2011 NATIONAL SURVEY ON DRUG USE AND HEALTH

IMPUTATION REPORT

Prepared for the 2011 Methodological Resource Book

Contract No. HHSS283200800004C
RTI Project No. 0211838.207.006.007
Deliverable No. 39

Authors:

Peter Frechtel (Editor)
Heather Archambault
Lisa Carpenter
Susan Edwards
Simon King
Jeff Laufenberg (Editor)
Peilan Martin
Andrew Moore
Victoria Scott

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

February 2013

2011 NATIONAL SURVEY ON DRUG USE AND HEALTH

IMPUTATION REPORT

Prepared for the 2011 Methodological Resource Book

Contract No. HHSS283200800004C
RTI Project No. 0211838.207.006.007
Deliverable No. 39

Authors:

Peter Frechtel (Editor)
Heather Archambault
Lisa Carpenter
Susan Edwards
Simon King
Jeff Laufenberg (Editor)
Peilan Martin
Andrew Moore
Victoria Scott

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

February 2013

Acknowledgments

This report would not be possible without the guidance and input of staff from the Substance Abuse and Mental Health Services Administration (SAMHSA). In particular, Jonaki Bose provided useful comments. Special thanks are also due to several current and former RTI International (a trade name of Research Triangle Institute) 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 copyedited the report, and Valerie Garner formatted the report in preparation for publication. Teresa Bass provided web production.

Table of Contents

Chapter **Page**

Table of Contents (continued)

Chapter **Page**

Table of Contents (continued)

Chapter **Page**

Table of Contents (continued)

Chapter **Page**

Appendix

List of Tables

List of Tables (continued)

Table **Page**

List of Figures

Figure **Page**

1. Introduction

Conducted annually, the National Survey on Drug Use and Health (NSDUH) is the primary source of information on the prevalence, patterns, and consequences of alcohol, tobacco, and illegal drug use and abuse among all U.S. civilian, noninstitutionalized residents of the 50 States and the District of Columbia, aged 12 or older. In the 2011 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 persons with no fixed household address (e.g., homeless transients not in shelters), residents of institutional group quarters (e.g., jails and hospitals), children younger than 12, and active military personnel. As it has since 1999, the 2011 implementation utilized a 50-State, multistage cluster design that enables the Substance Abuse and Mental Health Services Administration to provide representative estimates for each State and the District of Columbia. Both direct and model-based estimates are produced on a variety of measures based on a combination of multiple years of data.

Although the design of NSDUH has not changed significantly since the introduction of computer-assisted interviewing (CAI) technology in 1999, four important methodological changes were introduced in the 2002 survey that affected the estimates from the survey years that followed. First, the name of the survey was changed from the National Household Survey on Drug Abuse (NHSDA) to its present form in the 2002 survey year. Second, beginning in 2002, each NSDUH respondent has received an incentive payment of $30 for their participation on the survey; prior to that survey year, there were no such payments. Third, improved data collection quality control procedures were introduced in the survey during 2001 and 2002. Finally, since 2002 when it first became available, information from the 2000 U.S. Decennial Census has been used in the NSDUH weighting procedures. (Surveys conducted in previous years relied upon information from the 1990 U.S. Decennial Census.) Because of these changes, the 2002 survey year is considered the "baseline year" from which all trends are currently measured.

This report focuses on the predictive mean neighborhood (PMN) imputation methodology employed for the 2011 NSDUH. It also includes information about some logical editing procedures. 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.[1] The editing and coding section (Section 10) of the 2011 Methodological Resource Book (MRB) contains documentation for most procedures for logically editing data in the 2011 NSDUH.[2] This imputation report (Section 11 of the 2011 MRB) contains descriptions of additional editing procedures for variables that subsequently undergo statistical imputation. For some variables, these procedures may involve assumptions (e.g., picking the midpoint from a

[1] There are a few situations where data from outside the respondent's record is used in logical editing. 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. In the editing procedures for the roster and (especially) the roster pair variables (see Table 1.1), the record of the other pair member is consulted. A pair occurs when two members of the same dwelling unit are selected for the survey and complete the interview. This enables the gathering of information about relationships among household members.

[2] See Kroutil, Handley, and Bradshaw (2013); Kroutil, Handley, Bradshaw, Chien, and Felts (2013); and Kroutil and Chien (2013).

range) or other final edits to prepare the data for imputation. Table 1.1 provides a crosswalk to identify the location of relevant documentation on editing procedures in the 2011 MRB.

Table 1.1 Location of Documentation on Editing in the 2011 Methodological Resource Book

Variable Group	Location within MRB Section 10 (Editing and Coding)[1]	Location within MRB Section 11 (Imputation) Where Logical Edits/Additional Edits Are Discussed
Core Demographics	Report 3: Interviewer-Administered Data, Section 3.1	Sections 3.2 and 4.2
Drugs	Report 1: General Principles/Core Drug Data Report 2: Supplementary Self-Administered Data, Section 3.1.1	Section 5.2
Nicotine Dependence	Report 2: Supplementary Self-Administered Data, Section 3.4	(No additional edits discussed)
Immigrant Status	Report 3: Interviewer-Administered Data, Section 3.2.2	Sections 4.2.2 and 4.2.3
Employment	Report 3: Interviewer-Administered Data, Section 3.2.4	Section 4.2.1
Roster	None	Sections 7.2 to 7.4
Health Insurance	Report 3: Interviewer-Administered Data, Section 3.2.5	Section 9.2
Income	Report 3: Interviewer-Administered Data, Section 3.2.6	(No additional edits discussed)
Roster Pair	None	Sections 10.2 to 10.4

MRB = methodological resource book.

[1] Section 10 contains three reports. The first report describes general editing principles and procedures for editing the core substance use data. The second report describes procedures for editing supplementary self-administered interview data. The third report describes procedures for editing interviewer-administered data.

1.1 Organization of this Report

This report was reorganized significantly between 2009 and 2010. The 2011 version follows the organization of the 2010 report.

The PMN imputation methodology is described in detail in Chapter 2. Chapters 3 and 4 describe the imputation procedures applied to the core and noncore demographic variables, respectively. Chapter 3 also describes editing procedures for age, interview date, birth date, gender, race, and Hispanicity. The drug imputation procedures are discussed in Chapter 5. The imputation procedure for nicotine dependence differs from the procedures used for other drug variables and is described in Chapter 6. Chapter 7 describes the edits applied to the household roster, the creation of imputation-revised versions of the roster-derived household composition variables, and the creation of respondent-level variables with individual roster information. Chapter 8 summarizes the editing and imputation procedures applied to the income variables. Procedures for the imputation of missing values in the health insurance variables are described in Chapter 9. The editing and imputation processing of pair relationships and their accompanying multiplicities are detailed in Chapter 10.

This report also contains 11 appendices. Appendix A of this report is identical to that in the 2011 Editing and Imputation Evaluation Report (Scott et al., 2012). It contains a number of tables that quantify the amount of imputation and logical assignment (or editing) that selected analytic variables underwent during imputation processing in 2011. Appendix B provides details on the recoding of "other-specify" responses to some of the demographic questions, so that the data could be summarized in a meaningful way. (Coding of other-specify responses for other variables is summarized by Kroutil and Chien [2013]). The covariates in each of the imputation models are listed in Appendix C. The tables in Appendix C 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 D provides details on each final hot-deck step. The quality control measures used in the imputation procedures are summarized in Appendix E. Reasons that interviewers gave for overriding consistency checks in the household roster are presented in Appendix F, along with evaluations of their legitimacy and the resulting actions in editing the roster. The rules for determining pair relationships are defined in Appendix G. The conditions used for reconciling differing multiplicity counts between pair members are described in Appendix H, and the conditions used for reconciling differing household-level person counts between pair members are described in Appendix I. Appendix J details the priority conditions for creating household-consistent covariates. Appendix K contains detailed information about dwelling unit-level and person-level eligibility and the completeness criteria used to construct the household-level and person-level files.

1.2 Changes from the 2010 Survey to the 2011 Survey

Overall, the changes implemented to the imputation procedures between the 2010 and 2011 surveys were minor. These changes are described in four sections below. The first section describes the impact of an oversample in the Gulf Coast area on editing and imputation. The second section describes changes to the imputation procedures that were implemented to handle unusual cases appearing in the 2011 data. The third section describes two modifications to the imputation procedures that reduced the time required to process the data, with little impact on the final imputation-revised variables. The fourth section lists a few minor corrections and improvements to the imputation procedures that also had little impact on the final imputation-revised variables.

1.2.1 Gulf Coast Oversample

Because there was interest in investigating the effects of the gulf oil spill on substance use and mental health in the region, the NSDUH sample size in 2011 included approximately 2,000 additional cases, concentrated in Alabama, Florida, Louisiana, and Mississippi. Table 1.2 shows the final number of completed NSDUH interviews from 2007 to 2011. The 2011 editing and imputation procedures processed these extra cases in the same manner as all other cases. Because this is a change to the sampling procedures and not to the imputation procedures, no effects from this change are expected to impact estimates in this report. The additional cases likely made it slightly easier to fit the imputation models and slightly easier to find suitable donors for item nonrespondents. For details about the Gulf Coast Oversample, see Morton, Martin, Shook-Sa, Chromy, and Hirsch (2012).

Table 1.2 Number of Completed NSDUH Interviews, 2007-2011

Year[1]	Completed Interviews
2007	67,870
2008	68,736
2009	68,700
2010	68,487
2011	70,109

[1] The number of completed interviews for 2007 to 2010 includes some cases that were later discarded due to data errors. These cases are included in the counts in this table because they underwent editing and imputation processing originally; all the 2007-2010 imputation reports include them in their respective processing counts.

1.2.2 Changes to Accommodate Unusual Cases

In almost every year of the NSDUH, a few cases arise with patterns of response that have not been seen in recent years. Often, the imputation procedures have to be modified slightly to accommodate these unusual cases. In 2011, there were two changes to the imputation procedures that were implemented when unusual cases were encountered. These changes are described below.

There was one case in 2011 involving a pattern of nonresponse in the smokeless tobacco recency and frequency-of-use variables that had not been encountered in previous years. The respondent in this case was known to have used chewing tobacco in the past year, known to have been a past year but not past month user or a past 3 years but not past year user of snuff, and was missing a response for chewing tobacco 30-day frequency. New code was written to handle this case and is documented as missingness pattern 45 in Table D.56.

An edit was added to deal with a 2011 case for hallucinogens recency and frequency. This case reported lifetime nonuse of any hallucinogens other than LSD, PCP, and Ecstasy, but lifetime use of all three of the child drugs. The ages at first use for the parent and child drugs were as follows: 14 for overall hallucinogens, 14 for LSD, missing for PCP, and 15 for Ecstasy. Since the respondent was not a user of any hallucinogens other than LSD, PCP, and Ecstasy, logically the minimum of the three child ages at first use should be equal to the parent age at first use. But, since the age at first use for LSD was equal to the age at first use for overall hallucinogens, there was originally no restriction on the age at first use for PCP. However, the year and month of first use for LSD were later than the year and month of first use for overall hallucinogens. This suggested that the respondent must have used either PCP or Ecstasy before LSD. Since the Ecstasy age at first use was greater than the age at first use for overall hallucinogens, the drug that was used before LSD must have been PCP. Therefore, the respondent was edited to have a PCP age at first use of 14, and a PCP year and month of first use equal to the year and month of first use for overall hallucinogens. This edit is documented in Section 5.2.1.

1.2.3 Changes to Streamline Imputation Procedures

Two changes were made to the 2011 imputation procedures that reduced processing time without significant impact on the results. One was a simple change in the hot-deck step for the race variable. The other was a more involved change to the algorithm for the core-plus-noncore methamphetamine and stimulant variables.

In the hot-deck step for the race variable, sometimes it is difficult to find a donor for certain cases. An example of such a case occurred in the 2009 data, where the respondent was known to be Asian but did not give a specific Asian category (e.g., Chinese, Asian Indian) and reported being of Hispanic/Latino origin and Dominican. In past years, the likeness constraint involving the Hispanic/Latino group was never loosened, so a donor would be found from another age group. This step took extra time and ad hoc programming. Beginning in 2011, the likeness constraint involving the Hispanic/Latino group was dropped as a last resort, and the donor was selected from within the recipient's age group without regard to the Hispanic/Latino group. This change is expected to have no impact on the estimates because only one or two cases a year, at most, are affected.

Imputation-revised lifetime use and recency-of-use variables are created for core-plus-noncore methamphetamine and stimulants (Section 5.3.7). Before 2011, the procedures involved the reimputation of many variables that were not used in subsequent steps anywhere in the NSDUH. Lifetime drug use models were refit for stimulants, sedatives, cocaine, crack, and heroin, and provisional imputations were performed. After these models were refit, all the lifetime use indicators were reimputed using the PMN Type 3 methodology outlined in Section 2.4.3 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 core-plus-noncore 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 core-plus-noncore recency and 12-month frequency variables were imputed in a final hot-deck step that incorporated additional noncore variables as logical constraints. In 2011, the procedures were simplified so that only four core-plus-noncore variables underwent imputation: lifetime use of core-plus-noncore methamphetamine, lifetime use of core-plus-noncore stimulants, recency of use of core-plus-noncore methamphetamine, and recency of use of core-plus-noncore stimulants. As part of the simplification process, the final core-plus-noncore lifetime and recency-of-use variables for stimulants and methamphetamine were defined based on the provisional hot-deck steps. This simplification prevented the need for the imputation of additional variables (i.e., lifetime use indicators for all other drugs and 12-month frequency-of-use variables for core-plus-noncore stimulants and methamphetamine) that were not used in subsequent analyses. An assessment of the impact of this change on the prevalence estimates suggested that the effects would be minimal.

1.2.4 Minor Corrections and Recodes

Five changes were made to the 2011 imputation procedures that involved minor recodes or corrections. None of these have a significant impact on the estimates. Three changes were associated with the income variables, and the other two were associated with the household-level person count variables. These are described below.

The income State-rank variable, detailed in Section 2.5.2, is designed to rank the States based on an estimate of the proportion of the population from families with an income of greater than $20,000 per year. Before 2011, this variable, a covariate used in the income imputation models, was calculated using the edited variable FINC1. In 2011, the edited variable FAMINC1 was used to calculate the State rank instead of FINC1. Both FINC1 and FAMINC1 are variables that identify whether the respondent lived in a household where the total family income was equal to or greater than $20,000. However, the difference between the two variables is that for FINC1, single-person "families" and families with respondents who individually make more than $20,000 per year were assigned a skip code, instead of a value of 1 (family income at least $20,000) or 2 (family income less than $20,000). FAMINC1 does not have any skip codes, and respondents living in single-person households who had previously indicated they made at least $20,000 are logically assigned to have a family income of at least $20,000. When FINC1 was used in the calculation of the State rank, cases with a skip value were incorrectly treated as if the family made less than $20,000 per year. In 2011 processing and beyond, the variable FAMINC1 will be used instead. An impact assessment was done, and the effect on the 2010 imputed income variables was negligible. Because the State-rank variable is only used as a covariate in the imputation models, this correction was expected to also have a negligible impact on the results of the imputation.

During income imputation processing, the set of seven imputation-revised binary income variables (e.g., whether the family received income from a job) are used as covariates in both the response propensity and prediction models for the finer category income variables. In 2010, these were coded as 1 for yes and 2 for no. The result is that they were treated as continuous responses by the models. In 2011, they were coded as 1 for yes and 0 for no, so that they would be treated as discrete responses by the models. As with the change to the State-rank variable, this is a change to the covariates in the imputation models and therefore only indirectly affects the final imputed values. An impact assessment of this change indicated that the impact was low enough that trends would not be affected if the changes were made for 2011.

Before 2011, the domains for the variables "Total Family Income greater than or less than $20,000" and "Total Family Income (Finer Categories)" in Table A.26 did not include all respondents. The domain for the variable "Total Family Income greater than or less than $20,000" was limited to respondents with income less than $20,000 and to respondents with other family members in the household. The domain for the variable "Total Family Income (Finer Categories)" was limited to respondents with other family members in the household. The domains were defined as such because individuals without family members in the household were not asked the questions about their family income, and those with family members in their household, but with a personal income of more than $20,000 per year, were not asked whether their family income was more than $20,000. In general, a skip code means that the question does not apply. However, in this case, the questions are skipped to reduce respondent burden, and the values are logically assigned. For 2011, therefore, the domains were corrected to include "All Respondents." This change only affects Table A.26 of this report. Because there was no change to the imputation processes, no impact assessment was necessary.

