insurance variables, other variables, which were used as likeness constraints, are identified in the tables only by their variable names. These include SERVICE (an indicator of whether the respondent had ever been in the military), GOVTPROG (an indicator of whether the respondent's family participated in government public assistance programs), INCOME (a four-level categorical family-income variable, with levels of less than $20,000, $20,000 to less than $50,000, $50,000 to less than $75,000, and $75,000 or more), IRFAMIN1 (a two-level family income variable, with levels of less than $20,000 and $20,000 or more), and IRFAMSOC (an indicator of whether the respondent's family in the household received income from social security).

For the multivariate predictive mean neighborhood (MPMN) method, the likeness constraints, which were applied to the variables, differed between missingness patterns, and sometimes the constraints differed between age groups within the same missingness pattern.

Table E.104 Likeness Constraints for Health Insurance (MPMN)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means for all missing insurance variables (CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN)
LikC2	GOVTPROG of donor = GOVTPROG of recipient
LikC3	IRFAMSOC of donor = IRFAMSOC of recipient
LikC4	SERVICE of donor = SERVICE of recipient (if nonmissing)
LikC5	INCOME of donor = INCOME of recipient
LikC6	IRFAMIN1 of donor = IRFAMIN1 of recipient
LikC7	Donor must match recipient for all nonmissing insurance variables
LikC8	If recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then (1) if recipient has no job due to disability (JBSTATR = 14), then donor must have no job due to disability (2) if recipient has JBSTATR not equal to 14, then donor must have JBSTATR not equal to 14

MPMN = multivariate predictive mean neighborhood.

Table E.105 Constraints and Portion of the Predictive Mean Vector for Health Insurance (MPMN)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Only missing CAIDCHIP	300	None	1. HI1	1,2,7: 96	1,2,7: 121 1,2: 1	1,2,7: 62 1,2: 2	1,2,7: 16 1,2: 2
2	Only missing MEDICARE	83	None	1. HI2	1,3,7: 23 1,3: 1	1,3,7,8: 35	1,3,7,8: 17	1,3,7: 6 1: 1
3	Missing CAIDCHIP and MEDICARE	44	None	1. HI1 2. HI2	1-3,7: 7 1-3: 2 1: 2 None: 1	1,2,7,8: 11 1,2,8: 2 1,8: 1 8: 3	1,2,7,8: 8 1,2,8: 0 1,8: 2 8: 2	1,2,7: 1 1,2: 0 1: 0 None: 2
4	Only missing CHAMPUS	77	None	1. HI3	1,4,7: 23	1,4,7: 35 1,4: 1	1,4,7: 14	1,4,7: 3 1,4: 1
5	Missing CAIDCHIP and CHAMPUS	22	None	1. HI1 2. HI3	1,2,4,7: 7 1,2,4: 2 1,4: 0 4: 1	1,2,4,7: 10 1,2,4: 1	1,2,4,7: 1	No Cases
6	Missing MEDICARE and CHAMPUS	8	None	1. HI2 2. HI3	1,3,4,7: 3	1,4,7,8: 0 1,4,8: 2	1,4,7,8: 2	1,4,7: 0 1,4: 1
7	Missing CAIDCHIP, MEDICARE, and CHAMPUS	13	None	1. HI1 2. HI2 3. HI3	1-4,7: 3 1-4: 1 1,4: 0 4: 2	1,2,4,7,8: 3 1,2,4,8: 0 1,4,8: 0 4,8: 2	1,2,4,7,8: 0 1,2,4,8: 0 1,4,8: 0 4,8: 2	No Cases
8	Only missing PRVHLTIN	240	None	1. HI4	1,5,7: 75 1,5: 0 1,6: 1	1,5,7: 107	1,5,7: 40 1,5: 0 1,6: 0 1: 0 None: 1	1,6,7: 16
9	Missing CAIDCHIP and PRVHLTIN	52	None	1. HI1 2. HI4	1,2,5,7: 24 1,2,5: 0 1,6: 1 1: 0 None: 2	1,2,5,7: 17 1,2,5: 0 1,6: 1 1: 1 None: 1	1,2,5,7: 3 1,2,5: 1	1,2,6,7: 0 1,2,6: 0 1: 1
10	Missing MEDICARE and PRVHLTIN	12	None	1. HI2 2. HI4	1,3,5,7: 5 1,3,5: 0 1,6: 0 1: 2	1,5,7,8: 1 1,5,8: 1 1,6,8: 0 1,8: 2 8: 1	No Cases	No Cases

Table E.105 Constraints and Portion of the Predictive Mean Vector for Health Insurance (MPMN) (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
11	Missing CAIDCHIP, MEDICARE, and PRVHLTIN	19	None	1. HI1 2. HI2 3. HI4	1-3,5,7: 3 1-3,5: 0 1,6: 0 1: 0 None: 6	1,2,5,7,8: 0 1,2,5,8: 0 1,6,8: 1 1,8: 2 8: 5	1,2,5,7,8: 1 1,2,5,8: 0 1,6,8: 0 1,8: 0 8: 1	No Cases
12	Missing CHAMPUS and PRVHLTIN	16	None	1. HI3 2. HI4	1,4,5,7: 3	1,4,5,7: 5 1,4,5: 2 1,4,6: 1 1,4: 2	1,4,5,7: 2 1,4,5: 0 1,4,6: 0 1,4: 0 4: 1	No Cases
13	Missing CAIDCHIP, CHAMPUS, and PRVHLTIN	30	None	1. HI1 2. HI3 3. HI4	1,2,4,5,7: 8 1,2,4,5: 0 1,4,6: 1 1,4: 0 4: 7	1,2,4,5,7: 5 1,2,4,5: 0 1,4,6: 0 1,4: 2 4: 4	1,2,4,5,7: 1 1,2,4,5: 0 1,4,6: 0 1,4: 0 4: 2	No Cases
14	Missing MEDICARE, CHAMPUS, and PRVHLTIN	4	None	1. HI2 2. HI3 3. HI4	1,3-5,7: 1 1,3-5: 0 1,4,6: 0 1,4: 0 4: 2	1,4,5,7,8: 1	No Cases	No Cases
15	Missing CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN	92	None	1. HI1 2. HI2 3. HI3 4. HI4	1-5: 18 1,4,6: 0 1,4: 0 4: 10	1,2,4,5,8: 6 1,4,6,8: 0 1,4,8: 2 4,8: 12	1,2,4,5,8: 4 1,4,6,8: 1 1,4,8: 0 4,8: 31	1,2,4,6: 2 1,4: 0 4: 6

MPMN = multivariate predictive mean neighborhood.

[1] The predictive mean vector components are defined by the following:
1. HI1 = P(Respondent has any other health insurance)
2. HI2 = P(Respondent has Medicare)
3. HI3 = P(Respondent has CHAMPUS)
4. HI4 = P(Respondent has private health insurance)

Table E.106 Constraints for Health Insurance (UPMN), Any Other Health Insurance

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient is between 26 and 64 years old, donor must also be between 26 and 64 years old; if recipient is aged 65 years or older, donor must also be aged 65 years or older

UPMN = univariate predictive mean neighborhood.

Table E.107 Constraints and Portion of the Predictive Mean Vector for Health Insurance (UPMN), Any Other Health Insurance

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26 or Older
1	Missing ANYOTHER	200	None	1. HI1	1: 27	1: 122	1,2: 50 2: 1

UPMN = univariate predictive mean neighborhood.

[1] The predictive mean vector components are defined by the following:
1. HI1 = P(Respondent has any other health insurance)

E.7 Roster Pair Variables

Tables E.108 through E.127 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the pair variables. Tables E.108 through E.111 correspond to the first stage of imputation, where the pair relationship variable (IRPRREL) is created. Tables E.112 through E.119 are for the second stage, imputation of the multiplicity variables. Finally, Tables E.120 through E.127 correspond to the third stage, where the household counts are imputed.

There are a few instances where variable names are used without description for the purposes of brevity. In these cases, the variables are defined in a table at the beginning of the section in which they are used.

E.7.1 Stage One: Pair Relationship

Table E.108 Variables Used in Constraints for Stage One, Pair Relationship

Variable	Value	Constraint
MARIT1	1	At least one pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned
	0	Neither pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned
MARIT2	2	Both pair members have a marital status of "Married" that was not imputed or logically assigned
	1	One pair member has a marital status of "Married" that was not imputed or logically assigned
	0	Neither pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned

Table E.108 Variables Used in Constraints for Stage One, Pair Relationship (continued)

Variable	Value	Constraint
MARIT3	5	Both pair members have a marital status of "Married" that was not imputed or logically assigned
	4	One pair member has a marital status of "Married," and the other has a marital status of "Widowed" or "Divorced or Separated," neither of which was imputed or logically assigned
	3	One pair member has a marital status of "Married," and the other either has a marital status of "Never Married" or was younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned
	2	Both pair members have a marital status of "Widowed" or "Divorced or Separated" that was not imputed or logically assigned
	1	One pair member has a marital status of "Widowed" or "Divorced or Separated," and the other either has a marital status of "Never Married" or was younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned
	0	Both pair members either have a marital status of "Never Married" or were younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned

Table E.109 Logical Constraints for Stage One, Pair Relationship

Constraint #	Logical Constraint
LogC10	If recipient is a spouse-spouse pair, with or without children, then donor must be a spouse-spouse pair either with or without children
LogC15	If recipient could be either a parent-child pair where the child is aged 12 to 14 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 12 to 14 or some other clearly identifiable pair that is not of interest
LogC16	If recipient could be either a parent-child pair where the child is aged 15 to 17 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 15 to 17 or some other clearly identifiable pair that is not of interest
LogC17	If recipient could be either a parent-child pair where the child is aged 18 to 20 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 18 to 20 or some other clearly identifiable pair that is not of interest
LogC18	If recipient could be either a parent-child pair where the child is aged 21 or older or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 21 or older or some other clearly identifiable pair that is not of interest
LogC20	If recipient could be either a sibling-sibling pair where one sibling is aged 12 to 17 and the other is aged 18 to 25 or some other clearly identifiable pair that is not of interest, then donor must be either a sibling-sibling (12-17/18-25) pair or some other clearly identifiable pair that is not of interest
LogC21	If recipient could be either a sibling-sibling pair where the siblings are not in either age range of interest or some other clearly identifiable pair that is not of interest, then donor must be a sibling-sibling pair where the siblings are not in an age range of interest or some other clearly identifiable pair that is not of interest
LogC22	If recipient could be either a spouse-spouse pair with children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair with children or some other clearly identifiable pair that is not of interest

Table E.109 Logical Constraints for Stage One, Pair Relationship (continued)

Constraint #	Logical Constraint
LogC23	If recipient could be either a spouse-spouse pair without children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair without children or some other clearly identifiable pair that is not of interest
LogC24	If recipient could be a spouse-spouse pair with or without children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair with or without children or some other clearly identifiable pair that is not of interest
LogC25	If recipient could be either a grandparent-grandchild pair or another clearly identifiable pair that is not of interest, then donor must be a grandparent-grandchild pair or some other clearly identifiable pair that is not of interest

Table E.110 Likeness Constraints for Stage One, Pair Relationship

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient is not a possible parent-child pair, and recipient has at least one parent in the household, then donor must also have at least one parent in the household
LikC3	If recipient is not a possible parent-child pair, and recipient has no parents in the household, then donor must also not have any parents in the household
LikC4	If the number of children aged 0 to 18[1] is nonmissing for both recipient and donor, then the two values must be equal
LikC5	If the recipient pair members are of the same gender, then the donor pair members must be of the same gender. If the recipient pair members are of a different gender, then the donor pair members must also be of a different gender.
LikC6	The age of the donor's younger pair member must be the same age as the recipient's younger pair member
LikC61	If donor is a spouse-spouse pair, with or without children, then donor's younger pair member must be the same age as recipient's younger pair member
LikC62	If both donor and recipient have pair members of different ages, then donor's younger pair member must be the same age as recipient's younger pair member
LikC7	The younger pair member of both donor and recipient must fall within the same age group: 21-25, 26-34, 35-49, or 50 or older
LikC8	The older pair member of both donor and recipient must fall within the same age group: 26-34, 35-49, or 50 or older
LikC9	If recipient pair members are in the same age group (21-25, 26-34, 35-49, or 50 or older), then donor cannot be a parent-child pair where the child is aged 21 or older
LikC10	Neither of the donor's pair members can have had an imputed marital status
LikC101	If donor is a sibling-sibling (12-14/15-17) pair, then neither of donor's pair members can have had an imputed marital status
LikC11	If neither of the recipient's pair members' marital status was imputed or if recipient has MARIT1 = 1, then donor must have the same value for MARIT1 as recipient

Table E.110 Likeness Constraints for Stage One, Pair Relationship (continued)

Constraint #	Likeness Constraint
LikC12	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value for MARIT2 as recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT2 = 1, then donor must have had MARIT2 = 1 or 2.
LikC121	If donor is a sibling-sibling (12-14/15-17) pair, then apply Likeness Constraint 12
LikC13	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value for MARIT3 as recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT2 = 1, then donor must have had MARIT2 = 1 or 2.
LikC14	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value of MARIT3 as the recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT1 = 1, then donor must also have had MARIT1 = 1.

[1] For age group pairs 3 and 4, this constraint is on the number of children aged 0 to 11.

Table E.111 Constraints and Portion of the Predictive Mean Vector for Stage One, Pair Relationship

#	Age Group Pair[1]	Total Number of Cases	Logical Constraints	Predictive Mean Vector[2]	Likeness Constraints: Number of Cases
1	One pair member is aged 12 to 14, the other is aged 15 to 17	7	None	1. SIB	1-3: 7
2	One pair member is aged 12 to 14, the other is aged 18 to 25	6	15,20	1. SIB	1-3,10,11: 4 2,3,10,11: 2
3	Both pair members are aged 15 to 17	6	10,16,21-25	1. SIB	1-5,62,101,121: 3 2-5,62,10,12: 2 2-5,62: 0 2-5: 1
4	One pair member is aged 15 to 17, the other is aged 18 to 25	8	10,16,20-25	1. SIB	1-6,10,13: 4 2-6,10,13: 4
5	One pair member is aged 18 to 20, the other is aged 18 to 25	23	10,17,21-24	1. SPOUSE1 2. SPOUSE2	1-5,10,13: 8 2-5,10,13: 14 2-5,10,12: 0 2-5: 1
6	Both pair members are aged 21 to 25	33	10,18,21-24	1. SPOUSE1 2. SPOUSE2	1-5,10,12: 7 2-5,10,12: 25 2-5: 1
7	One pair member is aged 12 to 14, the other is aged 26 or older	10	15,21,25	1. PC	1-3,8,10,11: 8 2,3,8,10,11: 2
8	One pair member is aged 15 to 17, the other is aged 26 or older	9	16,21-25	1. PC	1-5,61,8,10,14: 3 2-5,61,8,10,14: 6
9	One pair member is aged 18 to 20, the other is aged 26 or older	14	10,17,21-25	1. PC	1-5,8,10,14: 5 2-5,8,10,14: 8 2-5,8,10,11: 1
10	One pair member is aged 21 or older, the other is aged 26 or older	87	10,18,21-25	1. SPOUSE1 2. SPOUSE2	1-5,7-10,12: 14 2-5,7-10,12: 73
11	Both pair members are aged 12 to 17	7	None	1. SIB	1-3: 7

[1] Because there was only one variable to be imputed, PAIRREL, there was only one missingness pattern. However, as the predictive mean vector and constraints vary by age group, the age groups are presented as missingness patterns in this table, and the column heading has been changed accordingly.

[2] The predictive mean vector components are defined by the following:
1. SIB = P(the pair relationship is sibling-sibling)
2. SPOUSE1 = P(the pair relationship is spouse-spouse, with children)
3. SPOUSE2 = P(the pair relationship is spouse-spouse, without children)
4. PC = P(the pair relationship is parent-child)

E.7.2 Stage Two: Multiplicity Counts

Table E.112 Logical Constraints for Stage Two, Multiplicity Counts

Constraint #	Logical Constraint
LogC2	If recipient is a parent-child pair and if recipient's household count(s) in the appropriate age group(s) (12-14, 15-17, and/or 12-20) is nonmissing, then donor's and recipient's count(s) must match. Also, donor's parent-focus multiplicity count(s) must fall within recipient's bounds.
LogC3	Donor's sibling-sibling (12-14/15-17) older focus multiplicity count must fall within recipient's bounds
LogC4	Donor's sibling-sibling (12-14/15-17) younger focus multiplicity count must fall within recipient's bounds
LogC5	Donor's sibling-sibling (12-17/18-25) older focus multiplicity count must fall within recipient's bounds
LogC6	Donor's sibling-sibling (12-17/18-25) younger focus multiplicity count must fall within recipient's bounds

Table E.113 Likeness Constraints for Stage Two, Multiplicity Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	Donor's pair relationship must match recipient's pair relationship
LikC3	Donor and recipient must have the same household size
LikC10	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's value must match
LikC11	If recipient's count of household members aged 12 to 14 is nonmissing, then donor's value must match
LikC12	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's value must match
LikC13	If recipient's count of household members aged 15 to 17 is nonmissing, then donor's value must match
LikC14	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's value must match
LikC15	If recipient's count of household members aged 26 to 34 is nonmissing, then donor's value must match
LikC16	If recipient's count of household members aged 26 or older is nonmissing, then donor's value must match
LikC17	If recipient's count of household members aged 35 to 49 is nonmissing, then donor's value must match
LikC18	If recipient's count of household members aged 50 or older is nonmissing, then donor's value must match

Table E.114 Constraints and Portion of the Predictive Mean Vector for Parent-Child, Child Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Completely missing	116	None	1. PRED1	1-3,10,12,15,17,18: 71 2,3,10,12,15,17,18: 34 2,3,15,17,18: 8 2,16: 3

[1] PRED1 is the predicted mean for the number of parents of a child aged 12 to 20 who is a member of a parent-child pair.

Table E.115 Constraints and Portion of the Predictive Mean Vector for Parent-Child, Parent Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Completely missing	3	2[2]	1. PRED1	1-3,10: 2 2,3,10: 0 2,3: 0 2: 1

[1] PRED1 is the predicted mean for the number of children aged 12 to 20 belonging to a parent who is a member of a parent-child pair.

[2] A donor could not be found for one pair. For that case only we removed the portion of the logical constraint that requires the household count (15-17) of donor and recipient to match.

Table E.116 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-14/15-17), Older Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	1	3	1. PRED1	1,3,10,11,12: 1

[1] PRED1 is the predicted mean for the number of siblings aged 12 to 14 for a respondent aged 15 to 17 who is a member of a sibling-sibling pair.

Table E.117 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	2	4	1. PRED1	1,3,10,11,12: 1 3,10,11,12: 0 3,13: 1

[1] PRED1 is the predicted mean for the number of siblings aged 15 to 17 for a respondent aged 12 to 14 who is a member of a sibling-sibling pair.

Table E.118 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	5	5	1. PRED1	1,3,10,12,14: 4 3,10,12,14: 0 3,12: 1

[1] PRED1 is the predicted mean for the number of siblings aged 12 to 17 for a respondent aged 18 to 25 who is a member of a sibling-sibling pair.

Table E.119 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	4	6	1. PRED1	1,3,10,12,14: 3 3,10,12,14: 1

[1] PRED1 is the predicted mean for the number of siblings aged 18 to 25 for a respondent aged 12 to 17 who is a member of a sibling-sibling pair.

E.7.3 Stage Three: Household Counts

Table E.120 Variables Used in Constraints for Stage Three, Household Counts

Variable	Value	Constraint
Males XXXX[1]	0	No household members in the specified age group are male
	1	All household members in the specified age group are male
	2	Some household members in the specified age group are male
	3	There are no household members in the specified age group

[1] XXXX denotes the 4-digit age range (e.g., 1217 denotes the age range 12 to 17).

Table E.121 Logical Constraints for Stage Three, Household Counts

Constraint #	Logical Constraint
LogC1	Donor's count of older siblings aged 15 to 17 must fall within recipient's bounds
LogC2	Donor's count of older siblings aged 15 to 17 cannot be larger than recipient's nonmissing count of household members aged 15 to 17
LogC3	Donor's count of older siblings aged 18 to 25 must fall within recipient's bounds
LogC4	Donor's count of older siblings aged 18 to 25 cannot be larger than recipient's nonmissing count of household members aged 18 to 25
LogC5	Donor's count of spouse-spouse pairs must fall within recipient's bounds
LogC6	Donor's count of spouse-spouse pairs with children must fall within recipient's bounds
LogC7	Donor's parent-child (12-14) parent count must be greater than 0
LogC8	Donor's parent-child (12-17) parent count must be greater than 0
LogC9	Donor's parent-child (12-20) parent count must be greater than 0
LogC10	Donor's parent-child (12-14) child count must be greater than 0
LogC11	Donor's parent-child (12-17) child count must be greater than 0
LogC12	Donor's parent-child (12-20) child count must be greater than 0
LogC13	Donor's parent-child (12-14) child count must fall within recipient's bounds
LogC14	Donor's parent-child (12-17) child count must fall within recipient's bounds
LogC15	Donor's parent-child (12-20) child count must fall within recipient's bounds
LogC16	Donor's parent-child (12-14) parent count must fall within recipient's bounds
LogC17	Donor's parent-child (12-17) parent count must fall within recipient's bounds
LogC18	Donor's parent-child (12-20) parent count must fall within recipient's bounds

Table E.122 Likeness Constraints for Stage Three, Household Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient lives in a multi-family home, then donor must also live in a multi-family home
LikC3	Donor's household size must be the same as recipient's household size
LikC4	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's count must match

Table E.122 Likeness Constraints for Stage Three, Household Counts (continued)

Constraint #	Likeness Constraint
LikC5	If recipient's count of household members aged 12 to 14 is nonmissing, then donor's count must match
LikC6	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's count must match
LikC7	If recipient's count of household members aged 12 to 20 is nonmissing, then donor's count must match
LikC8	If recipient's count of household members aged 15 to 17 is nonmissing, then donor's count must match
LikC9	If recipient's count of household members aged 15 or older is nonmissing, then donor's count must match
LikC10	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's count must match
LikC11	If recipient's count of household members aged 26 to 34 is nonmissing, then donor's count must match
LikC12	If recipient's count of household members aged 26 to 49 is nonmissing, then donor's count must match
LikC13	If recipient's count of household members aged 35 to 49 is nonmissing, then donor's count must match
LikC14	If recipient's count of household members aged 50 or older is nonmissing, then donor's count must match
LikC15	Donor's and recipient's counts of household members aged 0 to 17 must both be 0, 1, or greater than 1
LikC16	Donor's and recipient's counts of household members aged 0 to 17 must either both be 0 or both be positive
LikC17	Donor's and recipient's counts of household members aged 12 to 14 must either both be 0 or both be positive
LikC18	Donor's and recipient's counts of household members aged 12 to 17 must either both be 0 or both be positive
LikC19	Donor's and recipient's counts of household members aged 15 to 17 must either both be 0 or both be positive
LikC20	Donor's and recipient's counts of household members aged 18 to 25 must either both be 0 or both be positive
LikC21	If recipient's value of MALES15P is nonmissing, then donor's value of MALES15P must match
LikC22	If recipient's count of household members aged 15 or older is positive, then donor and recipient must have the same count of household members aged 15 or older and the same values for MALES15P
LikC23	If recipient's value of MALES1825 is nonmissing, then donor's value of MALES1825 must match
LikC24	If recipient's value of MALES1834 is nonmissing, then donor's value of MALES1834 must match
LikC25	If recipient's value of MALES2634 is nonmissing, then donor's value of MALES2634 must match

Table E.122 Likeness Constraints for Stage Three, Household Counts (continued)

Constraint #	Likeness Constraint
LikC26	If recipient's value of MALES2649 is nonmissing, then donor's value of MALES2649 must match
LikC27	If recipient's value of MALES2649 is nonmissing and not equal to 3, then donor's value of MALES2649 must match
LikC28	If recipient's value of MALES3549 is nonmissing, then donor's value of MALES3549 must match
LikC29	If recipient's value of MALES50P is nonmissing, then donor's value of MALES50P must match
LikC30	If recipient could possibly have a parent-child (12-14) parent count of 0, then donor's and recipient's counts of household members aged 12 to 14 must match
LikC31	If recipient could possibly have a parent-child (12-17) parent count of 0, then donor's and recipient's counts of household members aged 12 to 17 must match
LikC32	If recipient could possibly have a parent-child (12-20) parent count of 0, then donor's and recipient's counts of household members aged 12 to 20 must match
LikC33	If recipient could possibly have a parent-child (12-14) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC34	If recipient could possibly have a parent-child (12-17) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC35	If recipient could possibly have a parent-child (12-20) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC36	If recipient could possibly have a parent-child (12-14) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match
LikC37	If recipient could possibly have a parent-child (12-17) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match
LikC38	If recipient could possibly have a parent-child (12-20) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match
LikC39	If recipient's count of household members aged 12 to 14 is nonmissing, then donor and recipient must have the same screener count of household members aged 12 to 14
LikC40	If recipient's count of household members aged 12 to 17 is nonmissing, then donor and recipient must have the same screener count of household members aged 12 to 17
LikC41	If recipient's count of household members aged 15 to 17 is nonmissing, then donor and recipient must have the same screener count of household members aged 15 to 17
LikC42	If recipient's count of household members aged 18 to 25 is nonmissing, then donor and recipient must have the same screener count of household members aged 18 to 25
LikC43	Donor and recipient must have parent-child (12-14) parent counts that are both 0 or both positive
LikC44	Donor and recipient must have parent-child (12-17) parent counts that are both 0 or both positive
LikC45	If recipient's counts of household members aged 12 to 14 and 15 to 17 are both nonmissing, then donor and recipient must have parent-child (12-14 and 12-17) parent counts that are both 0 or both positive
LikC46	If recipient's counts of household members aged 12 to 17 and 18 to 25 are both nonmissing, then donor and recipient must have parent-child (12-17) parent counts that are both 0 or both positive

Table E.123 Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (12-14/15-17), Older Focus

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing	61	1,2	1. H1	1,3-5,8,39,41,43,44: 15 3-5,8,39,41,43,44: 7 5,8,39,41,43,44: 4	1,3-5,8,39,41,43,44: 20 3-5,8,39,41,43,44: 11 5,8,39,41,43,44: 3 17,19,45: 1

[1] H1 is the predicted mean count of older siblings aged 15 to 17.

Table E.124 Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (12-17/18-25), Older Focus

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing	81	3,4	1. H2	1,3,4,6,10,40,42,44: 16 3,4,6,10,40,42,44: 9 6,10,40,42,44: 7 8,20,46: 2	1,3,4,6,10,40,42,44: 22 3,4,6,10,40,42,44: 15 6,10,40,42,44: 7 18,20,46: 3

[1] H2 is the predicted mean count of older siblings aged 18 to 25.

Table E.125 Constraints and Portion of the Predictive Mean Vector for Household Spouse-Spouse Count

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing	81	5	1. SPOUSE1 2. SPOUSE2	1-3,9-11,13-15,21,23,25,28,29: 6 2,3,9-11,13-15,21,23,25,28,29: 32 2,15,22,23,25,28,29: 15 2,22,24,26,29: 8 2,22,27: 1	1-3,9-11,13-15,21,23,25,28,29: 8 2,3,9-11,13-15,21,23,25,28,29: 5 2,15,22,23,25,28,29: 6

[1] The predictive mean vector components are defined by the following:
1. SPOUSE1 = P(the number of spouse-spouse pairs in the household is 0)
2. SPOUSE2 = P(the number of spouse-spouse pairs in the household is 1)

Table E.126 Constraints and Portion of the Predictive Mean Vector for Household Spouse-Spouse with Children Count

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing	42	6	1. SPOUSE3	1-3,9-11,13-15,21,23,25,28,29: 9 2,3,9-11,13-15,21,23,25,28,29: 14 2,15,22,23,25,28,29: 4 2,16,22,24,26,29: 4 2,16,22,27: 1	1-3,9-11,13-15,21,23,25,28,29: 6 2,3,9-11,13-15,21,23,25,28,29: 1 2,15,22,23,25,28,29: 3

[1] The predictive mean vector components are defined by the following:
1. SPOUSE3 = P(there is at least one spouse-spouse pair with children in the household)

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing PPCP1220	10	9,18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,32: 0 2-4,10,11,13,14,32: 9 2,4,10,11,13,14,32: 0 2,12,32: 1	No Cases
2	Missing PPCP1217	3	8,17	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,31: 0 2-4,10,11,13,14,31: 2 2,4,10,11,13,14,31: 1	No Cases
3	Missing PPCP1217 and PPCP1220	25	8,9,17,18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,31,32: 3 2-4,10,11,13,14,31,32: 20 2,4,10,11,13,14,31,32: 1 2,12,31,32: 1	No Cases
4	Missing PPCP1214	2	7,16	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,30: 0 2-4,10,11,13,14,30: 1 2,4,10,11,13,14,30: 1	No Cases
5	Missing PPCP1214 and PPCP1220	0	7,9,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
6	Missing PPCP1214 and PPCP1217	9	7,8,16,17	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,30,31: 0 2-4,10,11,13,14,30,31: 8 2,4,10,11,13,14,30,31: 0 2,12,30,31: 1	No Cases
7	Missing PPCP1214, PPCP1217, and PPCP1220	49	7-9,16-18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,30-32: 7 2-4,10,11,13,14,30-32: 32 2,4,10,11,13,14,30-32: 5 2,12,30-32: 4	1-4,10,11,13,14,30-32: 0 2-4,10,11,13,14,30-32: 1
8	Missing PPCC1220	9	12,15	1. PC1 2. PC2 3. H3	1-4,7,10,35: 0 2-4,7,10,35: 5 2,4,7,10,35: 1	1-4,7,10,35: 0 2-4,7,10,35: 2 2,4,7,10,35: 1
9	Missing PPCC1220 and PPCP1220	21	15,18	1. PC1 2. PC2 3. H3	1-4,7,10,11,13,14,32,35: 0 2-4,7,10,11,13,14,32,35: 7 2,4,7,10,11,13,14,32,35: 0 2,7,12,32,38: 3 2,7,32: 2	1-4,7,10,11,13,14,32,35: 1 2-4,7,10,11,13,14,32,35: 8
10	Missing PPCC1220 and PPCP1217	0	8,12,15,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
11	Missing PPCC1220, PPCP1217, and PPCP1220	1	8,15,17,18	1. PC1 2. PC2 3. H3	1-4,7,10,11,13,14,31,32,35: 0 2-4,7,10,11,13,14,31,32,35: 1	No Cases
12	Missing PPCC1220 and PPCP1214	0	7,12,15,16	1. PC1 2. PC2 3. H3	No Cases	No Cases
13	Missing PPCC1220, PPCP1214, and PPCP1220	0	7,15,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
14	Missing PPCC1220, PPCP1214, and PPCP1217	0	7,8,12,15-17	1. PC1 2. PC2 3. H3	No Cases	No Cases
15	Missing PPCC1220, PPCP1214, PPCP1217, and PPCP1220	2	7,8,15-18	1. PC1 2. PC2 3. H3	1-4,7,10,11,13,14,30-32,35: 0 2-4,7,10,11,13,14,30-32,35: 0 2,4,7,10,11,13,14,30-32,35: 0 2,7,12,30-32,38: 1	1-4,7,10,11,13,14,30-32,35: 1
16	Missing PPCC1217	0	11,14	1. PC1 2. PC2 3. H3	No Cases	No Cases
17	Missing PPCC1217 and PPCP1220	0	9,11,14,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
18	Missing PPCC1217 and PPCP1217	0	14,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
19	Missing PPCC1217, PPCP1217, and PPCP1220	0	9,14,17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
20	Missing PPCC1217 and PPCP1214	0	7,11,14,16	1. PC1 2. PC2 3. H3	No Cases	No Cases
21	Missing PPCC1217, PPCP1214, and PPCP1220	0	7,9,11,14,16, 18	1. PC1 2. PC2 3. H3	No Cases	No Cases
22	Missing PPCC1217, PPCP1214, and PPCP1217	0	7,14,16,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
23	Missing PPCC1217, PPCP1214, PPCP1217, and PPCP1220	0	7,9,14,16-18	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
24	Missing PPCC1217 and PPCC1220	6	11,12,14,15	1. PC1 2. PC2 3. H3	1-4,6,7,10,34,35: 0 2-4,6,7,10,34,35: 5	1-4,6,7,10,34,35: 0 2-4,6,7,10,34,35: 1
25	Missing PPCC1217, PPCC1220, and PPCP1220	0	11,14,15,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
26	Missing PPCC1217, PPCC1220, and PPCP1217	1	12,14,15,17	1. PC1 2. PC2 3. H3	1-4,6,7,10,11,13,14,31,34,35: 0 2-4,6,7,10,11,13,14,31,34,35: 1	No Cases
27	Missing PPCC1217, PPCC1220, PPCP1217, and PPCP1220	30	14,15,17,18	1. PC1 2. PC2 3. H3	1-4,6,7,10,11,13,14,31,32,34,35: 3 2-4,6,7,10,11,13,14,31,32,34,35: 10 2,4,6,7,10,11,13,14,31,32,34,35: 0 2,6,7,12,31,32,37,38: 9	1-4,6,7,10,11,13,14,31,32,34,35: 2 2-4,6,7,10,11,13,14,31,32,34,35: 4 2,4,6,7,10,11,13,14,31,32,34,35: 0 2,6,7,12,31,32,37,38: 2
28	Missing PPCC1217, PPCC1220, and PPCP1214	0	7,11,12,14-16	1. PC1 2. PC2 3. H3	No Cases	No Cases
29	Missing PPCC1217, PPCC1220, PPCP1214, and PPCP1220	0	7,11,14-16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
30	Missing PPCC1217, PPCC1220, PPCP1214, and PPCP1217	0	7,12,14-17	1. PC1 2. PC2 3. H3	No Cases	No Cases
31	Missing PPCC1217, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	1	7,14-18	1. PC1 2. PC2 3. H3	1-4,6,7,10,11,13,14,30-32,34,35: 0 2-4,6,7,10,11,13,14,30-32,34,35: 1	No Cases
32	Missing PPCC1214	1	10,13	1. PC1 2. PC2 3. H3	1-5,10,33: 0 2-5,10,33: 0 2,4,5,10,33: 1	No Cases
33	Missing PPCC1214 and PPCP1220	0	9,10,13,18	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
34	Missing PPCC1214 and PPCP1217	0	8,10,13,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
35	Missing PPCC1214, PPCP1217, and PPCP1220	0	8-10,13,17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
36	Missing PPCC1214 and PPCP1214	5	13,16	1. PC1 2. PC2 3. H3	1-5,10,11,13,14,30,33: 0 2-5,10,11,13,14,30,33: 4 2,4,5,10,11,13,14,30,33: 0 2,5,12,30,36: 1	No Cases
37	Missing PPCC1214, PPCP1214, and PPCP1220	0	9,13,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
38	Missing PPCP1214 and PPCP1217	0	8,13,16,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
39	Missing PPCP1214, PPCP1217, and PPCP1220	0	8,9,13,16-18	1. PC1 2. PC2 3. H3	No Cases	No Cases
40	Missing PPCC1214 and PPCC1220	0	10,12,13,15	1. PC1 2. PC2 3. H3	No Cases	No Cases
41	Missing PPCC1214, PPCC1220, and PPCP1220	0	10,13,15,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
42	Missing PPCC1214, PPCC1220, and PPCP1217	0	8,10,12,13, 15,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
43	Missing PPCC1214, PPCC1220, PPCP1217, and PPCP1220	0	8,10,13,15, 17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
44	Missing PPCC1214, PPCC1220, and PPCP1214	0	12,13,15,16	1. PC1 2. PC2 3. H3	No Cases	No Cases
45	Missing PPCC1214, PPCC1220, PPCP1214, and PPCP1220	0	13,15,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
46	Missing PPCC1214, PPCC1220, PPCP1214, and PPCP1217	0	8,12,13,15-17	1. PC1 2. PC2 3. H3	No Cases	No Cases
47	Missing PPCC1214, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	0	8,13,15-18	1. PC1 2. PC2 3. H3	No Cases	No Cases
48	Missing PPCC1214 and PPCC1217	0	10,11,13,14	1. PC1 2. PC2 3. H3	No Cases	No Cases
49	Missing PPCC1214, PPCC1217, and PPCP1220	0	9-11,13,14,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
50	Missing PPCC1214, PPCC1217, and PPCP1217	0	10,13,14,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
51	Missing PPCC1214, PPCC1217, PPCP1217, and PPCP1220	0	9,10,13,14, 17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
52	Missing PPCC1214, PPCC1217, and PPCP1214	0	11,13,14,16	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
53	Missing PPCC1214, PPCC1217, PPCP1214, and PPCP1220	0	9,11,13,14, 16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
54	Missing PPCC1214, PPCC1217, PPCP1214, and PPCP1217	2	13,14,16,17	1. PC1 2. PC2 3. H3	1-6,10,11,13,14,30,31,33,34: 0 2-6,10,11,13,14,30,31,33,34: 2	No Cases
55	Missing PPCC1214, PPCC1217, PPCP1214, PPCP1217, and PPCP1220	0	9,13,14,16-18	1. PC1 2. PC2 3. H3	No Cases	No Cases
56	Missing PPCC1214, PPCC1217, and PPCC1220	8	10-15	1. PC1 2. PC2 3. H3	1-7,10,33-35: 0 2-7,10,33-35: 5 2,4-7,10,33-35: 1 2,5-7,36-38: 0 2,5-7: 0 2: 1	1-7,10,33-35: 0 2-7,10,33-35: 1
57	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1220	0	10,11,13-15, 18	1. PC1 2. PC2 3. H3	No Cases	No Cases
58	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1217	0	10,12-15,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
59	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1217, and PPCP1220	0	10,13-15, 17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
60	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1214	2	11-16	1. PC1 2. PC2 3. H3	1-7,10,11,13,14,30,33-35: 0 2-7,10,11,13,14,30,33-35: 1	1-7,10,11,13,14,30,33-35: 0 2-7,10,11,13,14,30,33-35: 1
61	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, and PPCP1220	0	11,13-16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
62	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, and PPCP1217	1	12-17	1. PC1 2. PC2 3. H3	1-7,10,11,13,14,30,31,33-35: 0 2-7,10,11,13,14,30,31,33-35: 1	No Cases
63	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	30	13-18	1. PC1 2. PC2 3. H3	1-7,10,11,13,14,30-35: 0 2-7,10,11,13,14,30-35: 13 2,4-7,10,11,13,14,30-35: 0 2,5-7,12,30-32,36-38: 6 2,5-7,30-32: 1	1-7,10,11,13,14,30-35: 0 2-7,10,11,13,14,30-35: 8 2,4-7,10,11,13,14,30-35: 0 2,5-7,12,30-32,36-38: 1 2,5-7,30-32: 1

[1] The predictive mean vector components are defined by the following:
1. H3 = Predicted mean count of children aged 12 to 20 with parents in the household
2. PC1 = P(there is one parent with at least one child aged 12 to 20 in the household)
3. PC2 = P(there are two parents with at least one child aged 12 to 20 in the household)

Appendix F: Quality Control Measures Used in the Imputation Procedures

Appendix F: Quality Control Measures Used in the Imputation Procedures

F.1 Introduction

For the 2014 National Survey on Drug Use and Health (NSDUH), a number of quality control (QC) measures were implemented for the imputation of demographic, drug use, income, health insurance, nicotine dependence, household composition (roster), and pair variables. These QC measures spanned the following three basic steps within the predictive mean neighborhood (PMN) imputation methodology: (1) response propensity adjustment, (2) prediction modeling, and (3) hot-deck imputation.

Checklists for the imputation of these variables were used during imputation processing for the 2014 survey and serve as formal documentation of the QC measures that were implemented. Most of these checks required both a technician check, where the individual who ran the computer program (technician) entered his or her name and the date the check was performed, and an independent reviewer check, where another person performed a check of the same items. The reviewer check ensured greater quality in the imputation procedures. Additionally, some checklist entries required the technician to document the procedures that were taken to run the programs, such as listing the variables that were dropped from the model in order to achieve model convergence.

The checks involved in each of the PMN steps are described in detail in the following sections. Variable-specific consistency checks were often created in addition to the checks of the PMN steps. However, they are not discussed in this appendix. The QC measures that were used in the editing process or the delivery of variables to other NSDUH teams also are not discussed.

F.2 Step 1: Response Propensity Adjustment

The first step of the PMN methodology, response propensity adjustment, adjusts the sampling weights for item nonresponse so that the item respondent weights that are used only during the imputation process are representative of the entire domain of interest. The item response propensity is modeled as a function of a predetermined set of covariates. The model can be thought of as a special case of the generalized exponential model (GEM)[1] developed for weighting procedures, in that imputations that are done at the item level are similar in nature to the weight adjustments made for entire units.