In the editing procedures of pair data (i.e., variables related to the presence of two respondents from a sampled dwelling unit), a variable called HHSIZE is created, which stores the number of people in the household and is reconciled across pair members if pair members disagree. This variable, described in Section 10.5.1.1, is used in many of the subsequent pair editing procedures, as a covariate in the imputation models and in constraints in the hot-deck steps. In stage three, where the household-level person counts are created, HHSIZE must be determined for households that included only one respondent. (Only households with responding pairs are processed in stages one and two.)

Before 2011, some households with only one member according to the edited variable TOTPEOP[3] were handled incorrectly with respect to the assignment of HHSIZE. For these single-member households, HHSIZE was assumed to be bad data and was assigned a value equal to the household size according to the screener. In most cases, this value was 1, consistent with the actual household size, so HHSIZE was assigned correctly. However, in cases where the screener listed a value other than 1 for these single-member households, HHSIZE was assigned an incorrect value. As a correction, in 2011, single-respondent households with TOTPEOP = 1 were directly assigned HHSIZE = 1. The impact assessment of this change showed that its effect on the final imputation-revised household-level person counts would be insignificant because (1) fewer than 100 households received an incorrect value for HHSIZE each year; (2) the relationships between HHSIZE and the final imputation-revised household counts is indirect; and (3) little imputation is done for the imputation-revised household counts.

Finally, the method for calculating the imputation indicators was changed for the household-level person counts for parent-child pairs.[4] For parent-child pairs in 2011, the household-level person counts are always equal to the corresponding multiplicity counts. For example, for a parent-child pair where the child is between the ages of 12 and 14, the household-level count of children aged 12 to 14 with at least one parent living with them is the same as the parent-focus multiplicity count that gives the number of children aged 12 to 14 belonging to a parent who is a member of a parent-child pair. Before 2011, for parent-child pairs with a missing household-level person count, the imputation indicator for the household-level person count always indicated statistical imputation (i.e., it had a value of 3). In 2011, the imputation indicator for the household-level person count was corrected to indicate logical assignment (i.e., it had a value of 2) when the corresponding multiplicity was not imputed. Because this correction only affected the imputation indicators and not the imputed values, a formal impact assessment was not conducted. In addition, this change affected so few records that there is no noticeable difference in the number and percentage of imputed or logically assigned cases for the affected variables from 2010 to 2011 (Table A.29).

1.3 How to Use this Report

It is recommended that readers who are unfamiliar with the PMN imputation methodology first read Chapter 2 of this report before reading the more substantive (subtask) chapters detailing particular subsets of variables (e.g., demographics, drugs, income, etc.) that underwent imputation in the 2011 NSDUH. Chapter 2 contains background information about the hot-deck imputation methods employed on the NSDUH before PMN was adopted in 1999 as well as technical details about the PMN methodology itself. This information will help set an appropriate context for readers as they familiarize themselves with the specific editing and imputation procedures that were employed on a particular variable set.

---

[3] TOTPEOP, like HHSIZE, stores the number of people in the household. However, it is not reconciled across pair members. See Section 7.3 for details.

[4] For definitions of "household-level person counts" and "multiplicity counts," refer to Chapter 10.

2. Imputation and the Predictive Mean Neighborhood Methodology

2.1 Introduction

As with most large-scale sample surveys, the respondent datasets for the 2011 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.

Once processed, imputed values cannot be distinguished from nonmissing values 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 2011 NSDUH have the prefix "IR" attached to their names.[5] 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 on the 2011 NSDUH, additional imputation indicators were created with the prefix "II2." These indicators gave more details about the source of the imputed or logically assigned value.

[5] 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.

2.2 Development of the Predictive Mean Neighborhood Methodology

Various methods of imputation have been used since the NSDUH was first administered in the early 1970s.[6] Starting in 1999, the predictive mean neighborhood (PMN) method for imputation was implemented for the NSDUH 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 2.1 provides a summary of the types of imputation procedures used for each of the variables imputed in the NSDUH samples from 1999 through 2011.

2.2.1 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, 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. For more information on the general hot-deck method of item imputation, see Little and Rubin (1987, pp. 62-67).

For the 2011 NSDUH, the only type of hot-deck method used for variables requiring imputation was PMN, described in greater detail later in this chapter. 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 3.5, Section 5.3.3.4, and Chapter 6, 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.[7]

Table 2.1 Summary of Item Imputation Procedure Used, by Variable and NSDUH Year

Variable	1999[1]	2000	2001	2002-2003	2004-2011
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

[6] 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).

[7] 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.

Table 2.1 Summary of Item Imputation Procedure Used, by Variable and NSDUH Year (continued)

Variable	1999[1]	2000	2001	2002-2003	2004-2011
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
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 3).
4 "USHD" refers to the unweighted sequential hot-deck method of item imputation described in this report (see Section 2.2.1.1).
5 "WSHD" refers to the weighted sequential hot-deck method of item imputation described in this report (see Section 2.2.1.2).
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.

2.2.1.1 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[8]—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), 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.

2.2.1.2 Weighted Sequential Hot Deck

The 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 final hot-deck assignment step, however, differs from that in USHD; under WSHD, the sampling weights are incorporated when replacing missing values among recipient records.

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:

---

[8] 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).

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).

2.2.1.3 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.

2.2.2 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). 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[9] 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.
- **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.

[9] Likeness constraints are flexible constraints that govern the similarity between donors and recipients. See Section 2.3.1.3 for details.

2.3 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 2.4), to complete imputation procedures.

2.3.1 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)[10] 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. 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 10 utilize yet another weight, PAIRWT. See Chen et al. (2013) for full descriptions of the preliminary and final analysis weights. See Westlake, Chen, and Gordek (2013) 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 to 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.

[10] 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 Chen et al. (2013).

3. **Item response indicator.** The item response indicator is set to zero for item nonrespondents and to 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 (Chen et al. 2013). Appendix C lists the starting and final covariate lists for each response propensity model fit in NSDUH imputation procedures.

2.3.2 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 8), 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 4), drugs (Chapter 5), and income (Chapter 8).

The goal of any prediction model is good prediction, so these models tend to start with long lists of covariates. Appendix C 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.[11] All covariates from the applicable starter list are utilized unless SUDAAN produces warning messages, which indicate nonconvergence or model instability. In these cases, the standard errors of the regression coefficients are used to make decisions about which covariates to drop from the models; covariates are dropped until SUDAAN no longer produces these warning messages. The primary advantage of using SUDAAN to fit prediction models is that the

[11] The two exceptions are the finer income categories, described in detail in Chapter 8, and the "old method" health insurance, described in detail in Chapter 9.

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 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 are more accurate.

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.

For the types of regression-based prediction models used for each variable that underwent imputation using PMN, see Table 2.2.

Table 2.2 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

Table 2.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood (continued)

Variable	Domain	Type of Regression Model	SAS/SUDAAN Procedure[1,2]
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 Persons	All respondents	Poisson	LOGLINK
Total Number of Children Younger than 18	All respondents	Poisson	LOGLINK
Total Number of Persons 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
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

Table 2.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood (continued)

Variable	Domain	Type of Regression Model	SAS/SUDAAN Procedure[1,2]
Total Income (Binary)	All respondents	Binomial Logistic	RLOGIST
Finer Income Categories	All respondents	Time-to-Event (Survival)	LIFEREG
Health Insurance			
Health Insurance (Old Method)	All respondents	Binomial Logistic	LOGISTIC
Health Insurance (Constituent Variables Method)	All respondents	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 (2008) for more information on all SAS-callable SUDAAN procedures in this table except LIFEREG and LOGISTIC. See SAS Institute Inc. (2004) for more information on the LIFEREG and LOGISTIC procedures.

2.3.3 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[12] of PMN is applied to select a donor for each item nonrespondent. The algorithm used to select donors is graphically displayed in the flow chart in Figure 2.1. 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 from a minimum of 30 candidate donors to supply imputed value(s) for a given recipient. Many of the hot-deck components used in PMN are described below and appear explicitly in the tables of Appendix D.

2.3.3.1 Logical and Likeness Constraints

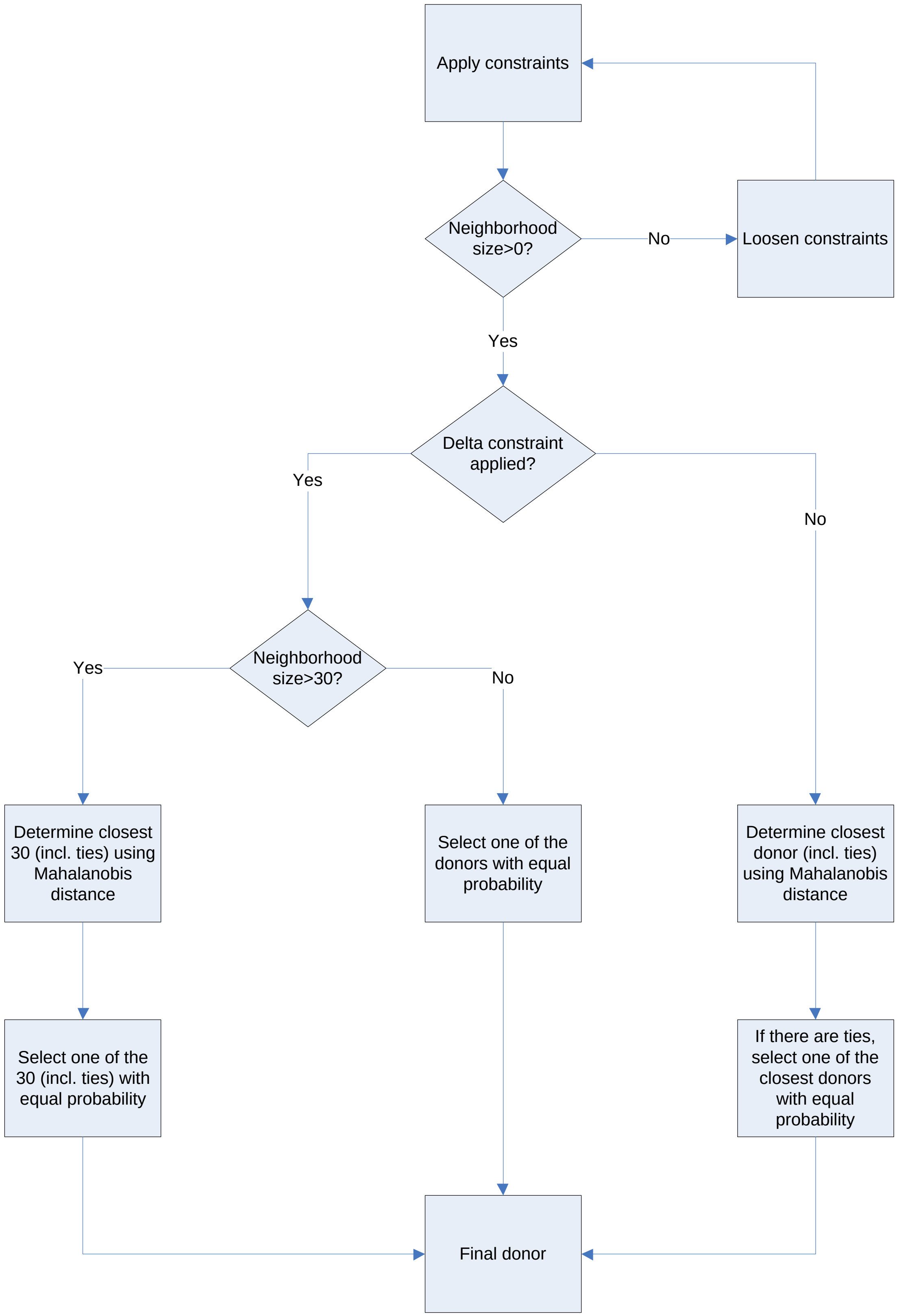
Logical constraints and likeness constraints are restrictions placed on the set of donors to make imputed values consistent with preexisting, nonmissing values of the item nonrespondents (recipients) and to make candidate donors as much like the recipients as possible. Logical constraints are fixed constraints 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.

---

[12] 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 PMN imputation evaluation report by Ault et al. (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 9.3.2.3).

Figure 2.1 Donor Selection Algorithm



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 2.3.

Table 2.3 Values of Delta for Various Predicted Probabilities

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

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

2.3.3.2 Predictive Mean Vector

The predicted means from the prediction step play a central role in the donor selection algorithm depicted in Figure 2.1, 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 D.

2.3.3.3 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 2.4 shows examples of NSDUH variables that were imputed using each of the four combinations of univariate/multivariate matching and assignment.

Table 2.4 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 3.9.3)	Finer Categories Income (Section 8.3)
Predictive Mean Vector Has More Than One Element (Multivariate Matching)	Marital Status (Section 3.9.1)	Lifetime Drug Use (Section 5.3.1)

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. 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.

2.3.3.4 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 D are segregated by missingness pattern for this reason.

2.3.3.5 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 2.1. 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).

2.4 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 2.4.1); Type 2, *multiple RP/multiple PRD* (Section 2.4.2); and Type 3, *single RP/multiple PRD* (Section 2.4.3). Each of the three PMN types is a coordinated application of the three basic steps of PMN discussed in Section 2.3.

In PMN, an imputation "set" is a set of variables for which a single donor is used in the final hot-deck step.[13] 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 2.5 lists the imputation sets for each variable group discussed in this report and the PMN type used to process each set.

Table 2.5 PMN Types Applied to Each Variable Group and Imputation Set

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

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

[13] Section 2.3.2.2 defines and discusses the differences between *provisional* and *final* hot-deck steps in the context of PMN.

2.4.1 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 2.3: 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 2.2 illustrates the single RP/single PRD type of PMN imputation.

Figure 2.2 PMN Type 1: Single Response Propensity/Single Prediction

Usually the single RP/single PRD type involves univariate assignment in the hot-deck step,[14] 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 (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.

2.4.2 Type 2: Multiple Response Propensity/Multiple Prediction

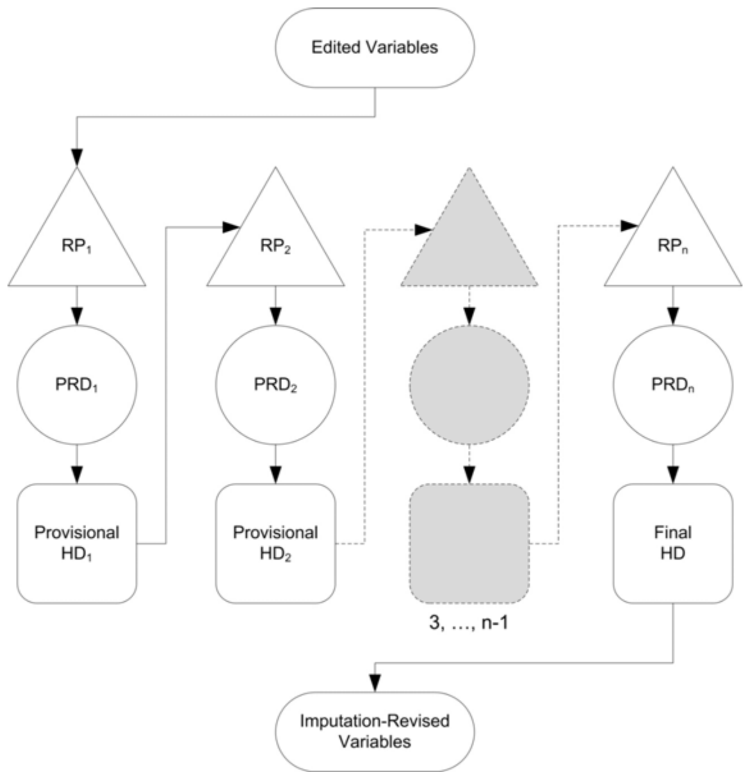
PMN Type 2, multiple RP/multiple PRD, involves multiple iterations of the single RP/single PRD type. However, for all iterations except the last, the hot-deck step is provisional instead of final and involves univariate matching and univariate assignment.[15] These provisional hot-deck steps tend to 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

[14] 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.

[15] 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 5.3.1.5.

be used as covariates in the RP and PRD models for variables later in the sequence.[16] 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 2.3 illustrates the multiple RP/multiple PRD type of PMN imputation.

Figure 2.3 PMN Type 2: Multiple Response Propensity/Multiple Prediction



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

[16] 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 9) and roster pairs (Chapter 10).

(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.