The following QC measures were conducted in Step 1:

[1] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures and is described in detail in Appendix A of the person-level sampling weight calibration report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2016a).

- The output of the response propensity modeling program was checked for singularities. Any singularities that occurred were investigated, and the model was corrected by removing the correlated covariates from the model.
- The values of the convergence flags were checked to see whether the GEM converged.
- The current distance to root was checked to make sure it was close to 0, indicating that GEM converged.
- When variables were reduced from the original model, the remaining levels of variables were checked to ensure appropriateness. An example of this check was to determine whether the base variables or lower order terms were present when interactions or higher order terms existed (e.g., "age" and "age squared" must be in the model if "age cubed" is in the model).
- An indicator was calculated in the response propensity program that measured the maximum adjustment to the weights. In most cases, the adjusted weights resembled the original design weights. If the maximum adjustment was too high (usually greater than 3.00), this was likely due to an overspecified model, where the adjustment was not performing at an optimum level. Large maximum adjustments were investigated and corrected, if possible, by removing extraneous variables from the models so that any final weight adjustment applied to a respondent was within acceptable bounds.
- After the weights were adjusted, the ratio of the maximum adjusted weight to the mean adjusted weight (called the "mmratio") was computed to monitor the variation among the weights. Any mmratio value that was greater than 25 percent was noted in the response propensity program checklist.
- The unequal weighting effect (UWE) was checked before and after adjustment to ensure that there was no significant variance increase due to the nonresponse adjustment. The difference in the UWE after-adjustment value should be no more than 20 percent of the UWE before-adjustment value. The difference was fairly small in most cases, and any difference greater than 20 percent was investigated and corrected, if possible.
- When using the SAS procedure PROC MEANS to examine summary statistics, the weighted totals for the independent variables in the model were compared both before and after the adjustment. If these weighted totals were equal, the adjustment procedures worked properly.

F.3 Step 2: Prediction Modeling

The prediction modeling step, using the weights from Step 1, calculates predicted means, which are then used in the hot-deck step(s) to create neighborhoods and select donors. The dependent variable in the model is usually the variable, or some transformation of that variable, that is undergoing imputation. Each model is built using only those cases within the domain with complete responses for that item. Predicted means are then calculated for all of the domain members, whether or not they were item respondents, using the values for the covariates and the estimates for the regression coefficients.

For categorical outcome variables, logistic regression models are used for the prediction models. For continuous variables, linear regression models are fit. For count variables, Poisson regression models are used. For response variables that are proportions (e.g., months on welfare), a logit transformation is applied to the proportion, and a linear regression model is utilized. For continuous nonnegative random variables (e.g., finer categories of income), a failure time model is used.

The models are discussed in detail in the imputation sections of Chapters 4 through 11 of this report. The following QC measures were conducted in Step 2:

- All models were checked for singularities and collinearities. Any singularities that occurred were investigated, and the model was corrected.
- For Poisson regression models, failure time models, and logistic models, the output was checked to ensure convergence had occurred.
- For logistic models, the log file was checked for "data warning" messages or other SUDAAN®-specific errors.[2] If there was a "data warning" message in the log, the SUDAAN model was determined to be unstable, and variables were removed to produce stability in the estimates. Similar to the response propensity model, if the main variable was dropped, its interaction variables also were dropped.
- When variables were reduced from the original model, the remaining levels of variables were checked to ensure appropriateness.
- Output was checked to verify that everything worked properly in the regression model.
- A check on the predicted means from the model was created to ensure that each respondent in the domain had a valid nonmissing predicted mean.

F.4 Step 3: Hot-Deck Imputation

The predicted means from Step 2 were used to determine the final assignments of imputed values in a hot-deck step. The goal of this step was to make donors and recipients as similar as possible. A neighborhood of potential donors was used, if possible, so that the donor selected was different each time the procedure was run. However, all potential donors in a neighborhood usually had very similar predicted means.

The QC measures in this step had two objectives: (1) ensure that the imputed values were consistent with preexisting nonmissing values, and (2) ensure that the imputed values were assigned as intended. Many variables had logical constraints that had to be adhered to in order to prevent logical inconsistencies. Specific QC checks were implemented to ensure this; however, only the more general checks that apply to all PMN imputations are listed in this appendix. For more detail on logical constraints, refer to Appendix E. The following QC measures were conducted in Step 3:

---

[2] Details can be found in the *SUDAAN® Language Manual, Release 11.0.1* (RTI International, 2013).

- Unusual imputed values were noted. If the imputed value was equivalent to one of the standard NSDUH missing value codes, this signaled a failure to obtain a donor, and measures were required (e.g., likeness constraints were loosened) to revise the programs so that a donor could be found. If the imputed value was otherwise unusual, the imputation process was examined to ensure that no errors occurred.
- The number of cases that were imputed within various levels of restrictiveness of the likeness constraints (as determined by the variable SMALLFLG) was noted.[3]
- The frequency of the variable "WORKED" was checked to ensure that no values were equal to 0. Values greater than 0 signified that the imputation procedure was able to find a donor for all missing cases.
- The distribution of edited variables was compared with the distribution of imputed variables to make sure that each imputed value was within the appropriate range corresponding to the value of the edited variable.
- The imputed values were crossed with the imputation indicators to ensure that the indicators were created correctly.
- After imputation was implemented, the distribution of values for nonrespondents was checked against the distribution of values for all respondents to ensure the similarity of these two subgroups.
- It was necessary to ensure that everyone to whom the variable did not apply received a skip code for the final imputed variable. For example, all those in the 12-to-14 age group should have a "Not Applicable (N/A)" value of 99 for the imputation-revised marital status variable IRMARIT.
- Any missing values were noted. This occurred when the program was unsuccessful in assigning a valid imputed value.
- Any cases where the imputed value was not consistent with preexisting nonmissing values were noted.
- It was necessary to ensure that any restrictions on the final imputed value for a given nonrespondent were honored. For example, some respondents were known to be employed, but either full-time or part-time employment status was not known. Checks were conducted to ensure that these respondents had either full-time or part-time status assigned to the employment status variable (EMPSTAT4), but not unemployed or other statuses.
- When multiple missingness patterns existed, each pattern was treated separately, and the distribution of imputed values within each missingness pattern was investigated. For example, if it was known that a respondent was a past year user, both past month and past year users should have been included among the imputed values, not just past month users.

[3] Refer to Appendix E for more details about likeness constraints and the "SMALLFLG" variable.

Appendix G: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

This page intentionally left blank

Appendix G: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

G.1 Introduction

In the household roster for the 2014 National Survey on Drug Use and Health (NSDUH), the interviewer was supposed to enter a roster of the respondent's entire household, which included age, gender, and the relationship to the respondent. It was not uncommon for the interviewer to enter a relationship code, age, or gender that was illogical, based on the age and gender of the respondent given in the core part of the questionnaire. Before the computer-assisted interviewing (CAI) instrument was implemented in the 1999 NSDUH, such responses would have been flagged at the data processing stage.[1] However, beginning with the 2000 survey and in every survey year since then, consistency checks have been added to the CAI instrument that allowed the interviewer, if needed, to correct the error while giving the interview. Details about these consistency checks are presented in Chapter 8.

In general, two types of consistency checks were implemented in the 2014 survey. The first type compared the entry in the roster with previously entered questionnaire information, specifically the respondent's age (CURNTAGE) and gender. The second type checked for internal consistency within the household roster. In some cases, a consistency check would be triggered even though the response was legitimate. This occurred if CURNTAGE was considered incorrect, for example, or in extremely rare family situations such as when a stepmother was younger than her stepson. With the exception of the check against the previously entered respondent's gender, the interviewer could override the consistency check and explain why the response given was correct. In some cases, the interviewer was correct in overriding the consistency check. In others, however, it was clear that the interviewer misunderstood how the roster should have been put together and that the override to the consistency check was not legitimate.

This appendix summarizes the explanations given by interviewers for consistency check overrides in the household roster. It is divided into two parts: consistency check overrides involving CURNTAGE and those involving internal consistency checks.

G.2 Override Comments from Interviewers: Comparisons with CURNTAGE

When an interviewer entered the respondent's roster entry (the "self" entry), if the age did not match the age previously entered in the questionnaire, a consistency check was triggered. The comparison was between the roster age for the "self" and CURNTAGE, which was the value of the final questionnaire-edited age (AGE) stored by Blaise.[2] Explanations for consistency check overrides for the variable CURNTAGE are provided in Table G.1. Because CURNTAGE

---

[1] Because the age and gender of the respondent given in the core part of the questionnaire were not allowed to change, the relationship code and sometimes the age of the roster member were set to bad data.

[2] The Blaise program is the computer program within the CAI instrument that was used to direct the respondent and interviewer through the questionnaire.

had the potential to change constantly throughout the questionnaire, no final variable with this name was created. However, in most cases, the value of CURNTAGE when the roster commenced was equivalent to NEWAGE, the value of CURNTAGE after the drug modules had been completed. In theory, NEWAGE was not always equivalent to AGE, the derivation of which is described in Chapter 4.

In the 2002 survey, the explanations provided in Table G.1 were not reviewed when determining AGE, nor were they reviewed when determining the final value of the age for the "self" entry in the roster. However, beginning with the 2003 survey, these explanations have been carefully reviewed. In rare cases, the final value of AGE was set to the age of the self in the questionnaire roster (the "roster age"), based on these explanations as well as other evidence, even if it disagreed with the age as it would have been calculated in prior survey years. Details about this process are provided in Chapter 4.

Even in cases where the explanation seemed clear that CURNTAGE was incorrect, the value of AGE was not always set to the roster age. In most cases, this was because the difference between CURNTAGE and the roster age was 1 year or less. A difference of 1 year was tolerated, because legitimate differences could result from a birthday occurring in the time between the drug modules and the roster.[3] In other situations, the value of CURNTAGE was incorrect, but the original questionnaire-edited age was correct, so no change was necessary. In still other cases, not all the criteria that were necessary for changing the value of AGE to be equal to the roster age were met. Cases where the value of AGE was changed to roster age are denoted in the "Comment" column in bolded italics in Table G.1. Otherwise, the reason for not changing the value of AGE to roster age also is shown in this column. The "Respondent's Age in Roster of Other Pair Member" column indicates whether the roster of the other pair member, if it existed, supported CURNTAGE or the override age as the respondent's age.

G.3 Override Comments from Interviewers: Internal Consistency Check Overrides

Internal consistency checks were performed on the household roster for the 2014 NSDUH. Interviewer explanations for overrides to these internal consistency checks were evaluated individually to determine their legitimacy. In the 2014 NSDUH, there were 446 interviewer explanations for overrides to these internal consistency checks. Of these, 434 were considered legitimate; no edit was applied to the age or relationship code of the roster member. The remaining 12 explanations were not considered legitimate; thus, the override was overruled, and the relationship code (and sometimes the roster member’s age) was altered accordingly. Interviewer explanations for overrides to these internal consistency checks not considered legitimate are provided in Table G.2. Also provided in this table are the questionnaire-edited age of the respondent (AGE), the age and relationship to the respondent of the roster member in question, and, in the "Comment" column, a brief indication of the probable true relationship of the roster member to the respondent.

[3] It was not uncommon for an interview to be conducted in more than one sitting. This could have occurred if either the respondent or the interviewer did not have enough time for the interview or otherwise could not complete the interview in a single sitting.

Table G.1 Explanations for Overrides to Consistency Checks against CURNTAGE

#	NEW AGE	Original Roster Age for Self	Screener Age	Verbatim Explanation from Field Interviewers[1]	Respondent's Age in Roster of Other Pair Member	Comment[2]	AGE = Final Roster Age
1	28	27	27	the r is still 27 bday later in JUly	28	Diff. ≤ 1 year	28
2	45	44	44	the respondent gave a different age for the screening	Not in a pair	Diff. ≤ 1 year	45
3	51	52	52	wife made mistake on husb age	Not in a pair	Diff. ≤ 1 year	51
4	23	24	24	resp is 24 not 23	Not in a pair	Diff. ≤ 1 year	23
5	18	17	18	response given by interviewer	18	Diff. ≤ 1 year	18
6	16	15	16	IR says he is 15 yrs.	Not in a pair	Diff. ≤ 1 year	16
7	70	71	70	her age is 70	Not in a pair	Diff. ≤ 1 year	70
8	16	15	16	R states she turned 16 in Feb. She was thus 15 on her last birthday.	16	Diff. ≤ 1 year	16
9	16	15	16	screener gave r's age as 16 when she is 15	16	Diff. ≤ 1 year	16
10	37	38	38	respondent born 06/1976 currently 38	Not in a pair	Diff. ≤ 1 year	37
11	28	27	28	He now says he is 27, after first answering 28	28	Diff. ≤ 1 year	28
12	21	20	21	response per IR	21	Diff. ≤ 1 year	21
13	21	20	21	Respondent is 20.	Not in a pair	Diff. ≤ 1 year	21
14	24	25	25	He says he really is 25, as appears in the roster	24	Diff. ≤ 1 year	24
15	30	31	31	30	Not in a pair	Diff. ≤ 1 year	30
16	14	13	13	had a recent birthday	14	Diff. ≤ 1 year	14
17	20	19	19	was in error	Not in a pair	Diff. ≤ 1 year	20
18	47	48	47	47	47	Diff. ≤ 1 year	47
19	48	49	46	r is 49 years old	48	Diff. ≤ 1 year	48
20	28	29	29	FIRST R HAD AGE INCORRECT	28	Diff. ≤ 1 year	28
21	28	27	27	She doesn't remember if she is 27 or 28	Not in a pair	Diff. ≤ 1 year	28
22	49	48	49	he gave me wrong age	Not in a pair	Diff. ≤ 1 year	49
23	73	72	72	73	73	Diff. ≤ 1 year	73
24	19	18	18	unsure how to correct	19	Diff. ≤ 1 year	19
25	42	41	42	SR listed IR as 41 on 9/13 she just turned 42 on 9/10	42	Diff. ≤ 1 year	42
26	15	14	15	last birthday	15	Diff. ≤ 1 year	15
27	43	44	44	hit wrong button	43	Diff. ≤ 1 year	43
28	35	36	36	Do not find that he was entered as 35	Not in a pair	Diff. ≤ 1 year	35

Table G.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW AGE	Original Roster Age for Self	Screener Age	Verbatim Explanation from Field Interviewers[1]	Respondent's Age in Roster of Other Pair Member	Comment[2]	AGE = Final Roster Age
29	45	44	43	on last bday 44,45 this year	45	Diff. ≤ 1 year	45
30	38	37	37	gave age at next b day	Not in a pair	Diff. ≤ 1 year	38
31	55	54	55	IR mistake	Not in a pair	Diff. ≤ 1 year	55
32	46	47	47	man gave wron g answer	46	Diff. ≤ 1 year	46
33	47	48	48	r is actually 47	Not in a pair	Diff. ≤ 1 year	47
34	16	17	17	17	16	Diff. ≤ 1 year	16
35	33	34	34	m is 34	33	Diff. ≤ 1 year	33
36	46	47	47	Respondent is age 47	46	Diff. ≤ 1 year	46
37	45	44	44	respondent is currently 44.	45	Diff. ≤ 1 year	45
38	70	69	69	r states he is 69 almost 70.	Not in a pair	Diff. ≤ 1 year	70
39	14	13	14	the respondent states he is 14 years old now.	14	Diff. ≤ 1 year	14
40	45	44	45	45 is correct age	45	Diff. ≤ 1 year	45
41	20	19	19	screen r did not have right age.	20	Diff. ≤ 1 year	20
42	16	15	15	r will be 16 in october	16	Diff. ≤ 1 year	16
43	46	45	43	mother is 45	46	Diff. ≤ 1 year	46
44	41	40	41	respondent finished 40 in september	41	Diff. ≤ 1 year	41
45	26	25	25	birthday since screening	Not in a pair	Diff. ≤ 1 year	26
46	40	41	41	The R is actually 41 not 40	40	Diff. ≤ 1 year	40
47	22	23	22	IR accidently told me he was 22, but he is actually 23	22	Diff. ≤ 1 year	22
48	15	14	15	Screening done by step mom/this child doesn't know his own age. autistic	15	Diff. ≤ 1 year	15
49	50	49	49	24	50	Diff. ≤ 1 year	50
50	77	76	76	Respondant says she is 76 but calculations says she is 77	77	Diff. ≤ 1 year	77
51	24	23	23	Respondent is now 24	Not in a pair	Diff. ≤ 1 year	24
52	39	38	39	the respondent states his age is 39.	39	Diff. ≤ 1 year	39
53	26	25	25	IR is 25	Not in a pair	Diff. ≤ 1 year	26
54	58	59	59	verified actual age with respondent	Not in a pair	Diff. ≤ 1 year	58
55	30	29	30	self	30	Diff. ≤ 1 year	30

Table G.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW AGE	Original Roster Age for Self	Screener Age	Verbatim Explanation from Field Interviewers[1]	Respondent's Age in Roster of Other Pair Member	Comment[2]	AGE = Final Roster Age
56	80	81	81	she is 81 did break-off yesterday	Not in a pair	Diff. ≤ 1 year	80
57	24	23	23	error in age given in the beginning	Not in a pair	Diff. ≤ 1 year	24
58	21	22	22	R is 21	Not in a pair	Diff. ≤ 1 year	21
59	33	32	32	When screened she was entered as 32 by her mother in law	33	Diff. ≤ 1 year	33
60	32	33	33	resp told his age and birthday	Not in a pair	Diff. ≤ 1 year	32
61	49	48	48	not sure the info was correct	49	Diff. ≤ 1 year	49
62	29	28	29	respondant is 29	Not in a pair	Diff. ≤ 1 year	29
63	37	36	36	its his age	Not in a pair	Diff. ≤ 1 year	37
64	15	14	15	R told FI age was 14, mother gave age as 15 during screening.	Not in a pair	Diff. ≤ 1 year	15
65	68	67	65	IR said her age is 67, earlier response of 68 incorrect.	68	Diff. ≤ 1 year	68
66	31	30	30	30	Not in a pair	Diff. ≤ 1 year	31
67	51	50	45	he is saying his birthday is April 4 1964	51	Diff. ≤ 1 year	51
68	28	27	28	HE ANSWERED AS LAST BIRTHDAY r IS 28	Not in a pair	Diff. ≤ 1 year	28
69	42	43	42	Householder is actually 43. He stated 42 in roster.	42	Diff. ≤ 1 year	42
70	24	23	24	He made a mistake	24	Diff. ≤ 1 year	24
71	39	38	37	she put next birthdays age earlier	39	Diff. ≤ 1 year	39
72	16	15	16	R gave incorrect age at first	16	Diff. ≤ 1 year	16
73	56	58	54	56	Not in a pair	Diff. > 1 year, FI does not support change leave as NEWAGE	56
74	56	59	59	R age is actually 59 not 56.	Not in a pair	***AGE was changed to equal roster age***	59

Table G.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW AGE	Original Roster Age for Self	Screener Age	Verbatim Explanation from Field Interviewers[1]	Respondent's Age in Roster of Other Pair Member	Comment[2]	AGE = Final Roster Age
75	45	48	38	THIS IS RESPONDENT	Not in a pair	Diff. > 1 year, FI does not support change leave as NEWAGE	45
76	50	53	50	male is 53	50	Diff. > 1 year, other pair member supports NEWAGE	50
77	30	33	33	30	30	Diff. > 1 year, other pair member supports NEWAGE	30
78	25	33	33	The birth date given was incorrect, should have been 08-21-1981	33	***AGE was changed to equal roster age***	33
79	36	46	46	r is listed as 46 in roster	Not in a pair	Diff. > 1 year, FI does not support change leave as NEWAGE	36
80	47	37	47	R is definietly 47 age 37 is error	Not in a pair	Diff. > 1 year, FI does not support change leave as NEWAGE	47
81	69	997	68	The R refused to provide the ages of all the HH's including herself?	Not in a pair	FI does not support change leave as NEWAGE	69
82	47	994	42	error	Not in a pair	Roster Age Missing	47
83	46	997	39	refuses	46	other pair member supports NEWAGE	46
84	41	997	40	Refuse to say age	41	other pair member supports NEWAGE	41
85	38	997	37	respondent didn't want to give this informtion	Not in a pair	FI does not support change leave as NEWAGE	38

Table G.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW AGE	Original Roster Age for Self	Screener Age	Verbatim Explanation from Field Interviewers[1]	Respondent's Age in Roster of Other Pair Member	Comment[2]	AGE = Final Roster Age
86	25	997	25	r does not want to answer those questions, and the error box kept popping up	Not in a pair	FI does not support change leave as NEWAGE	25

[1] These entries came directly from the 2014 NSDUH field interviewers. Any typographical errors or misspellings were transcribed directly and were not corrected.