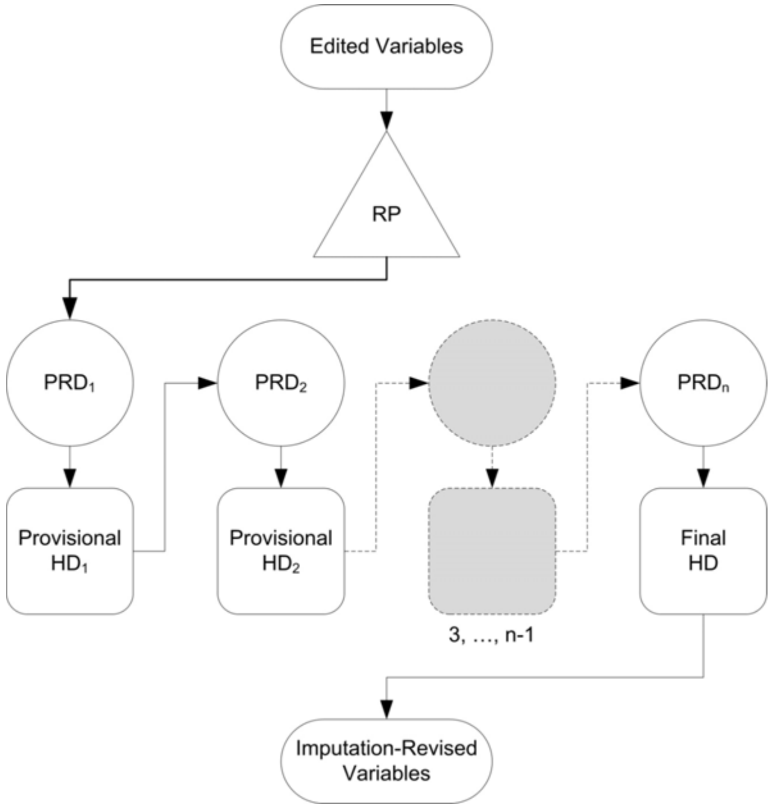
For variables imputed by PMN Type 2 and PMN Type 3 (single RP/multiple PRD), the order in which variables were modeled is of some importance because variables early in the sequence have the potential to be part of the set of covariates for variables later in the sequence, but variables late in the sequence 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 stigma 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 2.3.3.4). 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.

2.4.3 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 2.4 illustrates the single RP/multiple PRD type of PMN imputation.

Figure 2.4 PMN Type 3: Single Response Propensity/Multiple Prediction



2.5 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 3 through 10. 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 5) and the income variables (Chapter 8).

2.5.1 Age Groups

The variables related to drug use, household composition, income, and health insurance were highly correlated with age. This, along with the desire to use parallel processing to expedite implementation, 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.[17] 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.

[17] Age groups were sometimes aggregated for the health insurance procedures. See Chapter 9 for details.

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 facilitated 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 age groups of 12 to 17 and 18 or older. In the employment status models, the 15-to-17 and 18-to-25 age groups were aggregated. Finally, all age groups were aggregated for the Hispanic/Latino group, marital status, and immigrant age-of-entry models.

2.5.2 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.[18] 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.

[18] 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.

3. Imputation for the NSDUH Core Demographics Variables

3.1 Introduction

The 2011 National Survey on Drug Use and Health (NSDUH) included both "core" and "noncore" modules. The core modules remain in the questionnaire every year and are essential for trend measurements and prevalence estimates. Noncore modules follow the core modules and are subject to change. The core demographics variables in the 2011 survey discussed in this chapter include age, birth date, gender, marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level (highest grade completed). Although the interview date was not classified as a core demographic variable, its editing procedures also are included in this chapter. The only noncore demographic variables imputed were the immigrant variables and employment status variables, which are discussed in Chapter 4.

Prior to imputation, editing was performed on all of these core demographics variables. This editing could range from simply assigning legitimate skip codes, as was the case for marital status, to coding other-specify responses and resolving inconsistencies, as was the case for race.

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

- No imputation required: interview date, age, gender. These are described in Sections 3.2.1, 3.2.2, and 3.2.3, respectively. No values were missing after editing.
- Random assignment: birth date. This procedure is summarized in Section 3.2.4 because it is straightforward and does not involve the predictive mean neighborhood (PMN) method, described in Chapter 2.
- PMN: marital status, race, Hispanic/Latino origin, Hispanic/Latino group, education level. These are described in Sections 3.2.5, 3.2.6, and 3.2.7, respectively.

Overall, the core demographics variables discussed in this chapter tend to have high item response rates. Except for race, the item response rates were more than 99 percent. When given the opportunity to enter a race, many respondents entered "Hispanic" or some Hispanic/Latino group such as "Mexican," resulting in a considerable amount of missing data for the race question. As a result, the item response rate for the race variables tends to be about 96 to 97 percent.

3.2 Editing the Selected Core Demographics Variables

In this section, the editing procedures applied to the interview date, age, gender, birth date, marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level variables are described.

3.2.1 Creating the Edited Interview Date Variable (INTDATE)

The time and date of the interview start and completion were automatically saved by the computer-assisted interviewing (CAI) instrument after each questionnaire module was completed. These are referred to as time and date "stamps." In most cases, the time and date stamps were used to determine the interview date. The date stamp indicator (EIIDATE) specified the module date stamp that was used to create the edited interview date (INTDATE).

In some cases, the respondent's birthday occurred between the beginning and the end of the interview. In these cases, 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.

3.2.2 Creating the Edited Age Variable (AGE)

After a respondent had entered his or her birth date in the first part of the questionnaire, he or she had multiple opportunities to change his or her age in response to consistency checks throughout the questionnaire. Therefore, it was possible for the age recorded by the respondent at the beginning of the questionnaire (CALCAGE) to be different from the age captured at the end of the questionnaire (NEWAGE).

The final age variable, AGE, was determined using these two variables 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), and the pre-interview screener age. In most cases, when determining the final edited continuous age, priority was given to CALCAGE, NEWAGE, and the age 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 originally entered raw birth date (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[19] 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 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 the rules above were applied, this intermediate age value was compared with the age corresponding to the "self" in the household roster. In most cases, the final edited value for

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

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 entered in the roster or override the consistency check and provide an explanation as to 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 may not match the intermediate age value described above. However, if the explanations given for overriding the age consistency check were sufficient and other evidence pointed to the veracity of the roster age, and if 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.

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 analysis.

3.2.3 Creating the Edited Gender Variable (IRSEX)

As with previous surveys since 2002, it was mandatory in the 2011 survey that an interviewer enters the respondent's gender in QD01. As a result, it was not possible to have missing values for this question. To maintain continuity with previous surveys (1999-2001), the variable name IRSEX was used to describe gender in the 2011 survey. However, it was not necessary to create an imputation indicator, because IRSEX and QD01 were equivalent.

3.2.4 Creating the Edited Birth Date Variable (BRTHDATE)

To continue with the questionnaire, respondents were required to provide their date of birth and/or current age 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 raw 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 adjusted 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.

3.2.5 Creating the Edited Marital Status Variables (MARITAL, EDMARIT)

In the 2011 questionnaire, a single core question (QD07) asked about the respondent's marital status, among respondents aged 15 or older. The exact phrasing of the question 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

The creation of the edited variable derived from QD07, MARITAL, is described in Kroutil and Chien (2013). 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 skips were collapsed into a single legitimate skip code (99), and missing values were set to the SAS[20] missing code (.) so that they could be properly handled by the modeling programs.

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

In the 2011 questionnaire, two core questions focused on the respondent's ethnicity[21] (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?

3 MEXICAN / MEXICAN AMERICAN / MEXICANO / CHICANO
4 PUERTO RICAN
5 CENTRAL OR SOUTH AMERICAN

[20] SAS® software is a registered trademark of SAS Institute Inc.
[21] 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 CUBAN / CUBAN AMERICAN
7 DOMINICAN (FROM DOMINICAN REPUBLIC)
8 SPANISH (FROM SPAIN)
9 OTHER (SPECIFY)

QD05: Which of these groups describes you?

10 WHITE
11 BLACK / AFRICAN AMERICAN
12 AMERICAN INDIAN OR ALASKA NATIVE (AMERICAN INDIAN INCLUDES NORTH AMERICAN, CENTRAL AMERICAN, AND SOUTH AMERICAN INDIANS)
13 NATIVE HAWAIIAN
14 OTHER PACIFIC ISLANDER
15 ASIAN (FOR EXAMPLE: ASIAN INDIAN, CHINESE, FILIPINO, JAPANESE, KOREAN, AND VIETNAMESE)
16 OTHER (SPECIFY)

QD05ASIA: (Asked only if level 6 of QD05 was selected) Which of these groups describes you?

17 ASIAN INDIAN
18 CHINESE
19 FILIPINO
20 JAPANESE
21 KOREAN
22 VIETNAMESE
23 OTHER (SPECIFY)

As stated in the guidelines from the Office of Management and Budget (OMB),[22] "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.

[22] 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.

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 from the 2000-2002 surveys.[23] 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.

3.2.6.1 Categories Used in Race and Hispanic/Latino Variables

3.2.6.1.1 Race Categories

For editing purposes, the 5 specific categories in QD05 (white, black/African American, American Indian/Alaska Native, Native Hawaiian, and Other Pacific Islander) and the 6 specific categories in QD05ASIA (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese) were combined to produce 11 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 the creation of all the edited and imputation-revised race variables.

3.2.6.1.2 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),[24] 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

---

[23] 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).

[24] 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.

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.

3.2.6.2 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 the categories described in Section 3.2.6.1 and in this section, or to some other code that was informative in the final imputation described in Section 3.3. A summary of categories of other-specify codes and how they were handled is given in the following sections. Appendix B provides the individual other-specify codes and more details about how they were handled.

3.2.6.2.1 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 (see Section 3.2.6). 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.[25] 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

[25] See www.infoplease.com/ipa/A0855617.html (Ethnicity and Race by Countries) for more information.

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: (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 3.3.2. 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.

3.2.6.2.2 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 3.3.2.

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 given (non-other-specify) categories resulted in an unrestricted imputation.

3.2.6.2.3 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 non-other-specify 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-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 given (non-other-specify) categories of QD05 and QD05ASIA, it was necessary to reconcile these responses. In some cases, the combination of responses mapped to one of the multiple racial 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 non-other-specify categories. In particular, the other-specify response was always ignored if a non-other-specify category was selected, and the other-specify response was a geographic category code.[26] 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 persons 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.

3.2.6.2.4 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).

3.2.6.3 Edited Race Variables

3.2.6.3.1 Individual Race Categories (EDQD051-EDQD0513)

Edited variables were created that correspond to the 13 racial categories described in Section 3.2.6.1.1. These variables were called EDQD05*xx*, where *xx* represented a number between 1 and 13, corresponding to each of the 13 categories.

[26] 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 D05ASIA; 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.

EDQD0513 (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 EDQD0513 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.[27]

3.2.6.3.2 Broad Categories of Race (EDRACE)

The EDRACE is a 24-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 EDQD0513 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 EDQD0513 were nonmissing.

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

[27] 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.

- 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 EDQD0513 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 EDQD0513 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 EDQD0513 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 EDQD0513 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 EDQD0513 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 EDQD0513 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 Spanish), if "Puerto Rican" was mentioned in the QD05 and/or QD05ASIA other-specify responses, but QD03 = 2; else
- missing.

3.2.6.3.3 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 16^{th} 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 EDQD0513 was nonmissing,

EDNWRACE =

- 16 (Asian nonspecific only), if EDQD0513 was the nonmissing variable; else
- *xx* (one known racial category only), where EDQD05*xx* was the nonmissing variable out of EDQD051 through EDQD0512.

Scenario 2: If more than one of EDQD051 through EDQD0513 was nonmissing,

EDNWRACE =

- 13 (Native Hawaiian and Other Pacific Islander only), if both EDQD054 and EDQD055 were nonmissing, and all other EDQD05*xx* variables were missing; else
- 14 (Asian multiple category), if all of EDQD051 through EDQD055 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 EDQD0513 were missing,

EDNWRACE =

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

3.2.6.4 Edited Hispanic/Latino Variables

3.2.6.4.1 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 B.

3.2.6.4.2 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 3.2.6.1.2: 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.

3.2.6.4.3 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 3.2.6.1.2, 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.

3.2.7 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 completed by the respondent. No editing was performed on other questionnaire information, and although EDUC contained codes describing the type of nonresponse, EDEDUC was set to missing if no response was given to QD11.

In the 2011 questionnaire, a single core question (QD11) asked about the respondent's education level, in terms of the highest grade that the respondent had completed:

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

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

The creation of the edited variable derived from QD11, EDUC, is described in Kroutil and Chien (2013). The base variable for creating an imputation-revised version of education was called EDEDUC and was equivalent to EDUC, except that missing values were set to the SAS missing code (.) so that they were properly handled by the modeling programs.

3.3 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 2. Each imputation set is discussed in a separate section below.

3.3.1 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 ten missing values in the entire sample in a given survey year.

3.3.1.1 Response Propensity Step

The response propensity model for Imputation Set 1 utilized the preliminary analysis weight, PANALWT, as an input. The marital status question QD07 was only asked of respondents aged 15 or older. Therefore, the domain contained unit respondents with AGE ≥ 15. The creation of the AGE variable is described in Section 3.2.2. Unit respondents in the domain with nonmissing EDMARIT values were considered item respondents. The EDMARIT variable is described in Section 3.2.5. See Table C.1 in Appendix C for details of the covariates used in the RP model for this variable.

3.3.1.2 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.

3.3.1.3 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.

3.3.2 Race Variables (Imputation Set 2)

As discussed in Section 3.2.6, 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 potential confusion among Hispanic/Latino respondents. Practically all of the race nonrespondents reported being of Hispanic/Latino origin (Table 3.1). The likeness constraints involving Hispanic/Latino group strongly influenced the final imputed values.

3.3.2.1 Response Propensity Step

The response propensity models for Imputation Set 2 utilized the preliminary analysis weight, PANALWT, as an input. 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 C.1 in Appendix C for details of the covariates used in the RP models for this variable.

3.3.2.2 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 racial 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 3.3.3.4.

EDRACEFORMODEL is a recode of the variable EDRACE, described in Section 3.2.6.3.2:

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.

3.3.2.3 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 D.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 3.2.6.2.1.

Besides the segment and delta likeness constraints, the likeness constraints based on Hispanic/Latino group were important determinants of the final imputed value, because most of the item nonrespondents for race were Hispanic/Latino. Table 3.1 reports the distribution of Hispanic/Latino group among race item nonrespondents in 2011. Almost all are Hispanic/Latino and most (more than two thirds) of the Hispanic/Latino nonrespondents are Mexican only.

Table 3.1 Hispanic/Latino Status of Item Nonrespondents for Race

Hispanic/Latino Status	Item Nonrespondents for Race	
	Number	Percentage
Not Hispanic/Latino or Missing Hispanic/Latino Indicator	40	1.75
Hispanic/Latino	2,245	98.25
Mexican Only	1,553	67.96
Puerto Rican Only	245	10.72
Central/South American Only	175	7.66
Dominican Only	157	6.87
Other/Unknown	115	5.03

Depending on the missingness pattern, the item nonrespondent received values from the selected donor for some subset of EDRACE, EDNWRACE, EDQD051-EDQD055, and a collapsed version of EDQD056-EDQD0513. The collapsed version of EDQD056-EDQD0513 is simply the minimum of these variables; this 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) receive values for everything except EDNWRACE. The imputation-revised versions of these variables are shown in Table 3.2.

Table 3.2 Edited Race Variables and Their Imputation-Revised Counterparts

Edited Race Variable	Imputation-Revised Race Variable
EDQD051	IRRACEWH
EDQD052	IRRACEBK
EDQD053	IRRACENA
EDQD054	IRRACENH
EDQD055	IRRACEPI
EDQD056-EDQD0513 (collapsed)	IRRACEAS
EDRACE	IRDETAILEDRACE
EDNWRACE	IRNWRACE

IRDETAILEDRACE is the only imputation-revised variable that 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 3.3.4 below.

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.

3.3.3 Hispanic/Latino Origin Variables (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 3.2.6.4.1 above. The item response rate for this variable was much higher than for race.

3.3.3.1 Response Propensity Step

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

3.3.3.2 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.[28] The outcome variable was EDHOIND. The single predicted mean was the predicted probability that the respondent was of Hispanic/Latino origin.

3.3.3.3 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 D.3 through D.5 in Appendix D.

3.3.3.4 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 2011 detailed tables (Center for Behavioral Health Statistics and Quality, 2012).

3.3.4 Hispanic/Latino Group Variables (Imputation Set 4)

The edited variable EDHOGRP, described in Section 3.2.6.4.3, 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.

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

3.3.4.1 Response Propensity Step

The response propensity models for Imputation Set 4 utilized the preliminary analysis weight, PANALWT, as an input. 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 selected only a single Hispanic/Latino group. The multiple-group respondents whose EDHOGRP was assigned by the priority rule (Scenario 3, described in Section 3.2.6.4.3) were not used to fit the PRD model in the next step. See Table C.1 in Appendix C for details of the covariates used in the RP models for these variables.

3.3.4.2 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 3.3 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 3.3 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

3.3.4.3 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 do not generally require much imputation. The number of missing cases is usually fewer than 100 in each survey year.

3.3.4.4 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.

3.3.5 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.

3.3.5.1 Response Propensity Step

The response propensity models for Imputation Set 5 utilized the preliminary analysis weight, PANALWT, as an input. 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 C.1 in Appendix C for details of the covariates used in the RP models for this variable.

3.3.5.2 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-to-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 3.4.

Table 3.4 Mapping of EDEDUC Levels to RESPEDUC Levels

	RESPEDUC (Modeled Variable)	
EDEDUC (Base Variable)	**12-17**	**18+**
Never attended school	Less than elementary school	Less than high school
1st grade completed		
2nd grade completed		
3rd grade completed		
4th grade completed		
5th grade completed		
6th grade completed	Elementary school	
7th grade completed		
8th grade completed	Middle school	
9th grade completed		
10th grade completed	Some high school	
11th grade completed		
12th grade completed	High school	High school
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-to-17 hot-deck step and four for the 18-to-25 and 26-or-older hot-deck steps.

3.3.5.3 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-to-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.

3.3.5.4 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). This variable was often used as a covariate in later imputation models.

4. Imputation for the NSDUH Noncore Demographics Variables

4.1 Introduction

The 2011 National Survey on Drug Use and Health (NSDUH) included both "core" and "noncore" modules. The core modules remain in the questionnaire every year and are essential for trend measurements and prevalence estimates. Noncore modules follow the core modules and are subject to change. For the noncore demographics module of the 2011 NSDUH, three imputation-revised variables were created. The first was an employment status variable, EMPSTATY; 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. 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. The core demographics variables that were imputed in the 2011 survey are discussed in Chapter 3.

The variables describing current employment status were asked only of respondents aged 15 or older and were based on multiple questions in the noncore demographics module. Instead of a single question asking the respondent to describe his or her "current" employment status, several questions were asked regarding the respondent's employment situation during the week preceding the interview. A single imputation-revised variable, EMPSTATY, was then created from the series of employment status questions using the single response propensity (RP)/single prediction (PRD) type of the predictive mean neighborhood (PMN) method, which is described in Chapter 2. 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.

The two imputation-revised variables used to produce immigrant status (IRBORNUS and IRENTAG2) were created using the edited variables BORNINUS and ENTRYAG2 as base variables.[29] Like EMPSTATY, each was processed independently using the single RP/single PRD type of PMN. These variables recorded whether a respondent was born in the United States (BORNINUS), and if not, the variables recorded the age of entry (ENTRYAG2) into the United States. 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.

In the sections that follow, details are provided on the editing and imputation procedures applied to the three noncore demographics variables in the 2011 NSDUH.

[29] Although these variables are called "immigrant status" variables for convenience purposes, the immigrant questions also included information from eligible respondents who lived in the United States and were not born in the United States, but had no intention of staying permanently in the United States (e.g., foreign students are not immigrants). For this reason, respondents who indicated that they were not born in the United States are called non-U.S.-born respondents in this chapter.

4.2 Editing the Noncore Demographics Variables

Most of the editing procedures associated with the noncore demographic variables (employment status, born-in-U.S. indicator, and age of entry) are described by Kroutil and Chien (2013). Descriptions in this section are mostly limited to the creation of base variables for imputation from the edited variables described in Kroutil and Chien (2013).

4.2.1 Base Variable 1: Employment Status (EDEMPY)

The base variable EDEMPY, which was used to create the imputation-revised employment status variable EMPSTATY, was derived from JBSTATR and the edited variable WRKHRSUS. JBSTATR combined information from QD26, QD27, QD29, QD30, QD31, QD32, and QD33. The categories for JBSTATR are shown in Table 4.1. WRKHRSUS was an edited variable created from QD29, which asks, "Do you **usually** work 35 hours or more per week at **all** jobs or businesses?" WRKHRSUS was used in some cases to determine whether employed respondents were employed full time or part time. Both variables are described in more detail in Kroutil and Chien (2013).

Table 4.1 Categories of JBSTATR

Code	Employment Situation	Code	Employment Situation
1	Worked at full-time job, past week	**12**	No job: in school/training
2	Worked at part-time job, past week	**13**	No job: retired
3	Has job but out: vacation/sick/temp absence	**14**	No job: disabled for work
4	Has job but out: layoff, looking for work	**15**	No job: didn't want a job
5	Has job but out: layoff, not looking for work	**190**	Has full-time job, reason for not working unknown
6	Has job but out: waiting to report to new job	**191**	Has part-time job, reason for not working unknown
7	Has job but out: self-employed, no business past week	**199**	Has job, no further information
8	Has job but out: in school/training	**290**	No job, no further information
9	No job: looking for work	**299**	Other, not in labor force
10	No job: layoff, not looking for work	Remaining codes in the 900 series have their standard meanings in NSDUH: Don't know (994), Refused (997), Blank (998), Legitimate skip (999)	
11	No job: keeping house full time		

The base variable EDEMPY was derived from JBSTATR and the edited variable WRKHRSUS 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.

4.2.2 Base Variable 2: Born-in-U.S. Indicator (BORNINUS)

All respondents were asked in QD14 whether they were born in the United States (excluding U.S. territories). Responses were limited to "yes" or "no," and if the response was "no," the respondent was asked to name the country of origin in QD15. The edited variable BORNINUS was created using the responses to QD14. As part of the standard editing procedures, if the interviewer entered a U.S. State in QD15, the "no" in QD14 was overwritten with a logically assigned "yes." Other levels of BORNINUS were standard NSDUH missing data codes corresponding to "don't know," "refused," or "blank." More details about editing procedures are provided in Kroutil and Chien (2013).

4.2.3 Base Variable 3: Age of Entry (ENTRYAG2)

Beginning with the 2004 survey, respondents have been given the choice to write in the amount of time they had lived in the United States in years (in QD16b) or in months (in QD16c), depending upon their answer to QD16a (asking if they had lived in the United States for at least 1 year). The edited variables associated with QD16a, QD16b, and QD16c were called LIVUS1YR, LIVUSMOS, and LIVUSYRS, respectively. A legitimate skip code was assigned to LIVUSMOS if the respondent had lived in the United States for 1 year or more (LIVUS1YR = 1). Similarly, a legitimate skip code was assigned to LIVUSYRS, if the respondent had lived in the United States for less than 1 year (LIVUS1YR = 2). Codes for "don't know," "refused," "blank," and "bad data" also were applied to these variables at the editing stage. More editing details on these three variables are described by Kroutil and Chien (2013).

To compute the age at which a non-U.S.-born respondent entered the United States, the continuous form of the respondent's age and length of time living in the United States was produced for all non-U.S.-born respondents. Because QD16b and QD16c were designed to be mutually exclusive, the edited variables LIVUSMOS and LIVUSYRS were combined to create a continuous variable (LNGTHLIV) that indicated how many years a non-U.S.-born 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. The variable was set to missing when LIVUSYRS and LIVUSMOS had missing data codes. CONTAGE, the continuous age variable, was defined as CONTAGE = (interview date – birth date + 1)/365.25. Because interview date and birth date, as described in Chapter 3, 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.

The variable ENTRYAG2 is the base variable for creating the imputation-revised variable IRENTAG2 and represents 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. It also had a legitimate skip code (999) for respondents who were born in the United States.

4.3 Imputation for Noncore Demographics Variables

In this section, the imputation procedures applied to the three base variables EDEMPY, BORNINUS, and ENTRYAG2 are described. These variables comprised three independent manifestations of the single RP/single PRD type of PMN. Each single-member imputation set is discussed in a separate section below.

4.3.1 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 the respondent is known to be employed but part-time versus full-time status is not known. 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.

4.3.1.1 Response Propensity Step

The response propensity models for Imputation Set 1 utilized the preliminary analysis weight, PANALWT, as an input. 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 C.1 in Appendix C for details of the covariates used in the RP models for this variable.

4.3.1.2 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.

4.3.1.3 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, that donor's age must be within 5 years of recipient's age, was applied in the hot-deck step. See Appendix D for more details on missingness patterns and constraints for employment status.

4.3.1.4 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).

4.3.2 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.

4.3.2.1 Response Propensity Step

The response propensity models for Imputation Set 2 utilized the preliminary analysis weight, PANALWT, as an input. The domain indicator for the RP model included all unit respondents. Item respondents were those with a nonmissing value for BORNINUS. See Table C.1 in Appendix C 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 born-in-U.S. indicator was modeled using dichotomous logistic regression as implemented by the LOGISTIC procedure in

SUDAAN.[30] The outcome variable was BORNINUS. The single predicted mean was the predicted probability that the respondent was born in the United States.

4.3.2.3 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 D.

4.3.3 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.

4.3.3.1 Response Propensity Step

The response propensity model for Imputation Set 3 utilized the preliminary analysis weight, PANALWT, as an input. 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 C.1 in Appendix C for details of the covariates used in the RP model for this variable.

4.3.3.2 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 4.2.3. The expression of the proportion is $P_i = Y_i/N_i$, where $Y_i = \textit{Age at Entry}_i$ and $N_i = \textit{Continuous Age}_i$ (CONTAGE).

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

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

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

This transformation was nearly equivalent to the standard logit transformation,

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

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

4.3.3.3 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 Appendix D.

5. Imputation for the NSDUH Drug Variables

5.1 Introduction

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 2011 National Survey on Drug Use and Health (NSDUH) drug data. Consistent with prior years, the drug use measures collected in the 2011 NSDUH included lifetime usage, recency of use, frequency of use in the past 12 months, frequency of use in the past month, and age, year, and month of first use. However, depending on the drug in question, only a subset of these measures were collected and imputed. Table 5.1 summarizes the drugs and drug use measures that were imputed and indicates how these measures were segregated into imputation sets.

Note that some of the rows in Table 5.1 refer to a general drug category (e.g., hallucinogens) and one or more subcategories (e.g., LSD, PCP, and Ecstasy). These drug categories are described using the terms "parent drug" for the general drug category and "child drug" for the drug subcategory. For a drug to be considered a child drug, data must have been gathered on some combination of recency, frequency, and age at first use.

Parent-child drug pairs sometimes occurred in modules that included multiple gate questions (called "subgate" questions).[31] For example, the hallucinogens module includes numerous questions to determine the respondent's lifetime use or nonuse of hallucinogens. Included in these questions are specific "subgate" questions for LSD (LS01a), PCP (LS01b), and Ecstasy (LS01f). Parent-child drug pairs could also appear in separate modules (e.g., questions on cocaine and crack were included in two separate modules).

The parent-child drug combinations described in this chapter include smokeless tobacco (parent) and chewing tobacco and snuff (child); cocaine (parent) and crack (child); hallucinogens (parent) and LSD, PCP, and Ecstasy (child); pain relievers (parent) and OxyContin (child); and stimulants (parent) and methamphetamine (child). Smokeless tobacco differs from the other parent drugs in that respondents were not directly asked about smokeless tobacco; rather, respondents were asked only about its two child drugs (chewing tobacco and snuff). Any measures of smokeless tobacco, therefore, can be considered to be recoded variables of imputed variables because they were not directly imputed. Similarly, OxyContin differs from other child drugs in that there was no subgate question focusing solely on the child drug OxyContin. Instead, OxyContin was one of 18 types of pain relievers, which appeared both in question PR04a and on a card shown to the respondents by the interviewers when the respondents reached these questions. Table 5.2 shows all the drugs in parent-child relationships and the data that were gathered on them.

[31] A "gate" question is one that asks whether the respondent ever used the drug in his or her lifetime. The term is used because an affirmative response opens the "gate" to a series of other questions on use of the drug; a negative response leads to the skipping of all other questions on use of the drug.

Table 5.1 Drugs and Drug Use Measures, Imputation Sets

<table>
<tr><th rowspan="2">Drug</th><th colspan="6">Drug Use Measure</th></tr>
<tr><th>Lifetime Usage</th><th>Recency of Use</th><th>12-Month Frequency of Use</th><th>30-Day Frequency of Use</th><th>Age at First Use</th><th>Age at First Daily Use</th></tr>
<tr><td>Cigarettes</td><td>NM</td><td colspan="3">Set 2 (12-Month Frequency N/A)</td><td>Set 3</td><td>Sets 4, 5*</td></tr>
<tr><td>Smokeless Tobacco[1]</td><td rowspan="13">Set 1</td><td colspan="3">Set 6 (12-Month Frequency N/A)</td><td>Set 7</td><td rowspan="14">N/A</td></tr>
<tr><td>Cigars</td><td colspan="3">Set 8 (12-Month Frequency N/A)</td><td>Set 9</td></tr>
<tr><td>Pipes</td><td colspan="3">Set 10 (12-Month and 30-Day Frequency N/A)</td><td>N/A</td></tr>
<tr><td>Alcohol</td><td colspan="3">Set 11[5]</td><td>Set 12</td></tr>
<tr><td>Inhalants</td><td colspan="3">Set 13</td><td>Set 14</td></tr>
<tr><td>Marijuana</td><td colspan="3">Set 15</td><td>Set 16</td></tr>
<tr><td>Hallucinogens[2]</td><td colspan="3">Set 17</td><td>Set 18</td></tr>
<tr><td>Pain Relievers[3]</td><td colspan="3">Set 19 (30-Day Frequency N/A)</td><td>Set 20</td></tr>
<tr><td>Tranquilizers</td><td colspan="3">Set 21 (30-Day Frequency N/A)</td><td>Set 22</td></tr>
<tr><td>Stimulants[4]</td><td colspan="3">Set 23 (30-Day Frequency N/A)</td><td>Set 24</td></tr>
<tr><td>Sedatives</td><td colspan="3">Set 25 (30-Day Frequency N/A)</td><td>Set 26</td></tr>
<tr><td>Cocaine and Crack</td><td colspan="3">Set 27</td><td>Set 28</td></tr>
<tr><td>Heroin</td><td colspan="3">Set 29</td><td>Set 30</td></tr>
<tr><td>Core-Plus-Noncore Stimulants and Methamphetamine</td><td>Set 31</td><td colspan="3">Set 32 (12-Month and 30-Day Frequency N/A)</td><td>N/A</td></tr>
</table>

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.

Table 5.2 Drugs in a Parent-Child Relationship

Parent Drug	Child Drugs	Parent Data Collected	Child Data Collected	"Other" Lifetime Use Indicator[1]
Smokeless Tobacco	Chewing Tobacco, Snuff	None	Recency, 30-day frequency, age at first use	No
Hallucinogens	LSD, PCP, Ecstasy	Recency, 12-month frequency, 30-day frequency, age at first use	Recency, age at first use	Yes
Pain Relievers	OxyContin	Recency, 12-month frequency, age at first use	Recency, 12-month frequency, age at first use	Yes

Table 5.2 Drugs in a Parent-Child Relationship (continued)

Parent Drug	Child Drugs	Parent Data Collected	Child Data Collected	"Other" Lifetime Use Indicator[1]
Stimulants	Methamphetamine	Recency, 12-month frequency, age at first use	Recency, 12-month frequency, age at first use	Yes
Cocaine	Crack	Recency, 12-month frequency, 30-day frequency, age at first use	Recency, 12-month frequency, 30-day frequency, age at first use	No

[1] See Section 5.2.1.

5.2 Editing the Drug Variables

The majority of the editing procedures applied to the drug use variables are summarized in Kroutil, Handley, and Bradshaw (2013) and Kroutil, Handley, Bradshaw, Chien, and Felts (2013). However, additional edits were also applied during the imputation process and are discussed below. In general, these edits affected only a few records. They were implemented mostly to resolve inconsistencies, which prevented the determination of a valid interval for the assignment of the date of first use (see Section 5.3.3.4).