[2] "Diff." refers to the difference between CURNTAGE and the age of the "self" in the household roster, the "roster age." Bolded and italicized entries indicate that the criteria for changing the age to that given in the household roster for "self" were met.

Table G.2 Explanations for Overrides to Internal Consistency Checks

#	Consistency Check	AGE	Roster Member's Age and Relationship to Respondent	Verbatim Explanation from Field Interviewers[1]	Comment
1	Respondent is 16 or younger and has an in-law	15	35-year-old mother-in-law	R is having difficulty w roster. dad hd indicated has a little disability	Overrule; probable parent
2	Grandparent/Grandchild and respondent less than 30 years apart	15	9-year-old grandson	R is having difficulty w roster. dad hd indicated has a little disability	Overrule; probable sibling
3	Multiple spouses or partners	36	36-year-old unmarried partner	not a traditional household	Overrule; unsure of relationship (85)
4	Multiple spouses or partners	36	30-year-old unmarried partner	not a traditional household	Overrule; unsure of relationship (85)
5	Respondent is 16 or younger and has a spouse or partner	13	34-year-old unmarried partner	34	Overrule; probable parent
6	Respondent is 16 or younger and has a spouse or partner	13	27-year-old unmarried partner	unmarried partner	Overrule; probable parent
7	Respondent is 16 or younger and has an in-law	15	41-year-old father-in-law	The R minor thinks that his Mom's boyfriend is his Father-in-law	Overrule; unsure of relationship (85)
8	Respondent's female in-law is 16 or younger	26	7-year-old daughter-in-law	This HH is the daughter of the R's girlfriend.	Overrule; probable child
9	Respondent's female in-law is 16 or younger	25	5-year-old daughter-in-law	he is the stepfather	Overrule; unsure of relationship (85)
10	Respondent's mother is less than 13 years older than respondent	25	36-year-old mother	Respondant verified info is correct	Overrule; unsure of relationship (85)
11	Respondent's mother is less than 13 years older than respondent	12	23-year-old mother	he doesn't understand that she's not his real biological mother	Overrule; unsure of relationship (85)
12	Respondent's son is less than 13 years younger than respondent	15	8-year-old son	R states this is correct age	Overrule; probable sibling

[1] These entries came directly from the 2014 NSDUH field interviewers. Any typographical errors or misspellings were transcribed directly and were not corrected.

Appendix H: Rules for Determining Pair Relationships

This page intentionally left blank

Appendix H: Rules for Determining Pair Relationships

H.1 Rules for Determining Matching Pairs, in Priority Order

The following rules are used to determine the member of a respondent's household roster that corresponds to the other pair member. In these rules, an "age match" occurs if the questionnaire age of one pair member matches a roster age in the other pair member's roster, and a "gender match" occurs if the questionnaire gender of one of the pair members matches a roster gender in the other pair member's roster. In the table below, if the rules for Pair Member A and Pair Member B in a single row differ, then the count for that row includes the rules as listed, and the rules with Pair Member A and Pair Member B reversed. If the age and/or gender are off when finding these matches, the age and/or gender are defined by the questionnaire age and gender of the selected pair member when determining the pair domain. The rules, called match types, are listed in priority order in Table H.1, along with the number of pairs to which each rule was applied. Since the 2001 survey, it was technically impossible to identify more than one roster member as the "other pair member selected," resulting in either 0 or 1 MBRSEL for each responding pair. Rules involving situations where more than one MBRSEL existed are therefore not included in this table.

Table H.1 Rules for Determining Matching Pairs, in Priority Order

Match Type	Rule		
	Pair Member A	Pair Member B	Count
0.00	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, exactly one MBRSEL in right place	14,418
1.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	2,254
2.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	157
3.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	333
3.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	43
3.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	6

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
4.01	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	338
4.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	38
4.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	5
4.05	Age and gender match exactly, MBRSEL missing for all roster members	Age and gender match exactly, MBRSEL missing for all roster members	22
4.06	Age and gender match exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, relationship is "Other Relative"	0
4.08	Age within one, gender matches exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, relationship is "Other Relative"	0
4.09	Age within two, gender matches exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, relationship is "Other Relative"	0
4.10	Age and gender match exactly, MBRSEL missing for all roster members	Two matches for age and gender, MBRSEL missing for all roster members, relationship is "Other Relative"	0
4.11	Age and gender match exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, roommate type relationship	0
4.13	Age within one, gender matches exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, roommate type relationship	1
4.14	Age within two, gender matches exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, roommate type relationship	0
4.15	Age and gender match exactly, MBRSEL missing for all roster members	Two matches for age and gender, MBRSEL missing for all roster members, roommate type relationship	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	**Pair Member B**	**Count**
5.01	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	20
5.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	2
5.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	1
5.05	Age and gender match exactly, MBRSEL missing for all roster members	Age matches exactly, gender off, exactly one MBRSEL in right place	0
5.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	0
8.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	26
8.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	10
8.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	1
8.05	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender matches exactly, MBRSEL missing for all roster members	3
8.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	0
8.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age within one, gender matches exactly, MBRSEL missing for all roster members	0
10.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	262

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	**Pair Member B**	**Count**
10.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	38
10.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	9
10.05	Age and gender match exactly, MBRSEL missing for all roster members	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	4
10.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	1
10.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	1
10.11	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	1
11.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	5
11.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
11.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	0
11.05	Age and gender match exactly, MBRSEL missing for all roster members	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	2
11.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	0
11.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	0
12.01	Age and gender match exactly, exactly one MBRSEL in right place	Everything missing	14
12.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Everything missing	2
12.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Everything missing	1
12.05	Age and gender match exactly, MBRSEL missing for all roster members	Everything missing	0
12.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Everything missing	0
12.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Everything missing	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
13.01	Age and gender match exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	76
13.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	6
13.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	2
13.05	Age and gender match exactly, MBRSEL missing for all roster members	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	2
13.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	0
13.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	0
15.01	Age and gender match exactly, exactly one MBRSEL in right place	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0
15.03	Age within one, gender matches exactly, exactly one MBRSEL in right place		0
15.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0
15.05	Age and gender match exactly, MBRSEL missing for all roster members	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
15.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0
15.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0
16.01	Age and gender match exactly, exactly one MBRSEL in right place	Multiple matches on gender and age, MBRSEL missing for all roster members, pick one where relationship code matches	1
18.01	Age and gender match exactly, exactly one MBRSEL in right place	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.05	Age and gender match exactly, MBRSEL missing for all roster members	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
19.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	5
19.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	**Pair Member B**	**Count**
19.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	0
19.05	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender off, one MBRSEL, only two in household	0
19.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	0
19.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age within one, gender off, one MBRSEL, only two in household	0
19.11	Age within one, gender off, one MBRSEL, only two in household	Age within one, gender off, one MBRSEL, only two in household	0
20.00	No match, but no relationship codes are missing, and none involve domains of interest	No match, but no relationship codes are missing, and none involve domains of interest	22
21.01	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.05	Age and gender match exactly, MBRSEL missing for all roster members	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age matches exactly, gender off, MBRSEL missing for all roster members	0
22.01	Age and gender match exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	64

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	**Pair Member B**	**Count**
22.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	7
22.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	7
22.05	Age and gender match exactly, MBRSEL missing for all roster members	No match at all (often paired respondent is missing from roster)	0
22.06	Age matches exactly, gender off, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	0
22.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	No match at all (often paired respondent is missing from roster)	1
25.00	No match at all	No match at all	18

H.2 Rules for Identifying Pair Relationships among Pairs

Table H.2 summarizes the rules used to identify the pair relationships, using the relationship codes and questionnaire ages of the two pair members. Because the child (12 to 17)-parent and child (12 to 20)-parent relationships can be derived from relationships created using 12- to 14-year-olds, 15- to 17-year-olds, and 18- to 20-year-olds, these latter relationships are the ones referenced in the rules. The variable PAIRREL, which is the next to last column of the table, identifies the pair relationship as defined by Table 11.3 in the main body of this report. As with the rules for identifying which members of the roster belong to the pair, these rules—also called priority conditions because of their hierarchical nature—are shown in priority order. In the headers, the moniker "A" refers to pair member A, and "B" refers to pair member B. The relationship between A and B is described in the columns "A-B Relationship," from the perspective of pair member A ("B to A, According to A") and the perspective of pair member B ("A to B, According to B"). Any constraints on the pair members (other than FIPE3) are provided in the columns "Constraint on A" and "Constraint on B." These constraints include age constraints, where a range of ages (e.g., 12 to 17) indicates that the value of the questionnaire edited age (AGE) is between the numbers shown. Also in this column, "child" and "children" are defined as (a) roster member(s) with nonmissing ages less than 18. The question FIPE3 asks if the respondent is the parent of a selected 12- to 17-year-old. The responses provided in the table are either "yes" or "no." The column for RELMATCH indicates the quality of the match between pair members, as defined in Table 11.6 in the main body of this report. In the table, blank cells mean that no restrictions were placed on that variable to determine the pair relationship.

Table H.2 Rules for Identifying Pair Relationships among Pairs

Priority Condition	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, According to A	A to B, According to B						
1	Parent	Child	12-14				1	1
	Child	Parent		12-14				
2	Parent	Child	15-17				2	1
	Child	Parent		15-17				
3	Parent	Child	18-20				3	1
	Child	Parent		18-20				
4	Parent	Child	21+				4	1
	Child	Parent		21+				
5	Sibling	Sibling	12-14	15-17			5	1
	Sibling	Sibling	15-17	12-14				
6	Sibling	Sibling	12-17	18-25			6	1
	Sibling	Sibling	18-25	12-17				
7	Sibling	Sibling	No constraints, after considering #5 & #6					1
8	Spouse/partner	Spouse/partner	≥ 1 child	≥ 1 child			8	1
9	Spouse/partner	Spouse/partner	0 children, no bad data	0 children, no bad data			9	1
10	Spouse/partner	Spouse/partner	≥ 1 child	0 children, some bad data			8	1.5
	Spouse/partner	Spouse/partner	0 children, some bad data	≥ 1 child				
11	Spouse/partner	Roommate/nonrelative	≥ 1 child both sides, equal number each side					3
	Roommate/nonrelative	Spouse/partner	≥ 1 child both sides, equal number each side					

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, According to A	A-B Relationship: A to B, According to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
12	Partner	Partner	≥ 1 child	0 children, but other's children in household			8	3
	Partner	Partner	0 children, but other's children in household	≥ 1 child				
13	Spouse/partner	Spouse/partner	No constraints, after considering #8-12					1
14	Grandchild	Grandparent					11	1
	Grandparent	Grandchild						
15	Parent-in-law	Child-in-law					12	1
	Child-in-law	Parent-in-law						
	Other relative	Other relative						
	Roommate/boarder/ nonrelative	Roommate/boarder/ nonrelative						
16	Roommate/boarder/ other relative/ nonrelative/in-laws	Roommate/boarder/ other relative/ nonrelative/in-laws					13	1
17	Parent	Missing	12-14	B less than 10 yrs. older th. A		No	14	0
	Missing	Child						
18	Parent	Missing	12-14				1	2
	Missing	Child						
19	Child	Missing	A less than 10 yrs. older th. B.	12-14	No		14	0
	Missing	Parent						
20	Child	Missing		12-14			1	2
	Missing	Parent						
21	Parent	Missing	15-17	B less than 10 yrs. older th. A		No	14	0
	Missing	Child						
22	Parent	Missing	15-17				2	2
	Missing	Child						

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, According to A	A-B Relationship: A to B, According to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
23	Child	Missing	A less than 10 yrs. older th. B	15-17	No		14	0
	Missing	Parent						
24	Child	Missing		15-17			2	2
	Missing	Parent						
25	Parent	Missing	18-20				3	2
	Missing	Parent		18-20				
26	Child	Missing		18-20			3	2
	Missing	Child	18-20					
27	Parent	Missing	21+				4	2
	Missing	Parent		21+				
28	Child	Missing		21+			4	2
	Missing	Child	21+					
29	Sibling	Missing	12-14	15-17			5	2
			15-17	12-14				
	Missing	Sibling	12-14	15-17				
			15-17	12-14				
30	Sibling	Missing	12-17	18-25			6	2
			18-25	12-17				
	Missing	Sibling	12-17	18-25				
			18-25	12-17				
31	Sibling	Missing	No constraints, after considering #24, #25					2
	Missing	Sibling	No constraints, after considering #24, #25					
32	Spouse/partner	Missing	≥ 1 child	No spouse in roster			8	2
	Missing	Spouse/partner	No spouse in roster	≥ 1 child				

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, According to A	A-B Relationship: A to B, According to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
33	Spouse/partner	Missing	0 children, no bad data	No spouse in roster			9	2
	Missing	Spouse/partner	No spouse in roster	0 children, no bad data				
34	Spouse/partner	Missing	After #27, #28, no constraints	No spouse in roster			10	2
	Missing	Spouse/partner	No spouse in roster	After #27, #28, no constraints				
35	Grandchild	Missing	A at least 20 years older than B					2
	Missing	Grandparent						
	Grandparent	Missing	B at least 20 years older than A					
	Missing	Grandchild						
36	Roommate/boarder/ other relative/ nonrelative/in-laws	Missing			No		12	2
	Missing	Roommate/boarder/ other relative/ nonrelative/in-laws			No			
37	Roommate/boarder/ other relative/ nonrelative/in-laws	Missing					13	2
	Missing	Roommate/boarder/ other relative/ nonrelative/in-laws						
38	Nonmissing	Child	12-14			Yes	1	3
39	Nonmissing	Parent		12-14	Yes		1	3
40	Child	Nonmissing		12-14	Yes		1	3
41	Parent	Nonmissing	12-14			Yes	1	3
42	Nonmissing	Child	15-17			Yes	2	3

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, According to A	A-B Relationship: A to B, According to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
43	Nonmissing	Parent		15-17	Yes		2	3
44	Child	Nonmissing		15-17	Yes		2	3
45	Parent	Nonmissing	15-17			Yes	2	3
46	Parent	Roommate/boarder/ other relative/ nonrelative	12-14			No	13	3
						Missing	15	4
	Roommate/boarder/ other relative/ nonrelative	Parent		12-14	No		13	3
					Missing		15	4
47	Parent	Roommate/boarder/ other relative/ nonrelative	15-17			No	13	3
						Missing	16	4
	Roommate/boarder/ other relative/ nonrelative	Parent		15-17	No		13	3
					Missing		16	4
48	Parent	Roommate/boarder/ other relative/ nonrelative	18-20				17	4
	Roommate/boarder/ other relative/ nonrelative	Parent		18-20			17	4
49	Parent	Roommate/boarder/ other relative/ nonrelative	21+				18	4
	Roommate/boarder/ other relative/ nonrelative	Parent		21+			18	4

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, According to A	A-B Relationship: A to B, According to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
50	Nonmissing, not sibling	Child	12-14	21-75		No	13	3
			12-14, exactly one parent	21-75, exactly one spouse		Missing	1	3
			12-14, 0 or 2 parents, or B has 0 or 2 spouse	21-75, 0 or 2 spouses, or A has 0 or 2 parents		Missing	15	4
	Child	Nonmissing, not sibling	21-75	12-14	No		13	3
			21-75, exactly one spouse	12-14, exactly one parent	Missing		1	3
			21-75, 0 or 2 spouses, or A has 0 or 2 parents	12-14, 0 or 2 parents, or B has 0 or 2 spouse	Missing		15	4
51	Nonmissing, not sibling	Child	15-17	24-75		No	13	3
			15-17, exactly one parent	24-75, exactly one spouse		Missing	2	3
			15-17, 0 or 2 parents, or B has 0 or 2 spouse	24-75, 0 or 2 spouses, or A has 0 or 2 parents		Missing	16	4
	Child	Nonmissing, not sibling	24-75	15-17	No		13	3
			24-75, exactly one spouse	15-17, exactly one parent	Missing		2	3
			24-75, 0 or 2 spouses, or A has 0 or 2 parents	15-17, 0 or 2 parents, or B has 0 or 2 spouse	Missing		16	4

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, According to A	A-B Relationship: A to B, According to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
52	Nonmissing, not sibling	Child	18-20, exactly one parent	27-75, exactly one spouse		Missing	3	3
			18-20, 0 or 2 parents, or B has 0 or 2 spouse	27-75, 0 or 2 spouses, or A has 0 or 2 parents		Missing	17	4
	Child	Nonmissing, not sibling	27-75, exactly one spouse	18-20, exactly one parent	Missing		3	3
			27-75, 0 or 2 spouses, or A has 0 or 2 parents	18-20, 0 or 2 parents, or B has 0 or 2 spouse	Missing		17	3
53	Nonmissing, not sibling	Child	21+, exactly one parent	27-75, exactly one spouse		Missing	4	4
			21+, 0 or 2 parents, or B has 0 or 2 spouse	27-75, 0 or 2 spouses, or A has 0 or 2 parents		Missing	18	3
	Child	Nonmissing, not sibling	27-75, exactly one spouse	21+, exactly one parent	Missing		4	3
			27-75, 0 or 2 spouses, or A has 0 or 2 parents	21+, 0 or 2 parents, or B has 0 or 2 spouse	Missing		18	4
54	Spouse	Sibling	One is 12-14, other is 15-17, both sides have parents or spouses					3
	Sibling	Spouse						
55	Spouse	Sibling	One is 12-17, other is 18-25, both sides have parents or spouses					3
	Sibling	Spouse						

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, According to A	A-B Relationship: A to B, According to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
56	Spouse	Sibling	Ages neither 12-14/15-17 nor 12-17/18-25, both sides have parents or spouses				7	3
	Sibling	Spouse						
57	Other relative	Sibling	Both sides have 2 parents, ages of oldest parents on either side differ by > 5 years, age of youngest parents on either side differ by > 5 years				13	3
	Sibling	Other relative						
58	Nonmissing, not child	Sibling	15-17				19	4
	Sibling	Nonmissing, not child	12-14					
59	Nonmissing, not parent	Sibling	12-14				19	4
	Sibling	Nonmissing, not parent	15-17					
60	Nonmissing, not child	Sibling	18-25				20	4
	Sibling	Nonmissing, not child	12-17					
61	Nonmissing, not parent	Sibling	12-17				20	4
	Sibling	Nonmissing, not parent	18-25					
62	Nonmissing, not child	Sibling	Ages neither 12-14/15-17 nor 12-17/18-25, A older than B				21	4
	Sibling	Nonmissing, not child	Ages neither 12-14/15-17 nor 12-17/18-25, B older than A				21	4
63	Nonmissing, not parent	Sibling	Ages neither 12-14/15-17 nor 12-17/18-25, B older than A				21	4
	Sibling	Nonmissing, not parent	Ages neither 12-14/15-17 nor 12-17/18-25, A older than B				21	4

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, According to A	A-B Relationship: A to B, According to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
64	Sibling	Roommate, in-law, grandparent, grandchild, boarder, other relative, nonrelative	At least one is between 18 and 20				13	3
	Roommate, in-law, grandparent, grandchild, boarder, other relative, nonrelative	Sibling	At least one is between 18 and 20					
65	Sibling	Unusual in-law code	12-20	26 or older			13	3
	Unusual in-law code	Sibling	26 or older	12-20				
66	Spouse/partner	Not a child, parent, or sibling	≥ 1 child aged < 18	No spouse			22	4
	Not a child, parent, or sibling	Spouse/partner	no spouse	≥ 1 child aged < 18				
67	Spouse/partner	Not a child, parent, or sibling	15 or older, 0 children, no bad data	15 or older, no spouse			23	4
	Not a child, parent, or sibling	Spouse/partner	15 or older, no spouse	15 or older, 0 children, no bad data				
68	Grandparent, grandchild	Not grandparent, not grandchild					25	4
	Not grandparent, not grandchild	Grandparent, grandchild						
69	Any codes	Any codes	No constraints	No constraints			14	0

Appendix I: Conditions Used to Reconcile Differing Multiplicity Counts between Pair Members

This page intentionally left blank

Appendix I: Conditions Used to Reconcile Differing Multiplicity Counts between Pair Members

I.1 Introduction

In order to determine multiplicity counts, counts were obtained from each pair member. The count from the pair member who was the focus member of the domain was considered the direct count, and the count from the other pair member was considered the indirect count. Typically, these counts were in agreement, and the determination of the final multiplicity count was straightforward, provided neither roster had bad data codes. The strategy also was usually clear if one pair member had bad data in the household roster, or had a 0 count when the pair relationship precluded a value of 0. The count from the pair member with good, nonzero data was usually preferred in those cases. If the bad data was limited to bad relationship codes, then the member with good data was selected only if substituting the appropriate relationship codes for the bad data codes would have given a total that was equal to the count from the pair member with good data. There were instances where bad data codes existed in the roster, and this condition did not apply. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts still disagreed. The rules that were used to reconcile these disagreeing counts are outlined in this appendix.