5.2.1 "Other" Hallucinogens, "Other" Pain Relievers, and "Other" Stimulants Variables

The lifetime use question was not asked directly for hallucinogens, pain relievers, and stimulants. Instead, these modules all included subgate questions on lifetime use for their associated child drugs (i.e., for hallucinogens: LSD, PCP, and Ecstasy; for pain relievers: OxyContin; and for stimulants: methamphetamine). In addition to the subgate questions about specific child drugs, these modules also included subgate questions for many other drugs that fall under the parent drug category but are not one of the specific child drug variables analyzed for the NSDUH.[32] An "other" category was created by combining responses to all the subgate questions for these other nonchild drugs. Lifetime use of the parent drugs was then determined by combining any affirmative responses in the "other" category recode with any affirmative responses in the applicable child drug subgate questions for that parent.

1. The lifetime usage indicator for "other" hallucinogens was created using the lifetime usage information from all the hallucinogens subgate questions except LSD, PCP, and Ecstasy. If a respondent was a user of at least one of the other hallucinogens, then he or she was considered a user of other hallucinogens, even if some of the other hallucinogens subgate questions were unanswered. A missing value for other hallucinogens arose if at least one of the other hallucinogens subgate questions was unanswered and all the other hallucinogens subgate questions that were answered had a negative response.
2. The procedure for pain relievers was similar to the procedure used for hallucinogens. The major difference is that there was no subgate question focusing solely on the

[32] Peyote is one such example of an "other" (i.e., nonchild) type of hallucinogen, because recency, frequency, and age-at-first-use data are *not* collected for this substance.

child drug OxyContin. Specifically, OxyContin was one of 18 types of pain relievers, which appeared both in question PR04a and on a card shown to the respondents by the interviewers when the respondents reached these questions. Respondents could have selected any number of drugs listed on the card. A lifetime usage indicator for "other" pain relievers was created using information from all the pain relievers subgate questions, except the OxyContin item in PR04a. As with hallucinogens, a respondent's other pain relievers lifetime usage indicator was missing only if the subgate questions, other than the item that dealt with OxyContin, were all unanswered or if these questions were a combination of unanswered (i.e., "refused" and "don't know") questions and "no" responses.

3. The procedure for stimulants was almost identical to the procedure used for pain relievers. However, there was a specific subgate question on the child drug methamphetamine. Three lifetime usage indicators were created: one for "other" stimulants, one for methamphetamine, and one for stimulants.

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 (described below) were met. 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 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 age-at-first-use value was nonmissing, only one child age-at-first-use value was missing, and the minimum of the nonmissing child age-at-first-use values was greater than the parent age-at-first-use value, then the missing child age-at-first-use value was edited to the parent age-at-first-use value.
4. If the respondent was known never to have used "other" hallucinogens, the parent age-at-first-use value was nonmissing, only one child age-at-first-use value was missing, the minimum of the nonmissing child age-at-first-use values was equal to the parent age-at-first-use value, and the earliest of the nonmissing child months and years of first use was later than the parent month and year of first use, then the missing child age-at-first-use value was edited to the parent age-at-first-use value.[33]

[33] 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.

5.2.2 Respondents Imputed to Lifetime Use of Child Drugs Variables

As shown in Table 5.1, 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 raw 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. This edit only applied to OxyContin, methamphetamine, and crack, which are the only child drugs with 12-month frequencies. 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.
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 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).

5.2.3 Age-at-First-Use Variables

1. This edit applied to all parent age-at-first-use variables: cigarettes, overall hallucinogens, overall pain relievers, overall stimulants, and cocaine. If the parent age-at-first-use value was missing and the minimum of the child age-at-first-use values was 3 years, then the parent age-at-first-use value was edited to 3 years. This was done because respondents with age-at-first-use values of less than 3 years were ineligible to be donors (see Section 5.3.3.1).
2. This edit applied to all child age-at-first-use variables: daily cigarettes, LSD, PCP, Ecstasy, OxyContin, methamphetamine, and crack. If the parent age at first use was

equal to the respondent's current age, all missing child age-at-first-use values were edited to the same age.

3. This edit also applied to all child age-at-first-use variables. If the parent age at first use was equal to 1 less than the respondent's current age, the child recency[34] was lifetime but not past year (or, for cigarettes, past 3 years but not past year), and the child age-at-first-use value was missing, then the child age-at-first-use value was assigned to 1 less than the respondent's current age. This was done because the child age at first use cannot be less than AGE – 1, because the parent age at first use 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 age at first use cannot be less than AGE – 1 or greater than AGE – 1, it must be equal to AGE – 1.

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.

5.3 Imputation of the Drug Variables

The drug variables that undergo imputation are segregated into imputation sets, as shown in Table 5.1. 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 5.3.2, and the rest of the recency and frequency sets are described in Section 5.3.5 as deviations from the procedures described in Section 5.3.2.

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.

5.3.1 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 2.4.3. 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.

[34] Because there was no recency question associated with daily cigarettes, the overall cigarette recency was used instead.

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 5.1, with the exception of lifetime cigarette use.[35] 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.

5.3.1.1 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 1999–2010 surveys, the 2011 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. Based on the response to each gate question, the instrument either routed the respondent through the current drug module or skipped the respondent to the next module entirely. Thus, the respondent was not necessarily required to answer all questions in the questionnaire; the respondent could have skipped a module if he or she either indicated nonusage of the drug in the gate question or did not answer the gate question. Therefore, the gate question response was crucial to the range of responses available for subsequent questions in each module. Because these gate questions were asked of all respondents, the domain for the lifetime use indicators includes all respondents to the 2011 survey.

For certain categories of drugs, multiple gate questions within a drug module were used to assess lifetime use or nonuse of the overall group of drugs within that module (e.g., LSD, PCP, Ecstasy, and a number of other substances within the drug module for hallucinogens were used to assess usage of hallucinogens). For these drug groups, if any of the gate questions were answered "yes" (i.e., the respondent indicated using the drug once or more in his or her lifetime), then the lifetime use indicator for the overall drug group was set to "yes." For example, to assess lifetime use of the overall drug group "inhalants," the respondent was asked 11 different questions on whether he or she had ever, even once, inhaled any of the following with the intention of getting high: (1) amyl nitrite, "poppers," locker room odorizers, or "rush"; (2) correction fluid, degreaser, or cleaning fluid; (3) gasoline or lighter fluid; (4) glue, shoe polish, or toluene; (5) halothane, ether, or other anesthetics; (6) lacquer thinner or other paint solvents; (7) lighter gases, such as butane or propane; (8) nitrous oxide or "whippets"; (9) spray paints; (10) some other aerosol spray; and (11) any other inhalants. If the response to any of these questions was "yes," the respondent was deemed a lifetime user of inhalants, even if some of the other responses to the gate questions in the inhalants module were unanswered. Similarly, composite lifetime indications of use were formed for hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, and smokeless tobacco. To be considered a lifetime nonuser of a drug module with multiple gate questions, the respondent had to answer "no" to all of the gate questions. If none of the gate questions in a drug module was answered affirmatively, but some

[35] See Section 2.4.2 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.

of the gate questions were unanswered, then the individual was considered a nonrespondent for that module.

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 C.3 in Appendix C for details of the covariates used in the RP models for these variables.

5.3.1.2 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 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[36] to determine the parameter estimates and probability of use for both respondents and nonrespondents.

5.3.1.3 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), 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.[37] 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

[36] 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 (2008). SAS software is a registered trademark of SAS Institute Inc. SUDAAN is a registered trademark of Research Triangle Institute.

[37] Those with a missing lifetime use indicator for the drug were treated as lifetime nonusers.

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.

5.3.1.4 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:

- 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 2011 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.[38] 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.

5.3.1.5 Final Hot-Deck Step

Tables D.42 through D.45 in Appendix D 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

[38] 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.

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 subgate 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.

5.3.2 Imputation-Revised Cigarette Recency and Frequency of Use (Imputation Set 2)

As indicated in Table 5.1, 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.

5.3.2.1 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.

5.3.2.2 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 person was deemed an item respondent for cigarettes. Otherwise, he or she was an item nonrespondent. See Tables C.4, C.5, and C.6 in Appendix C for details of the covariates used in the RP models for this variable.

5.3.2.3 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.

5.3.2.4 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), 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.[39] 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 5.3. Logical constraints ensured that only donors with allowable specific recency categories were included in the neighborhood.

Table 5.3 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

[39] Those individuals whose past month use status was unknown were treated as if they were not past month users.

5.3.2.5 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 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 C.4, C.5, and C.6 in Appendix C for details of the covariates used in the RP models for this variable.

5.3.2.6 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.[40] 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

[40] 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.

subsequent hot-deck step. The logit was back-transformed into a proportion before use in the hot-deck step.

5.3.2.7 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 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 5.4, 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, 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 D.50. 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 D.50.

Table 5.4 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.

5.3.2.8 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).

5.3.2.9 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).

5.3.3 Imputation-Revised Cigarette Age at First Use (Imputation Set 3)

As indicated in Table 5.1, the third imputation set consisted of the cigarette age-at-first-use 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 age-at-first-use 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 age at first use) were asked the year and month of their first use. To have this information for all users, both missing year and missing month of first use for less recent initiates (and recent initiates who did not report year and month of first use) were replaced by assigning values consistent with the respondent's current age, interview date, imputation-revised age at first use, and imputation-revised recency and frequency variables. To have complete date-of-first-use information, day of first use was randomly assigned for all users. The combined data gave the respondent's age at first use along with the date of first use.

5.3.3.1 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 age at first use for cigarettes, 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 age-at-first-use measure,[41] the person was deemed an item respondent. See Tables C.4, C.5, and C.6 in Appendix C for details of the covariates used in the RP models for this variable.

5.3.3.2 Prediction Step (Cigarette Age at First Use)

The response variable in the model for age at first use, before a normalizing transformation, was the age at first use as a proportion of the current age. The numerator in this proportion was an integer representing age at first use. However, because this integer was in fact a truncated version of the real age at first use, 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

[41] 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.

$$Y_i = Age\ at\ First\ Use_i + Uniform(0,1)\ random\ number.$$[42]

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 age at first use was a modified 30-day frequency variable for cigarettes. It was defined as follows:

$\mathbf{new30_i}$	=	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 age at first use 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 C.4 through C.6 in Appendix C.

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 age at first use.

5.3.3.3 Hot-Deck Step

The imputation-revised cigarette age-at-first-use assignment was conducted using a single predicted mean: the predicted age at first use. Tables D.77 through D.79 provide a complete list of likeness and logical constraints applied to the cigarette age at first use imputations. The likeness constraints for age at first use 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.

[42] 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 [$Age\ at\ First\ Use_i$, N_i]. Thus, Y_i was prevented from being greater than N_i.

5.3.3.4 Date-of-First-Use Assignments

After the age-at-first-use imputations, all lifetime users of cigarettes had nonmissing age-at-first-use values. Using this age at first use (AFU), users were assigned values for year, month, and day of first use. Recent initiates, or those respondents whose AFU was within 1 year of his or her age, were asked for their year of first use (YFU) and month of first use (MFU). The day of first use (DFU) was not collected in the questionnaire and was missing for all respondents. The YFU, MFU, 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.

5.3.3.4.1 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 month and year of first use, in a way that was consistent with both the 30-day frequency/recency and age at first use. 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.

5.3.3.4.2 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 age-at-first-use 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 1st of the YFU."
- For the *latest possible date*, replace "last day of the month indicated by MFU/YFU" with "December 31st of the YFU."

5.3.3.4.3 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 year/month/day of first use were randomly assigned from a uniform distribution in a way that was consistent with both the cigarette 30-day frequency/recency and the age at first use. Again, the imputed YFU, MFU, and DFU were derived in the same manner as described in Missingness Pattern 1.

5.3.3.4.4 Missingness Pattern 4

The fourth missingness pattern occurred when the respondent reported, or was imputed to, an age at first use at least 2 years less than his or her age. This case is analogous to data prior to the 1999 survey, where month and year of first use 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.

5.3.3.5 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[43] and the final three letters identified the measure (AGE = age at first use, YFU = year of first use, MFU = month 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 year and month of first use.

5.3.4 Imputation-Revised Age at First Daily Cigarette Use (Imputation Sets 4 and 5)

In addition to age at first use, 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 age at first use, with some key exceptions as discussed below.

One such exception involved the domain of the age-at-first-use variable. Whereas the age-at-first-use 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

[43] 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 5.3.4.

respondent had a "yes" value for 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 Kroutil, Handley, and Bradshaw (2013) and Kroutil et al. (2013). 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 5.1, 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.

5.3.4.1 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[44] 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 person was deemed an item respondent. See Tables C.4, C.5, and C.6 in Appendix C for details of the covariates used in the RP models for this variable.

5.3.4.2 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.

5.3.4.3 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 D.46 and D.47.

5.3.4.4 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.

[44] 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 age at first use 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 age at first use. 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 age-at-first-use 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 cigarettes age at first use. 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 age at first use and age at first daily use were both initially missing, the imputed age at first use was set back to missing and reimputed simultaneously with the age at first daily use so that they were mutually consistent.[45]

Subject to these constraints, the age at first daily use of the randomly selected donor was then transferred to the recipient.

5.3.4.5 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 5.3.3.4 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.

5.3.4.6 Final Variables

The final imputation-revised date-of-first-daily-cigarette-use variables were named in the same manner as described in Section 5.3.3.5. 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.

5.3.5 "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 5.1 (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.

---

[45] Though it has occurred in prior years, the situation where no donors were available, even after loosening all constraints, did not occur in the 2011 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.

5.3.5.1 Recency of Use

5.3.5.1.1 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 5.2), 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 C.7 through C.36 and C.40 through C.48 in Appendix C for details of the covariates used in the RP models for these variables.

5.3.5.1.2 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.

5.3.5.1.3 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 5.3. 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.

5.3.5.1.4 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[46] 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[47] 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(past month use | lifetime use) = P(past month use | past year use) × P(past year use | lifetime use), and

P(past year, not past month use | lifetime use) = P(past year, not past month use | past year use) × P(past year use | lifetime use).

5.3.5.2 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.

[46] 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."

[47] The set of covariates used for these dichotomous logistic models were the same as those for logistic modeling given earlier in this section.

5.3.5.2.1 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 C for details of the covariates used in the RP models for these variables.

5.3.5.2.2 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 month-at-first-use question, the difference between the month of first use and the date of the interview indicated the total time period during which they could have been using drugs.[48] If the date of the interview was July 10, for example, and the month of first use 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 users.[49]

[48] 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.

[49] 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.

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.[50] 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.[51] Appendix C 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.

5.3.5.2.3 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, month of first use, 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), whether that recency of use was reported or imputed.[52] 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

---

[50] For item nonrespondents, where parameter estimates were used to determine predicted means, past year use was defined based on a provisional imputation.

[51] 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.

[52] 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.

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.

5.3.5.2.4 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 5.1).

5.3.5.3 30-Day Frequency of Use

5.3.5.3.1 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 those drug modules that asked the respondent to report this measure (Table 5.1). 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.[53] See Appendix C for details of the covariates used in the RP model for these variables.

5.3.5.3.2 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.

5.3.5.3.3 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).

A logical constraint required that the donated 30-day frequency was less than or equal to the respondent's preexisting 12-month frequency and greater than or equal to the respondent's preexisting 30-day binge drinking frequency. 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.

5.3.5.3.4 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.

5.3.5.4 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.

[53] 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.

5.3.5.4.1 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 C.16, C.17, and C.18 in Appendix C for details of the covariates used in the RP model for this variable.

5.3.5.4.2 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.[54] Using the adjusted weights, a linear univariate regression model was then fitted for the log-transformed variable Y_i within each age group.

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.

5.3.5.5 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 D provides

[54] 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.

detailed information on these hot-deck steps, and Table 5.5 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 5.5 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]
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

Table 5.5 Elements of Full Predictive Mean Vector (continued)

Drug Use Measure and Category of Interest	Predicted Mean	Substance
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.

5.3.5.6 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.

5.3.5.7 Recodes for Additional Analyses

Section 5.3.2.8 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 5.6 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 5.6 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.

5.3.6 "Other" Drugs Age at First Use

The age-at-first-use variables for Imputation Sets 7, 9, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30 in Table 5.1 (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 these "other" drugs age at first use as described below.

5.3.6.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the age-at-first-use variables for other drugs was the preliminary analysis weight, PANALWT. The RP step for age at first use for other drugs was very similar to the RP step for cigarettes age at first use (Section 5.3.3.1). 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 C provides a complete list of covariates used in each model to properly adjust the weights.

5.3.6.2 Prediction Step

The PRD step for age at first use for other drugs was also very similar to the analogous step for cigarettes (Section 5.3.3.2). 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

5.3.6.3 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 age-at-first-use variable was associated with a single predicted age at first use. 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 age at first use for smokeless tobacco was obtained by taking the minimum age at first use from chewing tobacco and snuff. Respondents were never asked directly for their age at first use for smokeless tobacco.
- For other substances with child drugs, respondents were asked for the age at first use for the parent drug and were also asked for their age at first use 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.

5.3.6.4 Year-of-First-Use, Month-of-First-Use, and Day-of-First-Use Assignments

The general principles described in Section 5.3.3.4 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 5.2). 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.

5.3.6.5 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, YFU = year of first use, MFU = month of first use, DFU = day of first use). Again, each IR variable was accompanied by an imputation indicator with the requisite II prefix.

5.3.6.6 Recodes for Additional Analyses

Section 5.3.5.6 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 5.6).

The age-at-first-use recodes were simply set to the minimum of the source age-at-first-use 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).

5.3.7 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, age at first use, 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. These versions of the methamphetamine variables were presented in a special section in the 2011 detailed tables (Center for Behavioral Health Statistics and Quality, 2012) but not in the main tables showing the standard list of drugs. For more information on the reporting of methamphetamine prevalence in the 2011 NSDUH, see Section B.4.8 of Appendix B of the 2008 national findings (Office of Applied Studies, 2009). New imputation-revised variables were created using the new questions in the noncore section of the questionnaire on 12-month frequency.

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 core-plus-noncore 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 2.4.3 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 core-plus-noncore 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 core-plus-noncore 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. The simpler method for lifetime use was run a single time. Only four cases received different imputed values for stimulants, and only one case received a different imputed value for methamphetamine. Given the high item response rates for these variables (see Table A.23) and the few differences, the impact was determined to be trivial. For recency of use, the simpler method was run three times, and the weighted proportion of respondents in each recency category was calculated for each run, for each of three age groups (12 to 17, 18 to 25, and 26 or older). The largest difference seen in the proportions was 0.0006. Because the impact was so low and the time saved using the simpler method was substantial, the simpler method was adopted.

5.3.7.1 Final Creation of Base Variables for Imputation

The edited recency-of-use variables MTHREC06 and STMREC06, created by the editing team, were used as a starting point for the final creation of the base variables for imputation. These variables are described in Kroutil et al. (2013). 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.

5.3.7.2 Reimputation of Lifetime Use Indicators (Imputation Set 31)

Using EDMTHLIFE and EDSTMLIFE, the processing of the lifetime use indicators proceeded, as described in Section 5.3.1. 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 5.1, 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.

5.3.7.3 Reimputation of Recency of Use

Using EDMTHREC and EDSTMREC instead of METHREC and STIMREC, the processing of the recency data proceeded, as described previously.[55] Final recency-of-use variables for methamphetamine and stimulants were created.

5.3.7.4 Recodes for Additional Analyses

In the manner described in Sections 5.3.2.8 and 5.3.5.6, some prevalence recodes were created that incorporated information from the noncore special drugs module. The core-plus-noncore methamphetamine recodes were CPNMTHFG, CPNMTHYR, and CPNMTHMN. The core-plus-noncore stimulants recodes were CPNSTMFG, CPNSTMYR, and CPNSTMMN. The core-plus-noncore psychotherapeutic recodes were CPNPSYFG, CPNPSYYR, and CPNPSYMN. No core-plus-noncore versions of the IEM or SUM recodes described in Table 5.6 were created for use in the detailed tables, even though the prevalence estimates would likely increase slightly if the noncore methamphetamine data were incorporated.

[55] Other core-plus-noncore edited variables also were used in these reimputation steps, in logical constraints of hot-deck steps.

6. Imputation for the NSDUH Nicotine Dependence Variables

6.1 Introduction

The 17 questionnaire items used to determine nicotine dependence and the methods used to measure nicotine dependence in the 2011 National Survey on Drug Use and Health (NSDUH) were first introduced in the 2001 survey and have been used in every survey since. As in 2010, only respondents who reported use of cigarettes in the past 30 days were asked the dependence questions in the 2011 survey.

The method for determining nicotine dependence involved calculating a continuous score from the Nicotine Dependence Syndrome Scale (NDSS) (Shiffman, Hickcox, Gnys, Paty, & Kassel, 1995; Shiffman, Waters, & Hickcox, 2003). As indicated in Table 6.1, the score was calculated from 17 of the 19 nicotine dependence questions that were asked of respondents who used cigarettes in the past 30 days. For details on how the NDSS was calculated, see Section B.4.2 in Appendix B of the 2009 summary of national findings (Office of Applied Studies, 2010a; 2010b).

Of the 344 eligible respondents in 2011 who did not answer all 17 NDSS questions, the majority (190) were missing a response for only one of the items (Table 6.2). Any respondent with more than one of the 17 items missing (44.8 percent of the eligible cases with incomplete data) did not have his or her missing responses replaced with imputed values, and no nicotine 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.

The imputations described in this chapter are unique in this report because they were not performed using the predictive mean neighborhood (PMN) methodology described in Chapter 2. The NDSS mean value was calculated from imputation-revised versions of the 17 nicotine dependence questionnaire variables. The majority of the editing procedures performed on these variables are described in Kroutil, Handley, and Bradshaw (2013).

6.2 Editing the Nicotine Dependence Variables

Table 6.1 shows the correspondence between the 17 raw variables used in the NDSS and their edited counterparts. The edited variables serve as the base variables for imputation. Among eligible respondents (those who had used cigarettes in the past 30 days), valid responses for the edited variables were as follows: 1 = Not at all true; 2 = Somewhat true; 3 = Moderately true; 4 = Very true; or 5 = Extremely true.

Table 6.1 Mapping of Raw Nicotine Dependence Question Variables to Edited Variables

Raw Variable	Question Text	Edited (Base) Variable
DRCGE01	After not smoking for a while, you need to smoke in order to feel less restless and irritable.	CIGIRTBL
DRCGE02	When you don't smoke for a few hours, you start to crave cigarettes.	CIGCRAVE
DRCGE03	You sometimes have strong cravings for a cigarette where it feels like you're in the grip of a force you can't control.	CIGCRAGP
DRCGE04	You feel a sense of control over your smoking – that is, you can "take it or leave it" at any time.	CIGINCTL
DRCGE05	You tend to avoid places that don't allow smoking, even if you would otherwise enjoy them.	CIGAVOID
DRCGE07	Even if you're traveling a long distance, you'd rather not travel by airplane because you wouldn't be allowed to smoke.	CIGPLANE
DRCGE08	You sometimes worry that you will run out of cigarettes.	CIGRNOUT
DRCGE09	You smoke cigarettes fairly regularly throughout the day.	CIGREGDY
DRCGE10	You smoke about the same amount on weekends as on weekdays.	CIGREGWK
DRCGE11	You smoke just about the same number of cigarettes from day to day.	CIGREGNM
DRCGE12	It's hard to say how many cigarettes you smoke per day because the number often changes.	CIGNMCHG
DRCGE13	It's normal for you to smoke several cigarettes in an hour, then not have another one until hours later.	CIGSVLHR
DRCGE14	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.	CIGINFLU
DRCGE15	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.	CIGNOINF
DRCGE16	Since you started smoking, the amount you smoke has increased.	CIGINCRS
DRCGE17	Compared to when you first started smoking, you need to smoke a lot more now in order to be satisfied.	CIGSATIS
DRCGE18	Compared to when you first started smoking, you can smoke much, much more now before you start to feel anything.	CIGLOTMR

6.3 Creating the Imputation-Revised Nicotine Dependence Variables

For respondents who used cigarettes in the past 30 days and provided complete data for all 17 of the nicotine dependence questions used to calculate the NDSS scale value, imputation-revised nicotine dependence variables were simply assigned the values from their base variable counterparts.

For respondents who had used cigarettes in the past 30 days and gave a valid response to exactly 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 our model, and the remaining 16 NDSS variables would serve as our 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.[56]

Respondents who had used cigarettes in the past 30 days but answered 15 or fewer of the nicotine dependence questions were left out of the modeling process entirely. The missing values in the NDSS variables for these respondents remained missing in the imputation-revised variables derived from them.

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. However, the ratio-adjusted, design-based weights were applied throughout.

6.4 Summary Information for Nicotine Dependence Variables

Imputations were necessary for the nicotine dependence variables to create an NDSS score for as many eligible persons as possible. The imputation method was devised to be easy to implement, given the complexities of handling this type of missing data. To avoid complicated models, imputations were limited to the 190 cases in the 2011 NSDUH where the respondent answered exactly 16 of the 17 questions. If an eligible respondent answered fewer than 16 questions, a situation that applied to 154 cases in the 2011 survey, no imputations were performed. Note that it was possible that the respondent was eligible to answer the questions about nicotine dependence because he or she was imputed to be a past month cigarette user. This situation occurred 10 times in the 2011 NSDUH.

Table 6.2 summarizes the eligibility of respondents to answer the nicotine dependence questions and reasons why respondents were classified as eligible or not eligible. Furthermore, this table provides details about the amount of nicotine dependence data that was 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 nicotine dependence data but would have missing data for all the nicotine dependence variables.

[56] 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.

Table 6.2 Summary of Response Patterns for 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]	54,121	77.20
N/A	No	Yes	No	N/A[1]	11	0.02
Subtotal					54,132	77.21
0	Yes	No	Yes	No	17	0.02
0	Yes	Yes	Yes	No	10	0.01
1-15	Yes	No	Yes	No	127	0.18
Subtotal					154	0.22
16	Yes	No	Yes	Yes	190	0.27
17	Yes	No	Yes	N/A[2]	15,633	22.30
Total					70,109	100.00

N/A = not applicable; NDSS = Nicotine Dependence Syndrome Scale.

[1] The respondent was not eligible to answer the NDSS questions.

[2] The NDSS variables for this scenario were not missing.

7. Imputation for the NSDUH Roster Variables

7.1 Introduction

This chapter describes the techniques used to edit and impute variables associated with the household roster for the 2011 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 persons 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 7.1 shows an example grid where the number of persons 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 7.2.

Table 7.1 Roster Grid Example Where Number of Persons in Household (QD54) Equals 4

Person #	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 7.2 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 7.3, which also shows that some of these variables underwent imputation. Note that among the three groups of variables described in this chapter, only the roster-derived household composition variables underwent imputation of any kind.

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), who was aged 18 years or older and who answered the health insurance coverage and income questions for the respondent. The edited versions of these variables and the questions to which they map are shown in Table 7.4.

Table 7.3 Roster-Derived Household Composition Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Total number of rostered persons	TOTPEOP	IRHHSIZE
Number of persons in household aged 17 or younger	KID17	IRKID17
Number of persons 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

Table 7.3 Roster-Derived Household Composition Variables (continued)

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
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
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.

Table 7.4 Proxy Variables

Raw Variable	Text of Survey Question Associated with Raw 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

7.2 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, ROSSEX1-ROSSEX25, ROSRLT1-ROSRLT25, ROSMSL1-ROSMSL25, PRNTYP1-PRNTYP25, SIBTYP1-SIBTYP25, CHDTYP1-CHDTYP25, and TWNTYP1-TWNTYP25. 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 7.2.1 describes the consistency checks programmed into the questionnaire. Section 7.2.2 describes the creation of a roster-level dataset for further processing. Section 7.2.3 describes roster edits involving the self. Section 7.2.4 describes roster edits for other household members, after the self has been established. Finally, Section 7.2.5 describes the creation of the final edited respondent-level detailed roster variables.

7.2.1 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 legitimacy are provided in Appendix F.

7.2.1.1 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 legitimacy are provided in Appendix F. 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 3. Strategies for the more common situation, where the original value for AGE was not set to the roster age, are discussed in Section 7.2.3.

7.2.1.2 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.

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 7.2.4. Explanations given by the interviewers for the overrides and evaluations of their legitimacy are provided in Appendix F.

7.2.2 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.

7.2.3 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 3). However, it was possible to have problems matching AGE (see Chapter 3 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 7.2.1.1). 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 an age and gender 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.[57] 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 7.1, 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

[57] 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.

cases, the original relationship codes were salvaged, depending upon the roster member who was used as a reference point.

7.2.3.1 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 reference person had been changed.

7.2.3.2 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.

7.2.4 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 7.2.1. 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.

7.2.4.1 Edits to Roster Age, Gender, and Relationship Codes: Changes to Different Values (Correct Reference Person)

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 still was believable even with the incorrectly entered age (e.g., "other relative"), then no inconsistency check was triggered and this type of error was difficult to detect. On the other hand, if an age entered this way triggered one of the consistency checks discussed in Section 7.2.1.2, 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 3) 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 3) 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. Because of the introduction of consistency checks (consistency checks #2 and #3 listed in Section 7.2.1.2), this did not occur in the 2011 survey. 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.

7.2.4.2 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 7.2.4.1 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 7.2.4.1 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 7.2.1.2.
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 7.2.1.2.
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 7.2.1.2. 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 7.2.1.2. 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 7.2.1.2.
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 7.2.1.2.
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 7.2.1.2).
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 check #11 listed in Section 7.2.1.2.
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 7.2.1.2.

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 7.2.1.2.
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 7.2.1.2.
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 7.2.1.2.
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 7.2.1.2.
16. The roster member was the respondent's sibling and the previous roster member[58] 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 7.2.1.2. 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 7.2.1.2. If the age difference was less than 30 years, this was covered by consistency check #12 in Section 7.2.1.2.
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.

7.2.4.3 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: Illogical Child Code)

In Section 7.2.4.2, illogical relationship codes were set to bad data. Often, this occurred because the interviewer used someone other than the respondent as the reference person 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.

[58] A "previous roster member" is the member who immediately precedes the member of interest in the roster.

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 7.2.4.2). 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 7.2 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 (see Section 7.2.4.2). 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[59] 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 7.2), then no uncles or aunts lived in the household. If a pair was 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.

7.2.4.4 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: Illogical Spouse Code)

The interviewer also could have used an incorrect reference person 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

[59] 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.

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. It should be noted that many of these edits were covered by consistency checks #4 and #5 listed in Section 7.2.1.2, 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 7.2.1.2.
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 7.2.1.2.

7.2.4.5 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: 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 7.2.1.2.

7.2.4.6 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: 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 7.2.1.2. 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 7.2.1.2.

7.2.4.7 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: 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 reference person. 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 7.2), 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.

7.2.5 Final Edited Respondent-Level Detailed Roster Variables

The raw 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 7.1 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 7.2 were incorporated into the values of the detailed roster variables, called 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 7.3.

7.3 Editing the 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 7.3. 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 7.5. 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 edit that of the other pair member.

The respondent's household size was assumed to equal the total number of rostered persons in the household, TOTPEOP, as shown in Table 7.3. 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 persons in the household aged 17 or younger) and HH65 (number of persons 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 7.5.2. 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 a little 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 7.5.2.

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 7.3. 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 7.4.

7.4 Editing the Proxy Variables

This section describes the creation of edited proxy variables, as listed in Table 7.4. Section 7.4.1 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 7.4.2 and 7.4.3 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 7.4 shows the correspondence between the five 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 7.5.

Table 7.5 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
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

7.4.1 Edited Indicator of Potential Proxies in Household (EDFAM18)

As described in Section 7.3, 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.

7.4.2 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 raw 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 raw 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.

7.4.3 Editing the Proxy Variables when EDFAM18 = 0

If EDFAM18 was 0, the proxy variables were edited as follows:

1. If the raw proxy variables had legitimate nonmissing values (i.e., not replaced by a logically assigned legitimate skip), the edited proxy variables (except PRXRELAT) were set to those nonmissing values.
2. If any of the raw proxy variables (except PRXRELAT) had a value of 2 ("no"), then all of the variables that followed were edited to legitimate skips.
3. If any of the raw proxy variables 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 raw 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).

7.4.4 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. 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. Once the proxy roster was established, the number selected in QP02 was matched to the corresponding person in the proxy roster.

7.5 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[60] 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 2.4.1, was used to impute nonmissing values among these eight variables in the order shown in Table 7.3. 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 7.5.1 describes the imputation process applied to the first edited variable, TOTPEOP. Section 7.5.2 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 7.5.2 will only list divergences from the process that was applied to the TOTPEOP variable.

7.5.1 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 7.5.1.1 describes the RP step, Section 7.5.1.2 describes the PRD step, and Section 7.5.1.3 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).

7.5.1.1 Response Propensity Step

The response propensity model for Imputation Set 1 utilized the preliminary analysis weight, PANALWT, as an input. 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 Tables C.49, C.50, C.51, and C.52 in Appendix C for details of the covariates used in the RP models for this variable.