Note that the reconciliation of differing counts was necessary for parent-child and sibling-sibling pairs but was not necessary for spouse-spouse pairs, since the multiplicity count for spouse-spouse pairs was always 1. As noted in Section 11.3.1, it was technically possible for a respondent to have multiple spouses, but these situations were not accounted for.

I.2 Parent-Child Counts

For parent-child counts, the screener and the FIPE3 variable were used to help reconcile disagreeing counts. The rules follow below, separated by the member of focus:

Parent-child pairs, child focus. The multiplicity counts in this domain reflected the selected child's parents and in most cases had values of 1 or 2. If neither side had bad relationship codes, and if the direct count exceeded the indirect count, the following rules applied:

1. The direct count might have exceeded the indirect count because one parent had left or entered the household between interviews. In this case, the ages in the rosters were matched to the screener roster to determine which count to believe. This was done in two ways. First, the total number of roster members between ages 30 and 39, 35 and 44, and 40 and 49 were compared between pair members and the screener. The pair member with age range counts closest to the screener was the one whose parent-child count was chosen for the final count. If neither side had age range counts equal to the screener, then the pair member with a parent-child count equal to the total number of screener roster members between ages 26 and 64 was chosen as the final count.

2. The direct count might have exceeded the indirect count because the selected parent did not consider the other "parent" a spouse or live-in partner. If the pair relationship was not imputed, the indirect count was selected. However, if the pair relationship was imputed and the older pair member called the younger pair member a child, then the older pair member considered the child's "true" parent as not a spouse or live-in partner, even though he or she claimed the "true" parent's children. In this case, the direct count was used (the child's adjusted count).

If the direct count was exceeded by the indirect count, then the child listed only one parent, and the parent listed a spouse (a "stepparent") or live-in partner in the household roster. The following rules applied:

1. The indirect count might have exceeded the direct count because the selected child did not accept a stepparent or live-in partner as his or her parent. If this stepparent or live-in partner was the other respondent selected, this was considered a parent-child pair based on the response of the "parent" to the FIPE3 question. If the FIPE3 question was answered "yes," and the RELMATCH variable had a value of 3, then the indirect count was selected as the multiplicity count. If the FIPE3 question was answered "no," the pair was not considered a parent-child pair and was not considered for these counts. Finally, if the FIPE3 question was not answered, the respondent was considered a "parent" if he or she was a stepparent. If the respondent was a live-in partner, the determination of the pair relationship was left to imputation. The multiplicity count was set to the indirect count to account for the possibility that the pair relationship would be imputed as parent-child.
2. Suppose the selected child did not accept a stepparent or live-in partner as his or her parent (as above), but the other respondent selected was the "true" or "original" parent. In this case, the stepparent or live-in partner was identified only in the "original" parent's roster, so there was no way to determine how the stepparent or live-in partner would have answered the FIPE3 question. A stepparent was considered a "parent" even if the child did not view him or her this way so that the indirect count was used. The case of live-in partners was less clear, and these cases were left to imputation.
3. If age range counts between the two pair members and the screener matched across a variety of age ranges (30 to 39, 40 to 49, and 50 to 59), but the child's roster had a bad relationship code among roster members of potential parent age (15 or older), or the child's roster had a value of MBRSEL that did not match what was finally determined to be the child's parent, then the multiplicity count for the parent—the indirect count—was selected as the final count.

Parent-child pairs, parent focus. The multiplicity counts in this domain reflected the selected parent's children and were limited to have values of at least 1. If neither side had bad relationship codes, the following rules applied:

1. If the count of children in the household within the relevant age ranges differed between the pair members, but one side had a count of children equal to the same

count from the screener roster, then the multiplicity count that corresponded to the pair member with the same count of children as the screener was used.

2. If the count of children in the household within the relevant age ranges differed between the pair members, and both sides had a multiplicity count that exceeded the count of all children from the screener roster, then the number of children in the screener roster was used as the multiplicity count. If the screener roster had missing exact ages, then the minimum multiplicity count from the two pair members' rosters was used as the final count.
3. The direct count and indirect count might differ because either the child listed a sibling that the parent considered "another relative" or the parent listed a child that the child considered "another relative." In either case, the parent was the one to answer the FIPE3 question. Because of this, the multiplicity count from the parent's perspective was selected as the final count, provided that the counts of children in the household within the relevant age ranges for each pair member were equal.
4. After considering the above situations, the multiplicity counts might have still differed without a resolution of which count should have been chosen as the final count. This occurred because the counts of children in the household differed between pair members, each of which differed from the screener count. Moreover, multiplicity counts did not exceed the screener age range count. In this instance, if one of the multiplicity counts equalled the screener age range count, then this multiplicity was selected as the final count. However, if this was not the case, then upper and lower bounds were created and the final multiplicity was left to imputation.

Because of the hierarchical nature of these counts, parent-child counts for 12- to 17-year-old and 12- to 20-year-old children could sometimes be derived if the 12- to 14-year-old parent-child count was already determined for both child focus and parent focus counts. In particular, if one pair member's count for 12- to 17-year-old children or 12- to 20-year-old children equalled or exceeded the final parent-child count for 12- to 14-year-old children and the other did not, then the pair member's count that equalled or exceeded the 12- to 14-year-old count was chosen as the final count.

I.3 Sibling-Sibling Counts

Although there were two types of sibling-sibling pairs under consideration, each associated with two domains, the same rules could be applied to all four domains. When the older sibling was the focus, the multiplicity count was a count of the number of siblings within the younger age group (12 to 14 or 12 to 17). Conversely, the multiplicity count was the number of siblings in the older age group (15 to 17 or 18 to 25) when the younger sibling was the focus. Deciding how to assign a final multiplicity count often involved looking at a count of household members within the age range of the siblings being counted. For example, if the older sibling was the focus and the age ranges were 12 to 14 and 15 to 17, the number of household members aged 12 to 14 were counted. The following general rules applied if the multiplicity counts for each pair member disagreed:

1. The counts disagreed if a household member left or entered the household between interviews. As before, the roster that was closest to the screener was used to

determine the count. In particular, depending upon the domain, the count of household members within the age range of the siblings being counted was compared between each pair member and the screener. The multiplicity count from the pair member with the count closest to the screener was used, provided that the member had no bad relationship codes within the relevant age range.

2. If the counts of household members within the age range of the siblings being counted differed between pair members and those counts were both exceeded by the screener count, then the multiplicity associated with the pair member with the age range count closest to the screener was chosen, provided that the member had no bad relationship codes within the relevant age range.

3. In some cases, the counts of household members within the age range of the siblings being counted were the same for the two pair members, but the multiplicity counts disagreed.

    a. If one pair member had bad relationship codes and the other did not, the disagreement could have been due to the bad relationship codes. If the sums of the multiplicity count and the number of bad relationship codes were equal across pair members, then the final count was set to equal the multiplicity of the pair member who did not have bad relationship codes.

    b. If one pair member identified the other as "sibling" but the other pair member did not reciprocate, then imputation was required to establish whether the relationship was sibling-sibling. The count associated with the pair member who indicated that the other pair member was a sibling should have been chosen as the final count. In effect, this was done by taking the maximum of the two pair members' counts.

4. If the counts of household members within the age range of the siblings being counted disagreed and both exceeded the screener count of household members within the relevant age range, then the multiplicity count was set to the screener count. If the screener roster had missing exact ages, then the minimum multiplicity count from the two pair members' rosters was used as the final count.

5. If the differing multiplicity counts could not be reconciled with the above rules, upper and lower bounds for the true multiplicity were determined using the two multiplicity counts, as well as the counts of children within relevant age ranges in both pair member's rosters and the screener roster. In rare cases, the values for these bounds were equal. These cases were investigated, and if the reasons were legitimate, then the final multiplicity count was set to this value. Otherwise, the final multiplicity was left to imputation.

Appendix J: Conditions Used to Reconcile Differing Household-Level Person Counts between Pair Members

This page intentionally left blank

Appendix J: Conditions Used to Reconcile Differing Household-Level Person Counts between Pair Members

J.1 Introduction

Household-level person counts for a particular domain were obtainable using the multiplicity counts if the pair belonged to a pair relationship that fit into that domain, provided only one family unit was in the household. No reconciliation between pair members was necessary in that case, since the reconciliation had already been done with the multiplicity counts. Other counts were obtained from single respondents for whom no reconciliation was necessary. This appendix discusses the conditions used to reconcile differing household-level person counts when the pair belonged to a pair relationship that corresponded to different pair domains than the one being counted. Typically, the counts between the two pair members were in agreement, and the determination of the final household-level count did not involve a reconciliation of counts, though assigning a final count meant ensuring that pair relationships were not hidden due to the relationships of the two pair members to other household members.[1] A similar situation occurred if one pair member had bad data in the household roster. The count from the pair member with good data was usually preferred in those cases, provided pair relationships of interest were not hidden. If bad data existed in either household roster, but the bad data was limited to bad relationship codes, then the member with good data was selected only if substituting the appropriate relationship codes for the bad data codes would have given a total that was equal to the count from the pair member with good data. There were instances where bad data codes existed in the roster, and this condition did not apply. There were other exceptions as well. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts still disagreed. In this appendix, the rules that were used to assign a final count, as well as to reconcile disagreeing counts, are outlined. For each pair domain, a set of general rules are given, each with specific conditions required for the general rule to be implemented. Within each general condition, if at least one of the specific conditions was not satisfied, upper and lower bounds were determined and the final count was left to imputation.

J.2 Parent-Child Counts

For parent-child counts where the pairs were not parent-child pairs of interest (e.g., sibling-sibling pairs, parent-child pairs where the child was 21 or older, etc.), the screener was used to help reconcile disagreeing counts. The rules follow below, separated by the member of focus:

Parent-child pairs, child focus. For the child-focus counts, the count is of the number of children of a parent in the household. The following general rules applied:

1. In most cases, the counts of children in the relevant age range with parent(s) in the household (abbreviated below as children with parent(s) in the household) for the two

[1] If a roster pointed to a household size of one, this was considered "bad data" since both pair members in the household were survey respondents.

sides agreed. However, both sides had to meet the following conditions in order for the final count to be set to one of the sides:

- Either no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or the counts of children with parent(s) in the household were equal to the screener age counts, or a side with good data indicated siblings within the relevant age range living together in a household without parents;
- No situations where parents were not identified in the household, but some in the household had bad relationship codes and were old enough to be parents;
- No counts of one child in the relevant child-age range when both members of the pair were in that range and the children were siblings;
- No pairs where the ages of the identified parents did not match, the pair members were not siblings, and both sides had relationship codes signifying "other relative" or a nonrelative, indicating more than one family unit in the household;[2] and
- The household size was greater than 1 and was nonmissing on both sides.

2. The counts of children with parent(s) in the household might have agreed even though the above conditions were not met. The final count of children with parent(s) in the household could still have been set to one of the sides, if any one of the following was true:
   - If the number of children within the relevant age ranges matched across both rosters and the screener and (at least) one side had all good age and relationship codes, provided the equal counts did not refer to different children;[3]
   - If both sides had a count of zero children with parent(s) in the household, both had a roster, and (at least) one side had all good age and relationship codes;
   - If both sides had a count of zero children with parent(s) in the household, both had a roster, and the number of respondents who were old enough to be parents in the household was zero according to the screener; or
   - If the counts of children with parent(s) in the household that agreed with each other equalled or exceeded the count of the number of children from the screener within the relevant age ranges.
3. The counts of children with parent(s) in the household might have agreed with a value of 1. If both pair members were children within the relevant age range, and both indicated they had parents even though the children were siblings, then they were not included in each other's rosters, but they were obviously in the screener roster, so the final count of children with parent(s) in the household was set to 2.

---

[2] Codes that indicate "other relative" or a nonrelative are 7 (roommate), 8 (child-in-law), 10 (parent-in-law), 12 (boarder), 13 (other relative), and 14 (other nonrelative).

[3] This was determined by excluding situations where the ages of the identified parents did not match, the pair members were not siblings, and both sides had relationship codes signifying "other relative" or a nonrelative, indicating more than one family unit in the household.

4. If one pair member did not have a valid roster but the other member did, the final count of children with parent(s) in the household was set to the other pair member's count under the following conditions:
   - No counts of one child with parent(s) in the household when both members of the pair were children in the relevant age range and the children were siblings, and
   - Either:
     - There were no bad relationship codes within the relevant child-age ranges and the respondent identified parents or children in the household,
     - There were no children within the relevant age range, or
     - No parents were identified in the household and nobody in the roster older than the respondent had a bad relationship code.
5. If one pair member did not have a valid roster but the other member did, and the above conditions were not met, it was still possible to use the other pair member's count of children with parent(s) in the household, if that count was 0, under any of the following conditions. Either:
   - The other roster was valid, did not have any bad ages, and had no ages in the relevant age range;
   - The other roster also was bad but the screener roster was valid and did not have any ages in the relevant age range; or
   - The respondent identified both grandchildren and grandparents in the roster where the "grandchild" relationship code(s) were incorrectly entered into the respondent's household roster. The "grandchildren" that these relationship codes were referring to were not the respondent's grandchildren, but, rather, they were the respondent's grandparent's grandchildren.[4]
6. When two different family units were in the household, the determination of the final count of children with parent(s) in the household had to be treated separately. This could have included the multigenerational families referred to earlier and the two siblings both with children in the relevant age range living in the household. The latter was more easily identified if it was not a parent-child pair (e.g., a cousin-cousin pair). The sum of the two counts of children with parent(s) in the household (one count might be 0) was used as the final count, provided the following conditions were satisfied on both sides:
   - There were no bad ages or relationship codes within the relevant age ranges;
   - Both had counts of children with parent(s) in the household pointing to two or fewer parents, meaning that the two family units were not identifiable on one side;

[4] This condition has not manifested itself since the 2001 survey. With the addition of a new consistency check added since the 2001 survey to address grandparent/grandchild code inconsistencies, this condition could be observed only if a respondent overrode this consistency check, which has not happened.

- The number of identified parents was not equal to the total number of household members older than 25 in the household on either side, meaning that parents could correspond to roster members identified by other relationship codes;
- The number of identified children was not equal to the total number within the relevant age range in the household on either side, meaning that children with parents could correspond to roster members identified by other relationship codes; and
- There were not three generations in the household with first and second generation parents both having children in the appropriate age range. This was already accounted for by the counts for one or both sides.

If the pair was a parent-child pair, the final count was determined using imputation.

7. Two family units might be in the household but the conditions given in item #6 were not met. If there were no bad ages or relationship codes within the relevant age ranges (for both children and parents), the two families in the household might have been already accounted for when the counts of children with parent(s) in the household were determined for each side. The maximum of the two counts was used as the final count if the household members in the roster older than 25 (of parental age) were either both equal to the number of household members older than 25 in the screener roster or both different than the number of members older than 25 in the screener roster. However, if the number of household members older than 25 in the screener roster was equal to the number of members older than 25 in one of the pair member's rosters but not the other, then the count of children with parent(s) in the household corresponding to the pair member with a roster matching the screener roster (among household members of potential parental age) was used as the final count of children with parent(s) in the household.
8. If one pair member did not have a valid roster and the pair member with a valid roster was within the valid age range and was a sibling to the other pair member, but the count of children with parent(s) in the household from his roster was only 1, then the final count was set to 2.
9. If the pair relationship was not parent-child nor was it sibling-sibling, but one side had nonzero counts of children with parent(s) in the household and the other did not, it was necessary to decide who to believe. This occurred often because one of the respondents was a relative outside the nuclear family unit—like a cousin or aunt/uncle—whose own parents did not live in the household, or the respondent was a boarder.[5] Selecting either the zero count or nonzero count in this instance required that the following conditions were met:
   - The respondent with a zero count of children with parent(s) in the household did not identify parents in the roster or he or she identified parents but was older than 20 and had no bad relationship codes within the relevant age ranges, and

[5] Even if there was disagreement between the respondents about whether a boarder or other family member was in fact a sibling, parent, or child, this would have been resolved at the pair relationship stage where it would have been determined whether this was in a domain of interest.

- Either the respondent with a nonzero count of children with parent(s) in the household had siblings or children within the relevant age range, or the respondent himself or herself was within that age range (with a count of 1).

When one count of children with parent(s) in the household was zero and the other was nonzero, the nonzero count was used under the following conditions:

- The respondent pair member with a nonzero count also did not have bad relationship codes within the relevant age ranges, and
- Either:
  - The count of children within the relevant age range in the household for the nonzero count pair member matched that of the zero count pair member, and the count of children with parent(s) in the household did not exceed the screener count of children within the relevant age range;
  - The count of children in the household within the relevant age range for the nonzero count pair member matched that of the screener;
  - The count of children in the household within the relevant age range for the zero count pair member matched that of the screener because a child was (or children were) listed as 11 years old in the nonzero count pair member's roster, when he or she (they) should have been 12 (according to the zero count pair member's and the screener roster) so that the final count was the nonzero count with this child (these children) added;
  - The respondent with a zero count had no household members with a family-type relationship code and the reported household sizes of the two pair members were equal (indicating that it was unlikely that anyone had entered or left the household between interviews);
  - The respondent with a nonzero count showed a parent-child relationship existed in the household, but the respondent with a zero count did not because he was not related to the other household members. However, the count of children within the relevant age range in the household for the zero count was closer to the screener age count. Nevertheless, the nonzero count was equal to or less than the screener age count; or
  - The other conditions had not already established a nonzero count, but a count for a subset age group had already been established as nonzero. For example, if the count for 12- to 14-year-olds was nonzero, then the 12- to 17-year-old count had to be nonzero.

The zero count of children with parent(s) in the household was used if the zero-count respondent had no bad relationship codes at all, and either:

- The household age composition among the relevant age ranges for the zero count pair member more closely matched the screener, or
- The pair was a grandparent-grandchild pair with an adult child of the grandparent living in the household. The nonzero count resulted from an assumption that a

respondent's adult child and grandchild within the relevant age range were a parent-child pair. If the grandchild identified the grandparent's child as "other relative" and did not identify any parents, this indicated that the grandparent's adult child was an uncle/aunt of the grandchild, not a parent.

10. If the pair relationship was not parent-child nor was it sibling-sibling, but one side had nonzero counts of children with parent(s) in the household and the other did not, taking the side that was closest to the screener sometimes meant that the count of children with parent(s) from neither pair member was chosen. As with the previous item, a zero count and a nonzero count often occurred because one of the respondents was a relative outside the nuclear family unit—like a cousin or aunt/uncle—whose own parents did not live in the household, or the respondent was a boarder. If neither the zero count nor the nonzero count was chosen, the final count could still have been determined using either the screener count, the count of children within the relevant age range for the respondent with a zero count, or one less than the nonzero count. One of these was chosen, provided that the following conditions were met:

    - The respondent with a zero count of children with parent(s) in the household did not identify parents in the roster or he or she identified parents but was older than 20 and had no bad relationship codes within the relevant age ranges, and
    - Either the respondent with a nonzero count of children with parent(s) in the household had siblings or children within the relevant age range, or the respondent himself or herself was within that age range (with a count of 1).

    The screener count was chosen if either:

    - The respondent pair member with a nonzero count also did not have bad relationship codes within the relevant age ranges. The count of children within the relevant age range in the household for the nonzero count pair member matched that of the zero count pair member, and the nonzero count exceeded the screener count of children within the relevant age range.
    - The respondent with a nonzero count showed a parent-child relationship existed in the household, but the respondent with a zero count did not because he was not related to the other household members. However, the count of children within the relevant age range in the household for the respondent with the zero count was closer to the screener age count, and the nonzero count exceeded the screener count of children.

    In situations where a respondent with a zero count had a roster more closely resembling that of the screener, but the screener included a household member within the relevant age range who was not part of the immediate family, neither the nonzero count of children with parent(s) in the household nor the screener count of children within the relevant age range could be used—a different count had to be used. Two strategies were employed:

    - For the respondent with a nonzero count of children with parent(s) in the household, the nonzero count was the same as the count of children within the relevant age range in the household, but it exceeded the number-of-children count for the zero-count respondent. However, the count of children within the relevant age range for the zero-count respondent, which was not zero, was closer to the screener age count than the nonzero-count respondent.