7.5.1.2 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

[60] An imputation set is a set of variables for which a single donor is used in the final hot-deck step. Chapter 2 describes this concept more fully.

software.[61] The single predicted mean used in the subsequent hot-deck step was the predicted number of people in the household.

7.5.1.3 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 D.

7.5.2 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 7.3: 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 C.49, C.50, C.51, and C.52 in Appendix C provide details of the covariates used in the RP models for these variables.

[61] Details about the LOGLINK procedure are discussed and additional references are provided in the *SUDAAN Language Manual, Release 10.0* (RTI International, 2008).

8. Imputation for the NSDUH Income Variables

8.1 Introduction

As with most of the imputation-revised variables discussed in the previous chapters of this report, imputations for the income variables in the 2011 National Survey on Drug Use and Health (NSDUH) were performed using the predictive mean neighborhood (PMN) methodology detailed in Chapter 2. The edits applied to the income variables prior to imputation are described in Kroutil and Chien (2013).

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 Chapter 2. The second imputation set for finer income categories 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, also described in Chapter 2.

As had been the case in the 2008–2010 surveys, separate questions to ascertain personal-level and other-family-level responses for the binary income variables were not asked in the 2011 survey, nor were there separate questions covering child support, interest/investment income, and other income. However, since 2008, respondents have been asked questions about binary and finer category actual annual income at both the personal and family levels. A comparison between the 1999–2007 and 2008–2011 sets of income questions asked is shown in Table 8.1. 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.

The income variables have lower item response rates compared with most of the other variables that undergo imputation. This is especially true of the variables for total personal income, total family income, and number of months on welfare. See Table A.26 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).

Table 8.1 **Comparison between 1999-2007 and 2008-2011 Sets of Income Questions**

Income Questions Included in NSDUH	**1999-2007 Surveys***			**2008-2011 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
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 2008-2011 surveys. For detailed explanations of these subsamples, consult the 2006 and 2007 Methodological Resource Books.

8.2 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 8.2.1. The PRD model for the first binary income variable imputed each year, family social security (FAMSOC), is described in Section 8.2.2, and the provisional hot-deck step for FAMSOC is described in Section 8.2.3. The remaining PRD and provisional hot-deck steps described in Section 8.2.4 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 8.2.5. Finally, a recode for the GOVTPROG variable, made from four imputation-revised binary income variables, is described in Section 8.2.6.

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 imputed is listed in Table 8.2.

Table 8.2 Imputation Order for Income Variables in the Binary Variable Phase and Edited Family Income Response Variables Used in Predictive Mean Models

Income Type	Variable Name
Family Social Security	FAMSOC
Family Supplemental Security Income	FAMSSI
Family Welfare Payments	FAMPMT
Family Other Welfare Services	FAMSVC
Family Wages	FAMWAG
Family Food Stamps	FSTAMP
Family Months on Welfare	WELMOS
Total Family Income[1]	FAMINC1

[1] The model for total family income used all of the variables above as covariates except the variable indicating months on welfare.

8.2.1 Response Propensity Step

The response propensity models for the binary income variables utilized the preliminary analysis weight, PANALWT, as an input. 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 (1) all of the variables listed in Table 8.2 were nonmissing, and (2) PINC1 and FINC1 were nonmissing. PINC1 is the person-level version of FAMINC1. FINC1 is the family-level version of FAMINC1, except that it has a skip code for respondents who have no other family members in their household. For more information on the variables PINC1 and FINC1 as well as on all other variables listed in Table 8.2, see Kroutil and Chien (2013). See Table C.53 in Appendix C for details of the covariates used in the RP models for these variables.

8.2.2 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, it 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.

8.2.3 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 2.4.2 describes the concept of a provisional hot-deck in more detail. The final hot-deck step is described in Section 8.2.5 and in tabular form in Tables D.95, D.96, and D.97 in Appendix D. 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 7.5.2 (LogC1 in Table D.95).
  - Donor's predicted mean must be within 5 percent of recipient's predicted mean (delta constraint; LikC3 in Table D.96).
  - 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 7.5.2 (LikC5 in Table D.96).
- 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.

8.2.4 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 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,[62] 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 on 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 D.96 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

[62] The Cox empirical logit was used when a person was on welfare for all 12 months.

Section 7.5.2. This likeness constraint was used in both the first and second attempts to find a donor.

- For FAMWAG, LikC7 in Table D.96 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 7.5. 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 4.3.1. 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.

8.2.5 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 D. 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 D.97. The remaining 384 cover patterns where WELMOS is missing; these are summarized as patterns 1 through 4 in Table D.97.

- 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.[63]
- 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 8.3). 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, 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,[64] and there is no need

[63] 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.

[64] 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 2).

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 both is one minus the probability of receiving neither, as expressed by $(1 - \text{PMT}) \times (1 - \text{SVC})$.[65]
- 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 8.3 correspond to pattern 1 in Table D.97, combination 4 corresponds to pattern 2, combination 5 corresponds to pattern 3, and combination 6 corresponds to pattern 4.

Table 8.3 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 D.97. 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

[65] 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.

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 of this chapter, the income variables have low item response rates compared with most of the other variables that undergo imputation. This is especially true of the variables for total personal income, total family income, and number of months on welfare. The constraints are fairly complex and extensive, especially when several of the other dichotomous income variables are missing. Also, 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.

The imputation-revised variables created by the final hot-deck program included imputation-revised versions of the eight variables listed in Table 8.2, plus imputation-revised versions of PINC1 and FINC1. Table 8.4 lists the imputation-revised personal and family income variables.

Table 8.4 Imputation-Revised Personal and Family Income Variables

Income Model	Variables
Social Security	IRFAMSOC
Supplemental Security Income	IRFAMSSI
Welfare Payments	IRFAMPMT
Welfare Services	IRFAMSVC
Wages	IRFAMWAG
Food Stamps	IRFSTAMP
Welfare Months	IRWELMOS
Total Family Income	IRPINC1, IRFINC1, IRFAMIN1

8.2.6 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 9 for details on the imputation of missing values in the health insurance variables.

8.3 Imputation for Finer Category Income Variables (Imputation Set 2)

Three income variables resulted from editing the questions in the finer income category 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 Kroutil and Chien (2013). All three imputation-revised variables derived from their edited counterparts were created using the single RP/single PRD PMN type described in Chapter 2. 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.

8.3.1 Response Propensity Step

The response propensity models for the finer category income variables utilized the preliminary analysis weight, PANALWT, as an input. All respondents were in the domain for this imputation set. Item respondents were those with nonmissing values for PINC2, FINC2, and FAMINC2. See Table C.58 in Appendix C for details of the covariates used in the RP models for these variables.

8.3.2 Prediction Step

Each of the three finer income category 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.[66] 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 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, 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 $\mathbf{X}\boldsymbol{\beta}$, 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, $\mathbf{X}\boldsymbol{\beta}$ was a translation of the estimated mean of log income.

8.3.3 Hot-Deck Step

The hot-deck step for the finer income category variables is an example of univariate matching, multivariate assignment (see Table 2.4 in Chapter 2). The only predicted mean used in

[66] Details about the LIFEREG procedure are discussed in the *SAS/STAT 9.1 User's Guide* (SAS Institute Inc., 2004).

the hot-deck step was related to FAMINC2, but the recipient also received values for PINC2 and FINC2 if either or both were missing. The imputation-revised versions of PINC2, FINC2, and FAMINC2 are called IRPINC2, IRFINC2, and IRFAMIN2. 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 income category 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.

8.3.4 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 9 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.

9. Imputation for the NSDUH Health Insurance Variables

9.1 Introduction

For the 2011 National Survey on Drug Use and Health (NSDUH), as has been the case since 2002, 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 imputes for three overall health insurance variables (e.g., any health insurance or any private insurance coverage) in a way that is consistent with iterations of the NSDUH questionnaire prior to 2002. The constituent variables method imputes for specific health insurance variables (e.g., Medicaid or Medicare coverage) and was first implemented for the health insurance module in 2002. As with other variable groups (i.e., demographics, drugs, etc.), the health insurance variables were imputed in sets, as follows:

- Imputation Set 1: IRINSUR, IRINSUR3, and IRPINSUR (Old Method)
- Imputation Set 2: IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT (Constituent Variables Method)
- Imputation Set 3: IROTHHLT (Constituent Variables Method)

Under the old method, three health insurance variables comprise Imputation Set 1. The first variable, IRPINSUR, was simply an imputation-revised version of the edited "private health insurance" variable, PRVHLTIN, detailed in Kroutil and Chien (2013). The second and third variables, IRINSUR and IRINSUR3, were both indicators of "any health insurance" coverage. These different versions of health insurance coverage indicators were created because the question set changed between 1999 and 2001. See section 9.2.1 for more detail on the creation of these variables.

Under the constituent variables method, indicator variables for more specific types of health insurance (i.e., coverage by Medicare or by Medicaid) were imputed in two additional sets. Imputation Set 2 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. Within Imputation Set 3, the "any other health insurance" variable, IROTHHLT, was created. Together, these five constituent imputation-revised health insurance variables were then used to create a third indicator of "any health insurance" coverage, IRINSUR4.

The constituent variables 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 2 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.

Regardless of whether the final health insurance variables were derived by the old method or the constituent variables method, imputations were performed using the same methodology, the PMN method. The health insurance variables that undergo imputation tend to have item response rates of more than 99 percent. See Chapter 2 for details on PMN; see Table A.27 in Appendix A for details on rates of missingness for the health insurance variables.

9.2 Editing the Health Insurance Variables

Under the old method, the three base variables (INSUR, INSUR3, and PINSUR) that undergo imputation are simple recodes created from six "source" variables, each of which maps to a single question in the NSDUH. In the 2011 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.

Under the constituent variables method, the six source variables are the same as those used in the old method, but they are not recoded and combined prior to being imputed. With the exception of MEDICAID and CHIPCOV, the variables remain separate to form the base variables for Imputation Sets 2 and 3.

Section 9.2.1 discusses the creation of base variables for imputation under the old method. These are the base variables used in Imputation Set 1. Section 9.2.2 discusses the creation of base variables for imputation under the constituent variables method. These are the base variables used in Imputation Sets 2 and 3.

9.2.1 Health Insurance Variables, Old Method (Imputation Set 1)

Under the old method, the three base variables (INSUR, INSUR3, and PINSUR) were created from the six source variables shown in Table 9.1. INSUR and INSUR3 indicated whether the respondent had any health insurance, and PINSUR indicated whether the respondent had any private health insurance. INSUR, which was created to maintain consistency with the 1999 survey, was coded as "yes" if any of MEDICARE, MEDICAID, CHAMPUS, or PRVHLTIN were coded as "yes"; it was coded as "no" if all four were coded as "no." In 2001, the questions associated with CHIPCOV and HLTINNOS were added to the questionnaire. The variable INSUR3 was created beginning with the 2001 survey to incorporate the addition of these two new questionnaire items. INSUR3 was coded as "yes" if any one of the six variables listed in Table 9.1 were coded as "yes"; it was coded as "no" if all six variables were coded as "no."[67] PINSUR was a direct recode of PRVHLTIN.

[67] In the 2000 survey, the variable INSUR2 was created to take advantage of the additional information provided by questions that did not exist in the 1999 questionnaire. However, because these additional questions were either replaced or reworded in later surveys, the variable INSUR2 has not been used in the surveys since 2000.

Table 9.1 Mapping of Raw Health Insurance Variables to Base Variables for Imputation Set 1 (Old Method)

Question Number[1]	Question Text[2]	Source Variable[3]	Used to Create INSUR	Used to Create INSUR3	Used to Create PINSUR
QHI01, QHI01v	Is the respondent covered by Medicare?	MEDICARE (1 = yes, 2 = no)	Yes	Yes	No
QHI02, QHI02v	Is the respondent covered by Medicaid or Medical Assistance?	MEDICAID (1 = yes, 2 = no)	Yes	Yes	No
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)	No	Yes	No
QHI03	Is the respondent currently covered by CHAMPUS or TRICARE, CHAMPVA, the VA, or military health care?	CHAMPUS (1 = yes, 2 = no)	Yes	Yes	No
QHI06	Is the respondent currently covered by private health insurance?	PRVHLTIN (1 = yes, 2 = no)	Yes	Yes	Yes
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 skip[5])	No	Yes	No

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 9.2.2 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.

9.2.2 Health Insurance Variables, Constituent Variables Method (Imputation Sets 2 and 3)

Under the constituent variables method, the editing process for the health insurance imputation process 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 9.1, except HLTINNOS, were used directly as base variables for imputation.

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 9.1. 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. Thus, the ANYOTHER variable was created using HLTINNOS and an additional edited variable (SKHLCCOV), which indicated whether a respondent was covered by any health insurance. SKHLCCOV and ANYOTHER were defined as follows:

SKHLCCOV =

- 1 (or 3[68]) if CAIDCHIP = 1, MEDICARE = 1, CHAMPUS = 1, or PRVHLTIN = 1; else
- 2 if CAIDCHIP = 2, MEDICARE = 2, CHAMPUS = 2, and PRVHLTIN = 2; else
- missing value code if the nonmissing values of CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN are all "2," and at least one of these variables had a missing response.

ANYOTHER =

- legitimate skip code (99) if SKHLCCOV = 1 or 3; else
- SKHLCCOV if SKHLCCOV = 2 or a missing value code.

9.3 Imputation for Health Insurance Variables (Imputation Sets 1 through 3)

Section 9.3.1 describes the creation of the three imputation-revised health insurance variables for the old method (Imputation Set 1). Section 9.3.2 describes the creation of the first four imputation-revised health insurance variables under the constituent variables method (Imputation Set 2). Section 9.3.3 describes the creation of the fifth and final imputation-revised health insurance variable under the constituent variables method (Imputation Set 3).

9.3.1 Health Insurance Variables, Old Method (Imputation Set 1)

The imputation type used for the three health insurance variables imputed under the old method differs from the three types described in Chapter 2, but it is most similar to single response propensity (RP)/multiple prediction (PRD). Even though there were three variables in the set, only one RP model was fit and only two PRD models were fit, but there was no provisional hot-deck step. Missing data were filled in for all three variables in a single, final hot-deck step. Since the sole purpose of the old method was to maintain consistency with earlier NSDUHs, the method has not been updated or revised to make it more consistent with the approaches used for other variables. For example, SAS procedures were used to fit the PRD models instead of SUDAAN procedures, again to maintain consistency with pre-2002 data processing. These are among the few PRD models in the NSDUH imputation process that continue to use SAS instead of SUDAAN.

[68] SKHLCCOV was coded as 3 if the respondent was covered by a State children's health insurance program but was not covered by Medicaid, Medicare, CHAMPUS, or private health insurance. Respondents with SKHLCCOV = 3 were treated in the same manner as those with SKHLCCOV = 1.

The sequence of imputation is not important for health insurance imputation under the old method, because no provisional hot-deck step is involved. In fact, the two PRD models can be run in parallel, though this is not done in practice. Rather, the PRD model associated with INSUR3 was fit before the PRD model associated with PINSUR was fit, and no PRD model was fit for INSUR. Since the response rates were high for these two variables, the single RP/multiple PRD type of PMN was appropriate for this imputation set.

9.3.1.1 Response Propensity Step

The response propensity models for Imputation Set 1 utilized the final analysis weight, ANALWT, as an input. All respondents were in the domain for the health insurance variables imputed under the old method. No respondents received skip codes for the final imputation-revised variables IRINSUR, IRINSUR3, and IRPINSUR. For the single RP model, a domain member was considered an item respondent if INSUR3 and PINSUR were nonmissing. See Tables C.65 and C.66 in Appendix C for details of the covariates used in the RP models for these variables.

9.3.1.2 First Prediction Step (INSUR3)

INSUR3 was a dichotomous variable; after imputation, its only values are "yes" and "no." Therefore, it was modeled using logistic regression and the adjusted weights that are outputs of the RP step. This PRD model was unusual in that the LOGISTIC procedure in SAS was used instead of the RLOGIST procedure in SAS-callable SUDAAN. The old method was implemented only to maintain consistency with earlier NSDUHs, in which the SAS procedure was used instead of the corresponding SUDAAN procedure.[69] The single predicted mean used in the later hot-deck step was the predicted probability that the respondent received overall health insurance, as defined by INSUR3.

9.3.1.3 Second Prediction Step (PINSUR)

The response indicator used in the single RP model was used in this PRD model, as were the adjusted weights. However, the domain indicator was different. Only respondents with INSUR3 = Yes were in the domain.