- If the count of children within the relevant age range for the zero-count respondent was the same as the nonzero count of children with parent(s) in the household, the number-of-children count for the zero-count respondent could not be used, since the nonzero count included a household member that was not in the appropriate age range at the time of screening. One less than the nonzero count of children with parent(s) in the household was therefore chosen as the final count.

11. Other situations with a zero and nonzero count did not necessarily mean that the relationship was something other than parent-child or sibling-sibling. This was usually due to one pair member having missing relationship codes for the roster member that would have been identified as a parent (i.e., relationship codes for roster members in a parental age range). If the count for the pair member with the entirely good roster was equal to the number within the relevant child age range for the pair member with bad relationship codes in the roster, the nonzero count was selected.

12. The two counts of children with parent(s) in the household might have disagreed where both were nonzero and both exceeded the screener count of children within the relevant age range. For the screener count to be chosen as the final household count of children with parent(s) in the household, the following conditions had to be met:
    - The pair member's household rosters had to have different numbers of children within the relevant age range,
    - The pair relationship could be neither parent-child nor sibling-sibling with a zero screener count of children within the relevant age range(s),
    - The total number within the screener roster (where the minimum age was 12 years) had to be at least two, and
    - The number of children in the screener roster within the relevant age range was valid and at least as large as the final count of children with parents in the household for the next smallest age range.

13. The two counts might have disagreed because one side had bad relationship codes within the relevant age range and the other did not. If the sum of the number of bad relationship codes with the smaller count equalled the larger count, the larger count was chosen.

14. The two counts might have disagreed because they disagreed on the ages of one or more household members, even though each respondent's count included all the children in their respective roster. If the roster for one respondent more closely matched the screener in terms of the distribution of ages within the roster, then that respondent's count was chosen.

15. The two counts might have disagreed because they disagreed on the ages of one or more household members and each respondent's count included all the children in their respective roster, but neither was closer to the screener count. If the screener

count differed from each respondent's count by the same amount, was greater than 1 but less than the other, then the screener count was used as the final count.

16. If the pair relationship was parent-child and the parent-child counts were associated with the same age range, then the household-level person counts were obtained using the parent-focus multiplicity counts corresponding to the appropriate age range. However, this did not occur if the age range for the pair relationship differed from the age range for the parent-child counts. If the pair relationship was imputed to be parent-child or it was deemed parent-child even though the child did not consider the parent a "parent," but the parent answered the FIPE3 question, then the nonzero count was used as the final count.
17. If, after all the above tests were done to find the final count, the minimum possible and maximum possible counts—considering both questionnaire rosters and the screener roster—were the same, then the final count was set to that value.
18. Remaining disagreeing counts were left to imputation, with appropriate bounds set on the imputed value.

Parent-child pairs, parent focus. For the parent-focus counts, the count is of the number of parents of at least one child in the household. The child-focus parent-child counts are processed first, so if the child-focus parent-child counts are 0, it necessarily means that the parent-focus counts will also be 0. Nonzero child-focus counts also point to nonzero parent-focus counts. After setting counts to 0 where necessary, the following general rules applied:

1. In most cases, the counts of parents with children in the household for the two sides agreed. However, both sides had to meet the following conditions in order for the final count to be set to one of the sides:
   - No situations where both pair members were children in the relevant age range but were in a spouse-spouse pair relationship and both identified the same roster member as a parent,
   - The household size was greater than 1 and nonmissing on both sides, and
   - Either:
     - No bad relationship codes for household members of an age to be parents,
     - The total count was 2 for two parents, or
     - The total count plus the number of grandparents equalled the total number of household members aged 26 or older, according to the screener roster.

   Note that it was not necessary to check for bad relationship codes in the child age ranges, since it was already known that the count had to be at least 1, and the number of children was not important for the parent counts.

2. The counts of parents with children in the household might have agreed even though the above conditions were not met. The final count could still have been set to one of the sides if it was a sibling-sibling pair, and the bad codes in the parental age range were on one side only. This would indicate that the side with bad codes were not missing parental codes.

3. If one pair member did not have a valid roster but the other member did, the final count of parents with children in the household was set to the other pair member's count if there were no bad relationship codes and no roster members with bad age and bad gender values. Other circumstances called for setting the final count to 0, which would necessarily be the case if the child-focus counts were 0.
4. When two different family units were in the household, the determination of the final count of parents with children in the household had to be treated separately. This could have included multigenerational families or two siblings both with children in the relevant age range living in the household. The latter was more easily identified if it was not a parent-child pair (e.g., a cousin-cousin pair). The sum of the two counts (one count might be 0) was used under the following conditions:
   - There were no bad ages or relationship codes within the relevant age ranges,
   - Both pair members had counts pointing to 2 or fewer parents, meaning that the two family units were not identifiable on a side,
   - The number of identified parents was not equal to the total number of household members older than 25 on either side, meaning that parents could correspond to roster members identified by other relationship codes, and
   - There were not three generations in the household, with first and second generation parents both having children in the appropriate age range. This was already accounted for by the counts for one or both sides.
5. Two family units might be in the household but the conditions given in item #4 were not met. If there were no bad ages or relationship codes within the relevant age ranges (for both children and parents), the two families in the household might have been already accounted for when the counts of parents with children in the household were determined for each side. The maximum of the two counts was used as the final count if the household members older than 25 (of parental age) in the roster were either both equal to the number of members older than 25 in the screener roster or both different than the number of members older than 25 in the screener roster. However, if the number of household members older than 25 in the screener roster was equal to the number of members older than 25 in one of the pair member's rosters but not the other, then the count of parents with children in the household corresponding to the pair member with a roster matching the screener roster (among household members of potential parental age) was used as the final count of children with parent(s) in the household.
6. If the pair relationship was a spouse-spouse pair and one of the pair members was within the relevant age range and had a positive count, then the count for that pair member was taken as the final count, provided there were no bad relationship codes in that roster for roster members aged 18 or older.[6]
7. The two counts might have disagreed with one nonzero count and the other equal to zero. Due to the fact that the counts of parent(s) in the household with children were

[6] For this condition, either the count for the other pair member was 0 or the count for the pair members was equal.

determined first and that the zero counts were handled separately, the final count of parents with children in the household determined at this stage of processing had to be nonzero. Counts arising from two or more families in the household also were handled in previous code. Hence, the final count had to be one or two parents.[7] The nonzero count was chosen as the final count if one of the following conditions were met:

- The count was 1 and there were no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or
- The count was 2.

8. The two counts might have disagreed where the number of roster members aged 26 or older disagreed between the two pair members. In these situations, one count was 1, and the other count was 2. The final count corresponded to the pair member with the number of roster members aged 26 or older closest to the screener number of roster members aged 26 or older, under the following conditions:
   - The difference between the screener count of the number of household members aged 26 or older and the pair members' counts of this number of household members was not the same between the two pair members,
   - Neither pair member had bad ages in their rosters, and
   - Each pair member either had no bad relationship codes in his or her roster or had a nonzero count with no bad relationship codes among respondents aged 26 or older.
9. The two counts might have disagreed if the bad relationship codes referred to missing parental codes. If one side had no bad relationship codes, and the sum of the number of bad relationship codes and the count on the side with the bad codes was equal to the count on the side with no bad relationship codes, then the count from the side with no bad relationship codes was used as the final count.
10. The two counts might have disagreed where one count was 2 and the other was 3. Since households with two family units had already been considered, the maximum number of parents possible was two, so the final count was set to 2.
11. If the pair relationship was parent-child and the parent-child counts were associated with the same age range, then the household-level person counts were obtained using the child-focus multiplicity counts corresponding to the appropriate age range.
12. If, after all the above tests were done to find the final count, the minimum possible and maximum possible counts—considering both questionnaire rosters and the screener roster—were the same, then the final count was set to that value.
13. Remaining disagreeing counts were left to imputation, with appropriate bounds set on the imputed value.

[7] This precluded the extremely unlikely possibility that the pair member with a zero count masked a situation where three parents in a single family unit lived in the household (two biological parents and a stepparent).

J.3 Sibling-Sibling Counts

The logic for the sibling-sibling counts did not depend upon whether the lower age range was 12 to 14 or 12 to 17 or whether the upper age range was 15 to 17 or 18 to 25. It also did not depend upon which pair member was the focus, though for the household-level person counts, the older member focus counts were the only ones considered. Hence, the counts of interest are of roster members in the upper age range. As with the parent-child pairs, the multiplicity counts could be used if the pair relationship was a sibling-sibling pair of interest. However, the counts had to be determined for all other pairs. The rules follow below, separated by the member of focus:

1. In most cases, the counts for the two sides agreed. However, both sides had to meet the following conditions in order for the final count to be set to one of the sides:
   - The pair could not be a sibling-sibling pair, where both respondents were in the upper age range, and could not have a younger sibling in the lower age range, and the count was 1. (This refers to a sibling-sibling pair that would not constitute a domain of interest.)
   - No bad relationship codes in the lower range if the count was 0.
   - Either:
     - No bad relationship codes in the upper range, or
     - The count matched the screener age count.
   - The household size was greater than 1 and nonmissing on both sides.
2. The counts might have agreed even though the above conditions were not met. The count could still have been set to one of the sides if any one of the following conditions was true:
   - If the number of children matched across both rosters and the screener for both the upper and lower age ranges, or
   - If the count was 0 and one of the following two conditions was true:
     - Neither side had bad relationship codes or ages, or
     - The number of household members aged 26 or older in the screener roster was zero.
3. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:
   - No bad relationship codes within the lower age range when the count was 0.
   - Either:
     - There were no bad relationship codes within the upper age range,
     - The count was equal to the screener age count within the upper age range, or
     - The count was 0, and the count of household members in the lower age range was 0.

4. If one pair member did not have a valid roster but the other member did, and the above conditions were not met, it was still possible to use the other pair member's count under the following conditions:
    - The count was 0,
    - The number of children in either the lower or upper age ranges was 0 with no bad ages in the roster.
5. If neither pair member had a valid roster, it was occasionally still possible to assign a final count. If the number of children in the screener roster in either the lower or upper age ranges was zero and the screener roster was valid, then it was not possible for a sibling-sibling pair in the relevant age ranges to be selected and the final count to be set to 0.
6. When two different sets of siblings were in the household, the determination of the final count had to be treated separately. The two sets of siblings refer to siblings where both parents from one set differ from the parents of the other set. The sum of the two counts (one count might be 0) was used, provided the following conditions were satisfied for both pair members:
    - The sum of counts of the number of sibling-sibling pairs equalled or exceeded at least one of the counts of household members in the upper age range for the screener roster or either of the pair member's rosters.
    - There were no bad relationship codes within the upper age ranges.
    - There were no bad relationship codes within the lower age range, or the count was nonzero.
7. If the counts from the two pair members did not agree, the following rules were used to assign the appropriate count, provided no bad relationship codes were evident in either age range on either side. These conditions are hierarchical, in that subsequent conditions require that the previous condition was not met.
    - If the number within the upper age range was the same on both sides, but the number in the lower age range was not, then the side with the number in the lower age range equal to the number in the screener roster within the lower age range was chosen. (In all cases, one side had a zero count and the other did not. This captured situations where it was necessary to discern whether the zero count was due to no children in the lower age range on one side and whether the screener also had no children in that range.)
    - For one pair member, the number of children in either the lower age range or the upper age range did not agree with the number in the screener roster in that range. However, for the other pair member, the number within both age ranges agreed with the screener count. The count was set to the side that agreed with the screener.
    - For both pair members, the numbers within the lower age range were either both zero or both positive. The number within the upper age range did not agree between pair members, but one pair member agreed with the screener. The final count was set to the count for that pair member.

- In the rosters for both pair members and the screener, the numbers within the upper age range for at least one of the three were nonzero but not necessarily equal. The numbers within the lower age range were not equal across any of the three rosters. The pair member with the number of children in the upper age range closest to the screener was selected.

8. If the counts from the two pair members did not agree, but one side had bad relationship codes within the upper age range and the other did not have bad relationship codes, and the sum of the count and the number of bad relationship codes on one side was equal to the count for the pair member with the good roster, then the count for the pair member with the good roster was selected.
9. If the counts from the two pair members did not agree, and the above conditions were not met, in many cases this was due to one of the pair members not being part of the immediate family unit, in which case his or her count was automatically 0. To identify these cases and assign the count to the other pair member, the following conditions had to be satisfied:
   - The pair relationship did not indicate an identifiable family-type relationship (e.g., sibling-sibling, parent-child, spouse-spouse, or grandparent-grandchild relationship).
   - Either:
     - One pair member did not have any relationship codes indicating parent, child, sibling, spouse, grandchild, or grandparent;
     - The other pair member had at least one relationship code indicating a relationship other than parent, child, sibling, spouse, grandchild, or grandparent;
     - For the pair member with family codes, either no bad relationship codes were within both the upper and lower age ranges or no bad relationship codes were within the upper age range, and the count was positive; or
     - There were no bad relationship codes within both the upper and lower age ranges for either pair member.
10. If one pair member had no bad relationship codes within both the upper and lower age ranges, but the other member had some bad codes, then the count associated with the pair member with no bad codes was selected if the count of immediate family members (parent, child, sibling, spouse, grandchild, or grandparent) was the same as the count of household members within both the lower and upper age ranges.
11. If one pair member had a zero count due to having no household members within the upper age range, but the number of household members within that age range was nonzero for both the screener and the other pair member (though not necessarily equal), and the count for the other pair member was equal to the number of household members within the upper age range for that pair member, then a nonzero count was selected. If the number of household members within that age range in the screener roster was nonzero, then that number was chosen as the final count. Otherwise, the

number of household members within the upper age range for the pair member with nonzero count was selected as the final count.

12. If the pair was a spouse-spouse pair, one count might have been zero while the other was nonzero because the spouse-spouse pair still lived with the parents of one pair member, and the pair member's younger siblings also lived in the household. In this case, the nonzero count was selected if the number of immediate family members (parent, child, sibling, spouse, grandchild, or grandparent) in the roster for the pair member with the zero count was less than his or her total household size.
13. In some cases, one pair member called the other pair member a parent or child, but the other pair member did not reciprocate. In the case of a child who did not reciprocate the parent's identification of him or her as a child, the child's count was always less than the parent's count. By the same token, in the case of a parent who did not reciprocate the child's identification of him or her as a parent, the parent's count was always less than the child's count. If the pair relationship was imputed to be "parent-child," then the pair member who did not acknowledge a parent-child relationship was overruled, and the maximum count of the two pair members was selected as final.

J.4 Spouse-Spouse Counts (with or without Children)

The multiplicity counts were not useful in the logic for the spouse-spouse household counts, since the spouse-spouse multiplicity counts were always 1.[8] If the household size was one, or the number of respondents aged 15 or older in the household was one or zero, then the final household person count was set to 0 since no spouse-spouse pairs could reside under those limits. If two family units had been previously identified in the household, the following rules were used to determine the final household person count:

1. When two different family units were already identified in the household, then two different parent sets were being referenced (one of the parent sets was often a single parent). The sum of the two counts (one count might be 0) was used, provided neither pair member had grandparents or grandchildren identified. This was to prevent spouse-spouse pairs from being counted twice, which would happen if grandparents were also parents of children younger than 18 years of age. If two family units were multigenerational families, then the final count was obtained by taking the maximum of the two pair members' counts.
2. It was possible for two different spouse-spouse pairs to be in the household, even though two different family units had not been identified. The final count was set to 2, even though two family units had not been previously identified, under the following conditions:
    - The pair relationship was not a spouse-spouse pair, and the total household size was at least four; and

[8] In rare cases, an individual might identify two spouses in the household. As noted in Section 11.2.2, the true multiplicity count in these cases was not determined; rather, the multiplicity count was set to 1, due to the complexity of determining the appropriate multiplicity count and the rarity of the occurrence of multiple spouses.

- Either:
  - Both sides identified a spouse,
  - Both sides identified a partner, or
  - One side identified a parent and the other side identified a parent-in-law.

3. If the conditions for the previous item were not met, it was still possible for two different spouse-spouse pairs to be in the household, even though two different family units were not previously identified. The final count was set to 2 under the following conditions:
   - One pair member had two parents with valid ages and both ages differed from the age of the spouse of the other pair member, and
   - The pair relationship was either sibling-sibling or a pair that was not a pair of interest.

Otherwise, reconciling the counts to a nonmissing value always required the following condition: There was no potential for two or more couples in the household that were not already obviously identified, whereby one of the pair members had at least four roster members of at least 15 years of age. This respondent had grandchildren younger than 18 years of age, did not have children-in-law, and had household members aged 12 or older who were not children, grandchildren, siblings, children, parents, spouses, or partners. For all remaining cases where a final household count needed to be assigned—in addition to the above condition—the final count was assigned using the following rules:

4. Among the majority of pairs, the counts for the two sides agreed. However, both sides had to meet the following conditions in order for the final count to be set to one of the sides:
   - The pair could not be a spouse-spouse pair where both respondents had a spouse or both respondents had a partner,
   - No bad relationship codes for roster members aged 15 or older for either pair member,
   - The number of spouse-spouse pairs was either one or zero for both pair members,
   - The household size was greater than 1 and nonmissing on both sides,
   - One pair member had at least two household members aged 15 or older, and
   - There were not two spouse-spouse pairs in the household according to the conditions given in item #3.
5. The counts might have agreed even though the above conditions were not met. The count could still have been set to one of the sides if any one of the following was true:
   - One pair member was younger than 18 and had no bad relationship codes for roster members aged 18 or older, but he or she did have bad relationship codes for roster members between the ages of 15 and 17 years old. The other pair member had no bad relationship codes for roster members aged 15 or older.

- One pair member had a single bad relationship code, and no other relationship codes could match it to make it a couple (i.e., the pair member did not have a single identified parent, grandparent, parent-in-law, or child-in-law). The other pair member had no bad relationship codes.
- One pair member had bad relationship codes among roster members aged 15 or older or had bad ages, and the other had no bad ages or relationship codes, where the pair member with no bad roster entries had the same number of household members aged 15 or older as the screener. The pair member with the bad roster entries would not have had the same age composition as the screener if the number of roster members aged 15 or older was added to the number of roster members with bad ages.
- One pair member had bad relationship codes among roster members aged 15 or older or had bad ages, and the other had no bad ages or relationship codes, where all the relationship codes for the pair member with no bad roster entries were immediate family codes (child, parent, sibling, spouse, partner, grandparent, or grandchild). For the pair member with bad roster entries, all the existing relationship codes were immediate family codes.

6. For those cases where the pair was imputed to be a spouse-spouse pair and both sides agreed that only one spouse-spouse pair was in the household, the count was set to 1 if any one of the following conditions was true:
   - Both sides had fewer than four people older than 15 in the household, or
   - One side had fewer than four people older than 15 in the household, and the other side had no bad relationship codes among roster members aged 15 or older
7. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under any one of the following conditions:
   - There were no bad relationship codes among roster members aged 15 or older, or
   - There were no bad relationship codes among roster members aged 18 or older and the pair member had parents.
8. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could have been because a couple entered the household or otherwise materialized after screening. The smaller count was chosen as the final count in this instance, which was identified if the following conditions were satisfied:
   - The screener count of roster members aged 12 or older was no larger than the count of roster members aged 12 or older in the roster of the pair member with the smaller spouse-spouse count.
   - The screener count of roster members aged 12 or older was smaller than the count of roster members aged 12 or older in the roster of the pair member with the larger spouse-spouse count.
   - The difference between the screener count of roster members aged 12 or older and the count of roster members aged 12 or older in the questionnaire rosters of the

pair members was smallest with the pair member with the smaller spouse-spouse count.

9. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could have been because a couple left the household or otherwise dissolved after screening. The larger count was chosen as the final count in this instance, which was identified if the following conditions were satisfied:
   - The screener count of roster members aged 12 or older was no larger than the count of roster members aged 12 or older in the roster of the pair member with the larger spouse-spouse count.
   - The screener count of roster members aged 12 or older was larger than the count of roster members aged 12 or older in the roster of the pair member with the smaller spouse-spouse count.
10. In many cases where the count of the number of spouse-spouse pairs did not agree between the two pair members, one side had a zero count and the other did not. The nonzero count was selected if the pair member associated with the zero count was not a close relative or somehow did not identify a spouse, partner, two parents, or two grandparents. The following conditions were required to select the nonzero count:
    - The pair member with a nonzero count either identified a spouse, a partner, two parents, or two grandparents.
    - The number of roster members aged 15 or older associated with the nonzero count pair member was no larger than the corresponding number associated with the zero count pair member.
    - If the side associated with the nonzero count identified a spouse, partner, or two parents, the following additional conditions were required:
      - The number of roster members between the ages of 26 and 44 was the same between the two pair members.
      - The number of roster members between the ages of 30 and 49 was the same between the two pair members.
      - The number of roster members between the ages of 35 and 54 was the same between the two pair members.
      - The number of roster members between the ages of 40 and 59 was the same between the two pair members.
    - If the side associated with the nonzero count identified two grandparents, the following additional condition was required:
      - The number of roster members aged 50 or older was the same between the two pair members.
11. The counts might not agree because a pair member's partner did not consider the other pair member's family as his or her own family. If at least one side identified a partner and the maximum count was 1, then the maximum was selected if both pair members had the same number of household members aged 15 or older. Otherwise, if

the pair members had a different number of household members aged 15 or older, the count belonging to the pair member with a count of household members aged 15 or older closer to that of the screener was used as the final count.

12. The counts might not agree because a pair member had two grandparents and an uncle/aunt husband-wife pair in the household. The maximum was selected if the pair member associated with the smaller count had a grandparent and had at least two roster members who were neither parents, siblings, children, spouses, partners, or grandparents, and the pair member with the larger count had children-in-law.
13. The count of the number of spouse-spouse pairs might not agree because one of the pairs was a sibling and sibling-in-law, and there are no codes for sibling-in-law. The maximum count was selected if the pair member with the smaller count did not have a spouse or partner but did have siblings aged 15 or older, and there were household members in his or her roster that were not parents, children, siblings, spouses, partners, grandchildren, or grandparents.
14. The count of the number of spouse-spouse pairs might not agree because one side had no nuclear family or grandparent-grandchild relationship codes, and one of the selected respondents was not in a child-parent, child-grandparent, or spouse-spouse relationship. The maximum count was selected if the following conditions were met:

    - The pair member's roster associated with the minimum count (usually 0) had no children, parents, siblings, spouses, partners, grandchildren, or grandparents among respondents aged 12 or older; and
    - The pair member's roster associated with the maximum count had some roster members who were not children, parents, siblings, spouses, partners, grandchildren, or grandparents.

    Note that this condition also nabbed cases where the relationship codes were not correctly identified on one pair member's roster. This occurred rarely, but when it did, the minimum count was 1 and the maximum count was 2.

15. The count of the number of spouse-spouse pairs might not agree because the pair members were siblings, and one sibling did not consider a stepparent or parent's partner as a "parent." The maximum count was selected if the following conditions were met:

    - The pair members were siblings,
    - The pair member associated with the maximum count had two parents,
    - The pair member associated with the minimum count had one parent, and
    - The roster associated with the pair member with the maximum count had more immediate family members (children, parents, siblings, spouses, partners, grandchildren, or grandparents) than the roster associated with the other pair member.

16. The count of the number of spouse-spouse pairs might not agree because the household changed after screening, which was not accounted for by previous conditions. In general, the count with a household composition closest to the screener

was selected. The age composition was defined by looking at age classes. The count for a given pair member was selected if any of the following properties held:

- The number of roster members between the ages of 26 and 44 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 30 and 49 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 35 and 54 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 40 and 59 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.

17. In some cases, neither pair member's household composition matched that of the screener. In that case, the household roster closest to that of the screener was selected. The maximum was selected if the number of screener roster members aged 12 or older exceeded the corresponding count from the questionnaire rosters of both pair members, which also differed from each other.

18. The counts might not agree because, on the rare occasion, one pair member in a spouse-spouse pair identified two grandparents of a different gender. Since there is no code for grandparents-in-law, they could not be identified, so the maximum count was selected. The following conditions were required:

    - The pair was a spouse-spouse pair.
    - The pair member with the maximum count had two grandparents of a different gender, and the pair member with the minimum count did not have any.

    The assumption here, of course, is that the grandparents of a different gender were in fact a spouse-spouse pair. There was no way to check whether a grandfather was the father's father and the grandmother was the mother's mother, for example.

19. Even though the household composition may match in terms of ages across the screener roster and the two pair members' rosters, the counts may disagree where two spouse-spouse pairs were clearly identified by one pair member but not the other. This may be because one of the in-laws was incorrectly identified on one side, or because a partner was not considered an in-law by a responding pair member, or because a partner did not consider other family members as "in-laws." The following conditions were required for the maximum count to be selected:

    - The number of screener roster members aged 12 or older matched the corresponding count from the questionnaire rosters of both pair members.
    - The pair member with the maximum number of spouse-spouse pairs had a spouse or partner and also had two parents.

- There were no bad relationship codes among roster members aged 15 or older on either pair member's roster.

20. If the counts for each pair member were not equal but the number of roster members aged 12 or older was the same between the two pair members, and the count for one pair member was the maximum possible in the household, then that number was selected as the final count. This condition was applied only after all other conditions, including conditions where the final count was ambiguous, had already been applied.
21. After accounting for all other rules, if the number of spouse-spouse pairs was still missing, but the lower and upper bounds for imputation were equal to each other, then the final household-level person count was set to one of those bounds.

J.5 Spouse-Spouse Counts (with Children)

The household counts for spouse-spouse counts with children obviously depended upon the counts obtained for spouse-spouse counts with or without children. The first two rules described in this section were determined directly from the spouse-spouse counts or from the household size, and no reconciliation of counts was necessary:

1. For a sizable proportion of cases, clearly no couples with children could be in the household, either because the spouse-spouse count was 0 or the household size was two or less. In these cases, the final spouse-spouse-with-children count was set to 0.
2. An additional small number of cases also could be readily determined by looking at the spouse-spouse count. If one pair member had a spouse-spouse-with-children count that exceeded the final spouse-spouse count, but the other pair member had a spouse-spouse-with-children count that was equal to or smaller than the final spouse-spouse count, then the final spouse-spouse-with-children count was set to the pair member's count that was consistent with the final spouse-spouse count.

The remainder of cases involved households with at least one spouse-spouse couple. After assigning values for the conditions described above, the assignment of values for these cases was done using the rules described in the rest of this section. If two family units had been previously identified in the household, the following rule was used to determine the final household person count:

3. When two different family units were already identified in the household, then two different parent sets were being referenced (one of the parent sets was often a single parent). The sum of the two counts (one count might be 0) was used, provided the spouse-spouse count was greater than 1. In that event, the maximum count was used.

Otherwise, reconciling the counts to a nonmissing value always required the following condition: There was no potential for two or more couples in the household that were not already obviously identified, whereby one of the pair members had at least four roster members of at least 15 years of age. This respondent had grandchildren younger than 18 years of age, did not have children-in-law, and had household members aged 12 or older who were not children, grandchildren, siblings, children, parents, spouses, or partners. For all remaining cases where a final household count needed to be assigned—in addition to the above condition (unless specifically noted below)—the final count was assigned using the following rules:

4. For cases that were not already determined by looking at the previous two conditions, the counts for the two pair members (if there were two pair members) were equal in the vast majority of cases. The final count could be set to each pair member's count under the following conditions:
   - Both pair members had valid rosters.
   - Either:
     – The counts were nonzero and equal to the final spouse-spouse count, or
     – There were no bad relationship codes for roster members younger than 18, and one of the following conditions held for at least one pair member:
       - The pair member's roster had no bad relationship codes for roster members aged 15 or older,
       - The pair member was older than 18 and had neither children nor siblings younger than 18 (covers zero counts since no bad codes were for members younger than 18), or
       - The pair member was younger than 18 and did not have parents, but there was one bad relationship code among roster members older than 18 in that pair member's roster (covers zero counts since only one bad relationship code could potentially be a single parent but not a pair of parents making a couple).
5. The pair members might both have had zero counts, but the above conditions did not apply. The final count could still have been 0 if the age counts for both pair members and the screener indicated nobody lived in the household who was younger than 18 and there were no bad roster ages. (In this case, it was not necessary to check for the potential of two or more family units in the household.)
6. The counts for both pair members might still have agreed with nonzero counts, even though none of the previous conditions applied. The final count could still have been set to one of the pair member's counts if the pair relationship was imputed to be a spouse-spouse pair with children.
7. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under one of the following conditions:
   - The count for the pair member with the valid roster was nonzero and equal to the final spouse-spouse count, or
   - There were no bad relationship codes for roster members younger than 18, and one of the following conditions held for the pair member with the valid roster. Either:
     – The pair member's roster had no bad relationship codes for roster members aged 15 or older,
     – The pair member was older than 18 and had neither children nor siblings younger than 18 (covers zero counts since no bad codes were for members younger than 18), or

– The pair member was younger than 18 and did not have parents, but there was one bad relationship code among roster members older than 18 in that pair member's roster (covers zero counts since only one bad relationship code could potentially be a single parent but not a pair of parents making a couple).

8. The pair member with the valid roster might have had a zero count, but the above conditions did not apply. The final count could still have been 0 if the age counts for both the pair member with the valid roster and the screener indicated nobody lived in the household who was younger than 18 and there were no bad roster ages. (In this case, it was not necessary to check for the potential of two or more family units in the household.)

9. If the spouse-spouse-with-children counts disagreed in the same manner as the spouse-spouse counts disagreed, then the choice was obvious: Use the count that corresponded to the correct spouse-spouse count. (In this case, it was not necessary to check for the potential of two or more family units in the household.) Details follow:

   - If the spouse-spouse-with-children counts were equal to the spouse-spouse counts for both pair members, even though they were unequal to each other, then the final spouse-spouse-with-children count was set to the final spouse-spouse count.

   - If the spouse-spouse counts exceeded the spouse-spouse-with-children counts by one for each pair member, even though they were unequal to each other, then the final spouse-spouse-with-children count was set to one less than the final spouse-spouse count.

10. Based on earlier conditions, households without couples were already excluded. Households with a possibility of two or more couples were also excluded. If the pair relationship was parent-child and at least one count was nonzero, then the identified couple corresponded to the parent-child relationship. The maximum of the counts was selected under the following conditions:

    - The sum of counts from the two pair members was 1.

    - Either:

      – The relationship was parent-child where the child was between the ages of 12 and 17, or

      – The relationship was parent-child where the child was between the ages of 18 and 20 and the child had siblings younger than 18.

11. In some cases, two couples were identified in the household where the household was multigenerational (one member of the younger couple was in a parent-child relationship with the older couple). If a sibling to the pair member in the younger couple was selected, or if a member of the younger couple was selected who "married into" the family, then he or she was not able to identify the nephews, nieces, brothers-in-law, or sisters-in-law—which could point to an appropriate accounting of all the couples with children—because of the relationship codes that were available. The maximum of the two counts was selected under the following conditions:

    - There were two couples in the household, as identified by the final spouse-spouse count.

- The difference between the pair members' counts was 1.
- Either:
  - The pair member with the smaller count had a spouse or partner and the pair member with the larger count had parents in the household, or
  - The pair member with the smaller count had parents-in-law or children-in-law in the household.

12. If a couple was in a marriage/partnership that occurred after an earlier marriage, the partner might not have considered the partner's children as his or her children, but the child (who also was selected) considered the spouse/partner a parent. Even though the pair relationship was not parent-child, these cases were still counted as spouse-spouse with children since they consisted of the children of one spouse/partner. The maximum count was selected under the following conditions:
    - The pair relationship was not one of interest.
    - One count was 0 and the other count was 1.
    - The pair member with the zero count had a spouse or partner.
    - The pair member with the nonzero count had parents.
    - The spouse-spouse final count was nonmissing.
13. The counts might have been unequal because children younger than 18 left, entered, or otherwise materialized or disappeared in the household after screening and between the time of the interviews. In general, the count was selected that corresponded to the pair member with a household composition closest to the screener household composition. If one pair member did not have children in the household and the other pair member did, the following conditions were required for the count corresponding to the pair member with a household composition closest to the screener:
    - One pair member had a nonzero count of children younger than 18 and the other pair member had a zero count of children younger than 18.
    - Either:
      - The screener composition indicated that children younger than 18 were in the household, whereupon the nonzero count was selected, or
      - The screener composition indicated that no children younger than 18 were in the household, whereupon the zero count was selected.
14. The counts might have been unequal with a count of 0 and a count of 1 because a pair member with a count of 0 was not part of the immediate family unit. The nonzero count was used under the following conditions:
    - The pair relationship was not a parent-child, sibling-sibling, spouse-spouse, or grandparent-grandchild relationship.

- Both pair members had relationship codes that were not parent, child, sibling, spouse, partner, grandparent, or grandchild codes among roster members who were aged 12 or older.

15. The counts might have been unequal because of bad relationship codes among roster members younger than 18. The following rules were used to determine if the count associated with the pair member did not have bad relationship codes:
    - The number of roster members younger than 18 was the same between both pair members.
    - The side with the smaller count had at least one bad relationship code for roster members younger than 18.
16. If, after considering all of the general conditions given above, the count was left to imputation, it was still possible that the lower and upper bounds were equal. In this instance, the final count was set to one of the bounds.

Appendix K: Priority Conditions Used to Create Household-Consistent Covariates

Appendix K: Priority Conditions Used to Create Household-Consistent Covariates

K.1 Household Size

In Table K.1, blank entries indicate that no conditions were required for that set of variables. The reported household size variable is QD54, and the edited household size variable is TOTPEOP, which cannot differ from the raw variable by more than 1. Any variable suffixed by "A" indicates that the variable corresponds to the value for pair member "A." A similar comment can be made with regard to the suffix "B." For example, "QD54A" reflects the reported household size for pair member A. The quality-of-roster counts are considered in the column "any roster missing?" The variables GOODAGEA and GOODAGEB are the total number of cases in the roster with valid ages. The variables that appear in the table are TGOODAGA and TGOODAGB, the total number of cases in the roster with valid ages, incorporating the minimum possible counts within the age categories 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older. Finally, the variable used to describe the screener household size is SHHSIZE. The conditions used to create the variable HHSIZE resulted in no missing values for this variable, and thus no imputation was required. The first column in Table K.1 shows the hierarchical priority condition, with the frequency of occurrence for each priority condition in parentheses.

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
1 (16,966)	Equal, both > 1, both nonmissing	Equal, both > 1, both nonmissing				TOTPEOPA
2 (0)	Equal, both > 1, both nonmissing	TOTPEOPB one more than TOTPEOPA	TGOODAGA ≤ QD54A	A: no		QD54A
3 (0)	Equal, both > 1, both nonmissing	TOTPEOPA one more than TOTPEOPB	TGOODAGB ≤ QD54B	B: no		QD54B
4 (0)	Equal, both > 1, both nonmissing	TOTPEOPA one more than TOTPEOPB	TGOODAGA = TGOODAGB TGOODAGA ≤ TOTPEOPA		SHHSIZE not equal to QD54A	TOTPEOPA
			TGOODAGA = TOTPEOPA		No condition	
5 (0)	Equal, both > 1, both nonmissing	TOTPEOPB one more than TOTPEOPA	TGOODAGA = TGOODAGB ≤ TOTPEOPB		SHHSIZE not equal to QD54B	TOTPEOPB
			TGOODAGB = TOTPEOPB		No condition	
6 (0)	Equal, both > 1, both nonmissing	Within one of each other			SHHSIZE at least as large or larger than screener roster, equal to QD54A	SHHSIZE
7 (0)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1, not equal to QD54B	QD54B ≥ TGOODAGB		SHHSIZE ≥ 2, closer to QD54B than TOTPEOPB	QD54B
8 (8)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1	TGOODAGB ≤ TOTPEOPB (no bad roster ages if equal)		SHHSIZE ≥ 2, TOTPEOPB is as close as QD54B	TOTPEOPB
9 (0)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1	TGOODAGB ≤ SHHSIZE		TGOODAGB ≤ SHHSIZE	SHHSIZE
10 (0)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1				TGOODAGB

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
11 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1, not equal to QD54A B: missing or 1	QD54A ≥ TGOODAGA		SHHSIZE ≥ 2, closer to QD54A than TOTPEOPA	QD54A
12 (11)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1	TGOODAGA ≤ TOTPEOPA (no bad roster ages if equal)		SHHSIZE ≥ 2, TOTPEOPA is as close as QD54A	TOTPEOPA
13 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1	TGOODAGA ≤ SHHSIZE		TGOODAGA ≤ SHHSIZE	SHHSIZE
14 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1				TGOODAGA
15 (2)	Both missing or 1	Both missing or 1			SHHSIZE ≥ 2, SHHSIZE at least as large or larger than screener roster	SHHSIZE
16 (34)	Not equal, both > 1	TOTPEOPB = QD54B	A: At least one age range variable less than min.[1] B: Age range variables all same or larger than min.			QD54B
		TOTPEOPA = QD54A	B: At least one age range variable less than min. A: Age range variables all same or larger than min.			QD54A
17 (5)	Not equal, both > 1		A: At least one age range variable less than min. B: At least one age range variable less than min.		Age range variables all same or larger than min.	SHHSIZE

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
18 (1,029)	Not equal, both > 1	QD54A is equal to at least one of TOTPEOPA or TOTPEOPB	A: Age range variables all same or larger than min., no bad roster ages		SHHSIZE at least as large or larger than screener roster, equal to QD54A	QD54A
		QD54B is equal to at least one of TOTPEOPA or TOTPEOPB	B: Age range variables all same or larger than min., no bad roster ages		SHHSIZE at least as large or larger than screener roster, equal to QD54B	QD54B
19 (0)	Not equal, both > 1	QD54A is equal to at least one of TOTPEOPA or TOTPEOPB	A: At least one age range variable less than min., or some bad roster ages		SHHSIZE at least as large or larger than screener roster, equal to QD54A	A: Maxima for each age range between given count and min.
		QD54B is equal to at least one of TOTPEOPA or TOTPEOPB	B: At least one age range variable less than min., or some bad roster ages		SHHSIZE at least as large or larger than screener roster, equal to QD54B	B: Maxima for each age range between given count and min.
20 (3)	Not equal, both > 1	Not equal, both >1	TGOODAGA = TGOODAGB, TGOODAGA = QD54A	A: no B: no		QD54A
			TGOODAGA = TGOODAGB, TGOODAGA = QD54B	A: no B: no		QD54B
21 (0)	Not equal, both > 1	Not equal, both >1	TGOODAGA = QD54A TGOODAGB > QD54B	A: no B: no	SHHSIZE > QD54B	QD54A
			TGOODAGB = QD54B TGOODAGA > QD54A	A: no B: no	SHHSIZE > QD54A	QD54B

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
22 (0)	Not equal, both > 1	Not equal, both > 1	TGOODAGA > GOODAGEA, TGOODAGB > GOODAGEB, TGOODAGA = SHHSIZE TGOODAGB = SHHSIZE	A: no B: no	TGOODAGA = SHHSIZE, TGOODAGB = SHHSIZE	SHHSIZE
23 (0)	Not equal, both > 1	Not equal, both > 1	TGOODAGA > GOODAGEA, TGOODAGB > GOODAGEB, TGOODAGA = TGOODAGB	A: no B: no		TGOODAGA
24 (8)	Not equal, both > 1	Not equal, both > 1		A: no B: no	SHHSIZE = sum of maxima for each age group across pair members	SHHSIZE
25 (161)	Not equal, both > 1	Not equal, both > 1		A: no B: no	SHHSIZE ≥ 2, at least as large or larger than screener roster, closer to one of the QD54's	QD54A if SHHSIZE closer to A, QD54B if closer to B
					SHHSIZE ≥ 2, at least as large or larger than screener roster, equidistant between the QD54's	QD54 of oldest pair member

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener HHSIZE Characteristics	HHSIZE Equals:
26 (2)	Not equal, both > 1	Not equal, both > 1		A fewer than B	SHHSIZE ≥ 2, at least as large or larger than screener roster, closer to QD54A than QD54B	QD54A
				B fewer than A	SHHSIZE ≥ 2, at least as large or larger than screener roster, closer to QD54B than QD54A	QD54B
				No condition	SHHSIZE ≥ 2, at least as large or larger than screener roster, equidistant between the QD54's	QD54 of oldest pair member
				No condition	SHHSIZE ≥ 2, at least as large or larger than screener roster	SHHSIZE
27 (0)			At least 3 of the age range variables are missing		SHHSIZE ≥ 2, at least as large or larger than screener roster	SHHSIZE

[1] "Min." refers to the minimum possible within each age range based upon the ages of the two pair members.

K.2 Age Variables

Table K.2 illustrates the hierarchical priority conditions ("priorities") used to create a new household-consistent 12 to 17 age group count. Similar priority conditions are used for the 0 to 11, 12 to 14, 12 to 20, 18 to 25, 26 to 34, 35 to 49, 50 or older, and 15 or older age groups. In this table, blank entries indicate that no priority conditions were required for that set of variables. As with the previous table, a variable followed by "A" indicates that the variable corresponds to the value for pair member "A." A similar comment can be made with regard to "B."

As stated earlier, the variables GOODAGEA and GOODAGEB are the total number of cases in the roster with valid ages, and the variables TGOODAGA and TGOODAGB are also the total number of cases in the roster with valid ages, but if the original adjusted count is less than the minimum required, the original count is replaced by the minimum within the age categories 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older. As noted in Section 11.3.1, these counts are adjusted so that the roster ages match what was entered in each pair member's questionnaire. Hence, AGE1217A is the adjusted count of 12- to 17-year-olds for pair member A, and AGE1217B is the adjusted count of 12- to 17-year-olds for pair member B. If AGE1217A or AGE1217B is less than the minimum possible, the count is replaced by the minimum, which is given by TAG1217A and TAG1217B, respectively. Otherwise, AGE1217A and TAG1217A are equivalent, as are AGE1217B and TAG1217B. The sum of AGE011A, AGE1217A, AGE1825A, AGE2634A, AGE3549A, and AGE50PA is GOODAGEA. Similarly, the sum of TAG011A, TAG1217A, TAG1825A, TAG2634A, TAG3549A, and TAG50PA is TGOODAGA. The same can be said for GOODAGEB and TGOODAGB.

The final 12 to 17 age count is denoted by AGE1217. The screener age count, denoted by SAGE1217, is used only if the age counts in each pair member's roster cannot conform to the minimum necessary or otherwise are not possible to incorporate. If after all edits the count for AGE1217 is missing, but the counts for other age groups are not missing, and the counts for the 0 to 11 age group are the same for both pair members, then the sum of the counts for the other age groups, plus the minimum possible for AGE1217, are given by EXC1217. If other means fail to determine the appropriate value for the age count, match measures are used. These are measures that summarize the quality of the match between the two pair members. A match label of "0" indicates a perfect match, where the pair member's roster has a household member who is identified as the other pair member with a perfect match on age and gender and is indicated as the other pair member by the MBRSEL variable. There are several levels of match measures where a lower number signifies a better quality match. These measures are explained in detail in Section 11.2.1.2.1. As a final check, if the age group counts do not equal HHSIZE, and the counts for the pair members are unequal, then the count is set to missing. As with Table K.1, the first column in Table K.2 shows the hierarchical "priority," with the frequency of occurrence for each priority in parentheses, for the AGE1217 count. In most cases, the frequencies corresponding to the other age ranges were the same as the frequency for AGE1217. In those cases where the frequency differed, footnotes provide details of the differences.

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
1 (0)[1]	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEB = HHSIZE, all nonmissing, all > 1	AGE1217A < min. (minimum), AGE1217B ≥ min.				AGE1217B
2 (0)[2]		AGE1217B < min. AGE1217A ≥ min.				AGE1217A
3 (0)		AGE1217A < min. AGE1217B < min.		SHHSIZE = HHSIZE, SAGE1217 ≥ min.		SAGE1217
4 (5)[3]		AGE1217A = AGE1217B, both ≥ min.	Another count except 12-17 < min.			AGE1217A
5 (0)[4]		AGE1217A not equal to AGE1217B, both ≥ min.	AGE1825A < min., AGE1825B ≥ min.			AGE1217B
6 (0)[5]			AGE1825B < min., AGE1825A ≥ min.			AGE1217A
7 (0)			Another count except 12-17 < min.		Fewer roster entries missing in A than B	AGE1217A
8 (0)					Fewer roster entries missing in B than A	AGE1217B
9 (1)[6]					A & B: none missing A has better match measure than B	AGE1217A

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
10 (0)[7]	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEB = HHSIZE, all nonmissing, all > 1	AGE1217A not equal to AGE1217B, both ≥ min.	Another count except 12-17 < min.		A & B: none missing B has better match measure than A	AGE1217B
11 (0)					A & B: none missing Age A ≥ Age B	AGE1217A
					A & B: none missing Age B > Age A	AGE1217B
12 (0)[8]						Missing
13 (16,022)		AGE1217A = AGE1217B	All other counts equal across pair members			AGE1217A
14 (555)		At least one age group has an unequal count between pair members		A: all age counts are equal to their screener counterparts	No missing roster entries on either side	AGE1217A
				B: all age counts are equal to their screener counterparts	No missing roster entries on either side	AGE1217B
15-22 (61)					A & B: none missing A has better match measure than B	AGE1217A
					A & B: none missing B has better match measure than A	AGE1217B
23 (82)					A & B: none missing Age A ≥ Age B	AGE1217A
					A & B: none missing Age B > Age A	AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
24 (4)	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEB = HHSIZE, all nonmissing, all > 1	At least one age group has an unequal count between pair members			Fewer roster entries missing in A than B A has good match measure (labels 0-7)	AGE1217A
					Fewer roster entries missing in B than A B has good match measure (labels 0-7)	AGE1217B
25 (0)					Fewer roster entries missing in A than B A is older than B	AGE1217A
					Fewer roster entries missing in B than A B is older than A	AGE1217B
26 (0)					Fewer roster entries missing in A than B B is older than A	AGE1217B
					Fewer roster entries missing in B than A A is older than B	AGE1217A
27 (0)					A & B: same number of roster entries missing (> 0) A is older than B	AGE1217A
28 (0)					A & B: same number of roster entries missing (> 0) B is older than A	AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
29 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEA = HHSIZE, GOODAGEB not equal to HHSIZE	AGE1217A < min. AGE1217B = min.				AGE1217B
30 (0)		AGE1217B < min. AGE1217A = min.				AGE1217A
31 (0)		AGE1217A < min. AGE1217B < min.		SAGE1217 ≥ min.		SAGE1217
32 (0)		AGE1217A = AGE1217B, both ≥ min.				AGE1217A
33 (0)		AGE1217A not equal to AGE1217B	AGE1825A < min. AGE1825B ≥ min.			AGE1217B
34 (0)			AGE1825B < min. AGE1825A ≥ min.			AGE1217A
35 (0)					Fewer roster entries missing in A than B	AGE1217A
36 (0)					Fewer roster entries missing in B than A	AGE1217B
37 (0)					A & B: same number of roster entries missing (> 0) A has good match measure (labels 0-7)	AGE1217A
38 (0)					A & B: same number of roster entries missing (> 0) B has good match measure (labels 0-7)	AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
39 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEA = HHSIZE, GOODAGEB not equal to HHSIZE	AGE1217A not equal to AGE1217B			A & B: same number of roster entries missing (> 0) A is older than B	AGE1217A
					A & B: same number of roster entries missing (> 0) B is older than A	AGE1217B
40 (0)		Priority conditions 29-39 not met				Missing
41 (0)		AGE1217 missing after priority conditions 29-40 invoked, other age range counts not missing				HHSIZE - sum of other age counts
42 (668)		Priority conditions 29-41 not met				AGE1217A
43 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEB = HHSIZE, GOODAGEA not equal to HHSIZE	AGE1217A < min. AGE1217B = min.				AGE1217B
44 (0)		AGE1217B < min. AGE1217A = min.				AGE1217A
45 (0)		AGE1217A < min. AGE1217B < min.		SAGE1217 ≥ min.		SAGE1217
46 (0)		AGE1217A = AGE1217B, both ≥ min.				AGE1217A
47 (0)		AGE1217A not equal to AGE1217B	AGE1825A < min. AGE1825B ≥ min.			AGE1217B
48 (0)			AGE1825B < min. AGE1825A ≥ min.			AGE1217A

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
49 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEB = HHSIZE, GOODAGEA not equal to HHSIZE	AGE1217A not equal to AGE1217B			Fewer roster entries missing in A than B	AGE1217A
50 (0)					Fewer roster entries missing in B than A	AGE1217B
51 (0)					A & B: same number of roster entries missing (> 0) B has good match measure (labels 0-7)	AGE1217B
52 (0)					A & B: same number of roster entries missing (> 0) A has good match measure (labels 0-7)	AGE1217A
53 (0)					A & B: same number of roster entries missing (> 0) A is older than B	AGE1217A
					A & B: same number of roster entries missing (> 0) B is older than A	AGE1217B
54 (0)		Priority conditions 43-53 not met				Missing
55 (0)		AGE1217 missing after priority conditions 43-54 invoked, other age range counts not missing				HHSIZE - sum of other age counts
56 (482)		Priority conditions 43-55 not met				AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
57 (285)	TGOODAGA = HHSIZE					TAG1217A
	TGOODAGB = HHSIZE					TAG1217B
58 (36)[9]	SHHSIZE = HHSIZE	AGE1217A, AGE1217B ≤ SAGE1217		AGE1217A & B ≤ SAGE1217		SAGE1217
59 (0)[10]	SHHSIZE = HHSIZE, HHSIZE = EXC1217	AGE1217 missing	Other counts not missing, AGE011A equals AGE011B			MIN1217
60 (13)[11]	Previous priority conditions for HHSIZE, TOTPEOP, GOODAGE, not met, either the two TOTPEOP's > 0, or SHHSIZE = HHSIZE			AGE1217A equals SAGE1217		AGE1217A
				AGE1217B equals SAGE1217		AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
61 (1)[12]	Previous priority conditions for HHSIZE, TOTPEOP, GOODAGE, not met, SHHSIZE = HHSIZE	AGE1217 missing	At least 3 of the other counts missing			SAGE1217
99 (0)[13]	All prior conditions were not met.					Missing

[1] The following frequencies were observed for priority condition #1: AGE1825 = 2; AGE1220 = 3; AGE011 = 10.

[2] The following frequencies were observed for priority condition #2: AGE1825, AGE3549, and AGE1220 = 1; AGE2634 and AGE50p = 2.

[3] The following frequencies were observed for priority condition #4: AGE011 = 0; AGE1220 = 6; AGE1825 and AGE2634 = 7; AGE1217 and AGE15p = 10; AGE1214 = 11.

[4] The frequency of priority condition #5 for AGE2634 = 1.

[5] The frequency of priority condition #6 for AGE3549 and AGE50p = 1.

[6] The following frequencies were observed for priority condition #9: AGE011, AGE2634, and AGE1214 = 0; AGE3549 and AGE50p = 2.

[7] The frequency of priority condition #10 for AGE3549 and AGE50p = 1.

[8] The frequency of priority condition #12 for AGE011 = 1.

[9] The following frequencies were observed for priority condition #58: AGE011 = 28; AGE1217 = 38; AGE15p = 34; AGE2634 and AGE1214 = 40; AGE3549 = 41; AGE50p = 43.

[10] The following frequencies were observed for priority condition #59: AGE3549 =1; AGE1217 = 2; AGE1825 = 3; AGE011 = 6.

[11] The following frequencies were observed for priority condition #60: AGE011 = 5; AGE50p = 6; AGE1214 = 12; AGE1217 = 15; AGE1825 = 9.

[12] The frequency of priority condition #61 for AGE15p = 0.

[13] The frequency of priority condition #99 for AGE2634 and AGE3549 = 1.

This page intentionally left blank

Appendix L: Creation of Household-Level and Person-Level Files

This page intentionally left blank

Appendix L: Creation of Household-Level and Person-Level Files

L.1 Introduction

For the 2014 administration of the National Survey on Drug Use and Health (NSDUH), a person was randomly selected for an interview through a five-stage sample selection process. States were first stratified into a total of 750 state sampling (SS) regions. Within each of these SS regions, a sample of census tracts was selected (i.e., the first stage of selection) with probabilities proportional to a composite size measure and with minimum replacement. For the second stage of selection, adjacent census block groups within selected census tracts were aggregated to meet the minimum dwelling unit (DU) requirements. One census block group was selected per sampled census tract with probability proportionate to a composite size measure and with minimum replacement. Within each selected census block group, adjacent census blocks were combined to form the third-stage sampling units or area segments. One area segment was selected within each sampled census tract with probability proportional to population size.[1] Once the sample segments were selected, specially trained field staff visited areas and created lists of all eligible DUs within the sample segment boundaries. These lists served as the frames for the fourth stage of sample selection. After the DUs were selected within each segment, an interviewer visited each selected DU to obtain a roster of all people aged 12 or older. This roster information was then used to select zero, one, or two people from the household at the fifth stage of sample selection.

At the end of the survey year, a household-level file and a person-level file were created to record the information obtained from the sampling processes described above. The household-level and person-level files were used in the final creation of the person-level and pair-level analysis weights. In addition, the person-level file was later subset into a smaller data file that contained only respondents who were considered "completed" cases; this file was used for analysis. Refer to Section L.3 for the definition of a completed case.

L.2 Dwelling Unit-Level Eligibility and Completeness Criteria

Before proceeding with the fifth stage of sample selection (i.e., within selected households), a set of rules was used to determine whether or not a DU was eligible to be selected. Eligibility of the DU was recorded in the binary variable DUELIG, where a value of 1 indicated eligibility. Two examples of DUs deemed to be ineligible included those defined as "vacant" and those determined to be "not a primary residence."

Occasionally, DUs were eligible but failed to complete the screening process. Reasons for not completing the screening process were recorded, including situations such as "language barrier," "refusal," and "denied access." Completeness of the screening process for the DU was recorded in the binary variable DUCOMP, where a value of 1 indicated completeness. For the

[1] Segments consist of clusters of the geographic aggregated adjacent census blocks. SS regions were formed through geographically partitioning each state into roughly equal-sized regions based on a composite size measure. The sample design report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2015d) contains more information regarding the sample design.

segments where all the DUs were from denied-access areas, such as gated communities, an adjustment was made in the final household-level file. Although the field interviewers could not obtain an accurate count of DUs from denied-access areas, these DUs were considered eligible. Therefore, DU information from the U.S. Census Bureau for these areas was used in the household-level file.

During the third stage of sampling, it was possible to select a sample segment more than once because samples were selected with replacement. These duplicated segments had different segment IDs (SEGIDs) for each duplicate. However, one SEGID contained all the DU information and the other had none. The number of eligible DUs was split as evenly as possible between the two SEGIDs, and this information was updated in the household-level and person-level files.

L.3 Person-Level Eligibility and Completeness Criteria

During screening, respondents were asked to identify all eligible household members so that only eligible individuals were listed and, therefore, potentially selected. Eligibility was determined according to the criteria provided in Section L.1. Eligible respondents at the time of screening were recorded in the binary variable PRELIG, which had a value of 1 if the household member was eligible. Respondents who were selected were recorded in the binary variable PRSEL, where 1 indicated a selected individual. It was possible for an individual to be selected, but at the time of the interview, to be determined ineligible. Examples of changes from eligibility to ineligibility included "the selected person turned out not to be a permanent resident in the DU" and "roster error." If this occurred, the value of PRELIG was changed from 1 to 0.

A summary of the number of selected, eligible, and completed dwelling units is shown in Table L.1. The number of eligible, selected, and interviewed individuals also is summarized in the table.

Table L.1 NSDUH Household, Person Eligibility, and Completed Interview Counts: 2014

	Selected Dwelling Units	Eligible Dwelling Units	Completed Screenings (Dwelling Units)	Eligible People	Selected People	Interviewed People	Completed Cases (Interviews)
CAI	185,013	154,533	127,605	270,252	91,640	67,969	67,901

CAI = computer-assisted interviewing.

To be considered a completed case for purposes of analysis, a respondent had to provide "yes" or "no" answers to the cigarette usage gate question and to at least 9 of the following additional drug usage gate questions: (1) chewing tobacco, (2) snuff, (3) cigars, (4) alcohol, (5) marijuana, (6) cocaine (in any form), (7) heroin, (8) hallucinogens, (9) inhalants, (10) pain relievers, (11) tranquilizers, (12) stimulants, and (13) sedatives.[2] Unlike the paper-and-pencil interviewing (PAPI) questionnaire in 1999 and surveys prior to 1999, no logical inference could be made from information within a section if the gate question was not answered. This was because the computer-assisted interviewing (CAI) instrument routed respondents out of a section if the gate question was not answered. Completeness of the survey for eligible individuals was

[2] For more details on the "usable case rule," see Section 2.2.3.

recorded in the binary variable PRCOMP, which had a value of 1 if the respondent was a completed case, and 0 if not. For a summary of the number of completed cases in the 2014 survey, see Table L.1.

L.4 Variables in the Household-Level and Person-Level Files

This section documents some of the important person-level variables that were created for the household-level and person-level files.

Screener-level demographic variables were created from the screener roster information in the household-level and person-level files. XAGE was the screener age, which either could be "continuous" (single-year ages) or categorical. A respondent could choose to give an age category instead of the actual age. The age categories with their accompanying codes were 199 = 12 to 17 years old; 299 = 18 to 25 years old; 399 = 26 to 34 years old; 499 = 35 to 49 years old; and 599 = 50 years old or older. Screener race (XRACE1-XRACE6), screener Hispanicity (XHISP), and screener gender (XSEX) also were produced from the screener roster information. XRACE1 through XRACE6 were indicator variables representing white, black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and other, respectively. The household-level variable PAIRSEL represented the number of people within each age group selected from a DU. It was a 20-level variable indicating whether zero, one, or two individuals were selected from the five age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older) in a given household. (If two people were selected from the household, this variable indicated the age groups of both pair members.) Similar to PAIRSEL, the household-level variable PAIRRESP had 20 levels, which indicated whether zero, one, or two people completed the interviews from the five age groups within a household.

As described in the sample design report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2015d), states were partitioned into SS regions, which were further partitioned into clusters of adjacent blocks called "segments." The variable SEGID (segment ID number) was a two-letter state abbreviation followed by a two-digit SS region and a two-letter segment identifier, with the fifth character indicating the survey year and the sixth character being the initial quarter assignment of the segment. Census region (REGION) was a four-level geographic variable recoded from the respondent's state of residence. The four levels were Northeast, Midwest, South, and West. The population density variables PDEN10 and PDEN210[3] classified respondents according to their living situation, whether it was in a rural or urban area, and, if urban, the size of the urban area. They were used to categorize segments where the respondents lived based on 2010 census data and the December 2009 Core-Based Statistical Area (CBSA) classifications.[4] PDEN210 had five levels: segment in CBSA with 1 million or more people; segment in CBSA with 250,000 to 999,999 people; segment in CBSA with fewer than 250,000 people; segment in urban area but not in CBSA; and segment in rural area (not in CBSA and not in urban area). PDEN10 had three levels: segment in CBSA with 1 million or more people; segment in CBSA with fewer than 1 million people; and segment not in CBSA.

[3] PDEN10 and PDEN210 were similar to PDEN/PDEN2 and PDEN00/PDEN200 in previous sample designs. The suffix "10" indicates that the variables were derived from the 2010 census.

[4] CBSAs include metropolitan and micropolitan statistical areas as defined by the Office of Management and Budget (2009).

The variable STATE represented the Federal Information Processing Standards (FIPS) state codes for the 50 states and the District of Columbia. The variable STATE was created at the sampling stage and did not contain any missing values. The variables VESTR and VEREP were created to capture the sampling design structure. Each SS region appeared in a different variance estimation stratum (VESTR) every quarter. Two replicates (VEREP) were defined within each variance stratum. Each replicate consisted of four segments, one for each quarter of data collection. The segment-level variable RURURB10[5] is derived from the 2010 census block-level designation of rural or urban.[6] In the NSDUH sample, if 100 percent of the blocks are rural, the segment is defined as rural (RURURB10 = 1). If one or more of the blocks within a segment is urban, the segment is defined as urban (RURURB10 = 2). Other sampling variables such as DIVISION, SSREGION, GQTYPE, ID, STNAME, STUSAB, and QUARTER[7] also were included in the household-level and person-level files.

---

[5] RURURB10 was similar to RURORURB and RURURB00 in previous sample designs. The suffix "10" indicates that this variable was derived from the 2010 census.

[6] The census classifies as urban all blocks located within urbanized areas (UA) and urban clusters (UC). UAs and UCs generally consist of core census block groups or blocks that have a population density of at least 1,000 people per square mile and surrounding census blocks that have an overall density of at least 500 people per square mile. In addition, under certain conditions, less densely settled territory may be part of each UA or UC.

[7] For more details on these sampling variables, refer to the sample design report in the 2014 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2015d).