Like INSUR3, PINSUR was a yes/no variable, and it was modeled using the LOGISTIC procedure in SAS. However, since the domain was restricted as described in the preceding section, the single predicted mean used in the later hot-deck step was conditional. In particular, it was the predicted probability that the respondent received private health insurance as defined by PINSUR, given that the respondent was known to have received overall health insurance as defined by INSUR3.

9.3.1.4 Final Hot-Deck Step

Details on the missingness patterns, constraints, and predictive mean vectors for the final hot-deck step for old method health insurance are available in Appendix D. This section explains the more general ideas behind the hot-deck step applied to these variables.

[69] See Section 2.3.2 for a brief comparison of SAS and SUDAAN with regard to prediction models.

There are only six possible missingness patterns for old method health insurance. The three base variables are INSUR, INSUR3, and PINSUR, each of which can have one of three values: yes, no, or missing. This leads to $3 \times 3 \times 3 = 27$ possible combinations. However, some of these 27 combinations do not require imputation (because none of the three are missing), and there are logical relationships that render other combinations impossible. A glance at how the variables are created reveals that PINSUR is "nested" within INSUR, and INSUR is nested within INSUR3. Because of this nesting, the following logical relationships hold:

- PINSUR = yes implies that INSUR = yes and INSUR3 = yes.
- INSUR = yes implies that INSUR3 = yes.
- INSUR3 = no implies that INSUR = no and PINSUR = no.
- INSUR = no implies that PINSUR = no.
- INSUR3 = missing implies that INSUR ≠ yes and PINSUR ≠ yes.
- INSUR = missing implies that PINSUR ≠ yes.

The only remaining possibilities are the six missingness patterns.

The logical constraints preserve the logical relationships between the three variables. The likeness constraints are based only on the predictive means and the AGE variable. The predictive mean vectors are, where appropriate, conditioned on what is known; since the predicted mean associated with PINSUR is conditional, the conditional part has to be undone for missingness patterns where INSUR3 is missing. Finally, since (1) the item response rate is high for these variables, (2) the domain includes all unit respondents, and (3) the constraints are not very restrictive, the vast majority of item nonrespondent cases found donors on the first attempt.

9.3.2 Health Insurance Variables, Constituent Variables Method, First Four Constituent Variables (Imputation Set 2)

The first four imputation-revised constituent 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.

9.3.2.1 Response Propensity Step

The response propensity models for Imputation Set 2 utilized the final analysis weight, ANALWT, as an input. All respondents were in the domain for this stage of the constituent variables method. 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 C.60 in Appendix C for details of the covariates used in the RP models for these variables.

9.3.2.2 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.

9.3.2.3 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 PMN imputation evaluation report (Ault et al., in press).

9.3.2.4 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-to-64 age group. This was done because (1) only a small proportion of respondents in these age groups had Medicare, particularly for the 18-to-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 (two missing values in the 2011 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.

9.3.2.5 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 under the constituent variables method are available in Appendix D. This section explains the general ideas behind the hot-deck step applied to these variables.

There are fifteen missingness patterns for the first stage of constituent variables method health insurance imputation. 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. No logical constraints were applied, since 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. As had been the case for variables imputed under the old method, 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.

9.3.3 Health Insurance Variables, Constituent Variables Method, Final Constituent Variable (Imputation Set 3)

The final imputation-revised constituent 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 under the constituent variables method, 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 65-or-older age group.

9.3.3.1 Response Propensity Step

The response propensity models for Imputation Set 3 utilized the final analysis weight, ANALWT, as an input. 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 variable ANYOTHER. See Table C.60 in Appendix C for details of the covariates used in the RP model for this variable.

9.3.3.2 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.

9.3.3.3 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.

9.3.3.4 Recodes for Additional Analyses

The overall health insurance variable associated with the constituent variables method, 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 according to the constituent variables method. 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 constituent health insurance variables or a reported "no" for all five of them; and "3" otherwise.

10. Imputation for the NSDUH Roster Pair Variables

10.1 Introduction

In each household selected for the National Survey on Drug Use and Health (NSDUH), zero, one, or two household members are selected for interviewing. When two members of the same household are selected, the data for the responding pair (i.e., pair data) are used to study outcome variables based on the relationship between those household members (i.e., the pair relationship). For these analyses, 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, where 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 roster pair variables (hereafter, "pair variables") for the 2011 NSDUH. The variables described in this chapter can be divided into three groups:

- pair relationship variables,
- multiplicity count variables, and
- household-level person count variables.

10.1.1 Pair Relationship Variable

The pair relationship variables are derived from the household composition (roster) variables, as described in Chapter 7 of this report. These variables include the edited and imputation-revised pair relationship variables PAIRREL[70] and IRPRREL, respectively, as well as the imputation indicator IIPRREL, which summarizes how the data in IRPRREL were obtained. In addition to these variables, the quality-of-match indicator RELMATCH and the pair indicator PAIRMEM (whether a respondent was part of a respondent pair) were created. Finally, four additional variables were created to aid in pair analyses: PRNTIND, AGEOTHER, SEXOTHER, and PAIRID. The variable PRNTIND identified whether the respondent was a parent in a parent-child relationship; 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.

10.1.2 Multiplicity Count Variables

For analyses 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

---

[70] The levels of PAIRREL are provided in Table 10.3.

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. "Multiplicity" is the term used to describe the complication that arises in the analysis of pair data in which the analysis is at the person level for a given pair domain, and several pairs in the household could be associated with the same person. 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. 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).

Consider the earlier example for person-level outcome variables where the proportion of interest is that of children who use drugs and whose parents report talking to them about drugs. In this case, if the household has two parents, the selected child has *two* inclusion possibilities (one with each parent) in the set of all such parent-child pairs. However, because children form the target population for this example, it is desirable to assign only one observation per child. A reasonable way to achieve this is to take an average of the two proportions of children who use drugs that together correspond to the two pairs associated with the child (i.e., one for each parent in this example). In other words, the response for each parent-child pair from two-parent households is divided by the number of parents. This divisor is known as the multiplicity factor.

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

$$\tau_y(d) = \sum_{h=1}^{H} \sum_{i=1}^{N_h(d)} \sum_{p=1}^{M_{hi}(d)} \frac{y_{hip}(d)}{M_{hi}(d)}$$

(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 persons in the household that are in domain *d*). For the sake of simplicity, the weights are not shown in the above estimator.

Multiplicity count variables were only created for specific relationships of interest. These variables are listed in Table 10.1. With the exception of the spouse-spouse multiplicity variables listed at the bottom of this table, these variables all underwent imputation.

Table 10.1 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.

10.1.3 Household-Level Person Count Variables

Whereas the multiplicity count variable is the number of pairs in the household that can be associated with the focus pair member, the household-level person count variable is the number of persons of focus in the household for a given pair domain, provided such a pair domain existed in the household, 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. Note that household person counts for all domains are 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 10.2, and all underwent imputation.

Table 10.2 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

10.1.4 Staged Variable Processing

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 2 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 creation and imputation were completed for the variables in each stage before moving on to the next stage.

10.2 Stage One Editing: Pair Relationships

10.2.1 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.

10.2.2 Creating the Pair Relationship Variable (PAIRREL)

Because the creation of the multiplicity factors requires complicated programming logic, multiplicities could not be created for all possible pair relationships. The following pair relationships were considered to be "of interest," requiring the creation of multiplicities in each case:

- 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;[71] and
- spouse-spouse (includes partner-partner), with or without children.

Even though these pair relationships were of the most interest, no restrictions were placed on the types of pairs that could be selected for inclusion in the NSDUH sample. However, 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.) The various pair relationships that could be identified are stored in the variable PAIRREL, the levels of which are summarized in Table 10.3. The levels in PAIRREL 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.

The process of identifying the pair relationships was a two-step process: (1) match the household rosters of the pair members, and (2) determine the pair relationships using the

[71] 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.

relationship codes and ages of the matched rosters, if they could be determined. The first step is described in Section 10.2.2.1 and Appendix G, and the second step is described in Section 10.2.2.2 and Appendix G. Any relationships that could not be determined by this process were imputed as described in Section 10.5.2.

Table 10.3 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
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

Table 10.3 Levels of the Variable PAIRREL (continued)

Value of PAIRREL	Interpretation	Domain of Interest
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.

10.2.2.1 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 G. In general, the conditions matched IRSEX and AGE for the one pair member against the age and gender 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, as described by the conditions listed in Table 10.4.

Table 10.4 Measures of the Quality of Definitive Roster Matches

Index Value	Description
0	Age and gender matched exactly, with exactly one MBRSEL correctly identifying the other pair member
1	Age and gender matched exactly, with MBRSEL correctly identifying the other pair member, but there was more than one MBRSEL[1]
2	Age within one, gender matched exactly, with exactly one MBRSEL correctly identifying the other pair member
3	Age within two, gender matched exactly, with exactly one MBRSEL correctly identifying the other pair member
4	Age and gender matched exactly, with MBRSEL missing for all roster members
5	Age matched exactly, gender off, with exactly one MBRSEL correctly identifying the other pair member
6	Age within one, gender matched exactly, with MBRSEL correctly identifying the other pair member, but there was more than one MBRSEL[1]
7	Age within two, gender matched exactly, with MBRSEL correctly identifying the other pair member, but there was more than one MBRSEL[1]
8	Age within one, gender matched exactly, with MBRSEL missing for all roster members
9	Age within two, gender matched exactly, with MBRSEL missing for all roster members[2]
10	Age within 10, gender matched exactly, with exactly one MBRSEL correctly identifying the other pair member[2]

[1] Since the 2001 survey, it was technically impossible to identify more than one roster member as the "other pair member selected," resulting in either zero or one MBRSEL for each respondent pair. As a result, index values #1, #6, and #7 did not occur in 2011.

[2] For pairs where one pair member had a match corresponding to index values #9 or #10, if the other pair member had a match no better than index value #9, an additional requirement was implemented where the reported household sizes for both pair members had to be equal to 2.

Given that at least one side of the pair indicated a match according to one of the index values provided in Table 10.4, the other side could have a match that was weaker (i.e., not definitive), using the index values in Table 10.5. Table 10.5 also shows the weakest match that was allowed (as denoted by the index value) for the other pair member.

In the cases where a single roster member had to be selected among duplicates (index values #14 through #18) 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.

Table 10.5 Measures of the Quality of Roster Matches That Are Not Definitive, Given That One Side of the Pair Had a Definitive Match (as Shown by the Conditions Provided in Table 10.4)

Index Value	Description	Weakest Index Value Allowed for Other Pair Member	
		In Code	Observed
11	Age within 10, gender matched exactly, with MBRSEL missing for all roster members, provided another roster member with a closer age could not have been chosen	8	2
12	Everything missing, but the other pair member had good data	9	05
13	Age missing, gender matched exactly, household sizes equal	9	4
14	Age, gender, and relationship code matched exactly for two roster members, with two MBRSELs identifying the two roster members (one was randomly selected)[1]	8	Not observed
15	Age, gender, and relationship code matched exactly for two roster members, with MBRSEL missing for all roster members (one was randomly selected)	8	4
16	Age and gender matched exactly for two or more roster members, with MBRSEL missing for all roster members (one with matching relationship code was randomly selected)	8	Not observed
17	Age within one and gender matched exactly for two roster members, both with the same relationship code, with two MBRSELs identifying the two roster members (one was randomly selected)[1]	8	Not observed
18	Age within one and gender matched exactly for two roster members, both with the same relationship code, with MBRSEL missing for all roster members (one was randomly selected)	8	Not observed
19	Age within one, gender off, with exactly one MBRSEL correctly identifying the other pair member, only two members in household	10	0
20	No matches possible, but relationship codes indicate the pair is not a part of a domain of interest	As with other pair member	As with other pair member
21	Age matches exactly, gender off, with MBRSEL missing for all roster members	9	0
22	No matches possible	9	4

[1] Since the 2001 survey, it was technically impossible to identify more than one roster member as the "other pair member selected," resulting in either zero or one MBRSEL for each respondent pair. As a result, index values #14 and #17 did not occur in 2011.

10.2.2.2 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 10.3. 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.[72]

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 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 10.3).

[72] The spouse-spouse pair relationship includes partner-partner pair relationships.

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 10.6. 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 2011 survey is provided in Table 10.7. The amount of imputation required was dependent upon the quality of the rosters. The attributes of the roster are described in Chapter 7.

Table 10.6 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.
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.

Table 10.6 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH (continued)

Value of RELMATCH	Values of PAIRREL	Interpretation
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 10.7 Frequencies of the Levels of the Variable RELMATCH: 2011

RELMATCH	Frequency (Percent)	Requires Imputation
0	17 (0.09)	Yes
1 (PAIRREL ≠ 10)	19,539 (97.81)	No
1 (PAIRREL = 10)	40 (0.20)	Yes
1.5	1 (0.01)	No
2 (PAIRREL ≠ 10)	96 (0.48)	No
2 (PAIRREL = 10)	0 (0.00)	Yes
3	99 (0.50)	No
4	184 (0.92)	Yes

10.3 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 10.2, 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 10.3.1, and the second step is described in Section 10.3.2 and Appendix H. Multiplicities that could not be determined through these steps were left to imputation as described in Section 10.5.3.

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.

10.3.1 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 10.8. 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.[73] 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 10.8 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

10.3.2 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 multiplicity count came from the pair member with good data.[74] 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 FIPE3 variable was used to help reconcile differences. Detailed rules for reconciling differences between pair members are provided in Appendix H.

---

[73] 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.

[74] There were some provisions 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.

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 2011 survey is shown in Table 10.9. 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 10.9 Amount of Imputation Required for Multiplicities in Various Pair Domains: 2011

Pair Domain	Multiplicity	Missing Cases
Parent-Child (12–14), Child Focus	Number of parents	76
Parent-Child (12–14), Parent Focus	Number of children	0
Parent-Child (15–17), Child Focus	Number of parents	76
Parent-Child (15–17), Parent Focus	Number of children	0
Parent-Child (12–17), Child Focus	Number of parents	152
Parent-Child (12–17), Parent Focus	Number of children	0
Parent-Child (12–20), Child Focus	Number of parents	184
Parent-Child (12–20), Parent Focus	Number of children	2
Sibling (12–14)-Sibling (15–17), Older Sibling Focus	Number of younger siblings	2
Sibling (12–14)-Sibling (15–17), Younger Sibling Focus	Number of older siblings	2
Sibling (12–17)-Sibling (18–25), Older Sibling Focus	Number of younger siblings	8
Sibling (12–17)-Sibling (18–25), Younger Sibling Focus	Number of older siblings	12

10.4 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 (Chen et al., 2013), it was necessary to identify the household-level person counts for each domain. This 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 10.3. For the parent-child counts where the child was between 15 and 17, calculating the household counts was unnecessary.[75] For the other two remaining sibling-sibling domains, the reason is historical: They were added after the procedures were first developed, and there was

[75] 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.

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 10.4.1, and the second step is described in Section 10.4.2 and Appendix I. Household counts that could not be determined by this process were left up to imputation as described in Section 10.5.4.

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 10.3. Moreover, as stated in that section, to implement such an adjustment would have 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.

10.4.1 Determining the Household-Level Person Count for Each Respondent

10.4.1.1 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.

10.4.1.1.1 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.

10.4.1.1.2 Obtaining Counts for Child-Focus Parent-Child Domains: Other Pairs and Single Respondents

For other pairs[76] and single respondents, the following conditions were required to determine the household count for the number of children of parents in the household:

[76] "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 10.4.1.1.3 and 10.4.1.2.2.

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.

10.4.1.1.3 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).
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.

10.4.1.2 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.

10.4.1.2.1 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 H). Note that a pair that was within one sibling-sibling pair domain had to be outside the other sibling-sibling pair domain.

10.4.1.2.2 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.

10.4.1.3 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[77] but had a spouse (or unmarried partner) and children younger than 18.

[77] 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.

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.

10.4.2 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 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 I. 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 I, 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 I.

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.

10.5 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 10.5.1. The imputation process for stages one, two, and three are described in Sections 10.5.2 through 10.5.4, respectively. The final covariates for the fitted models are listed in Appendix C.

10.5.1 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.

10.5.1.1 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,[78] 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 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 2011

[78] In households where a pair of respondents was selected, the household size had to be greater than or equal to two.

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 J.

10.5.1.2 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- to-25 count and adding 1 to the previously obtained 12-to-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 J.

10.5.1.3 Household Composition Age Counts of Males

For some pair variables, particularly spouse-spouse pairs, knowledge of the gender of the roster member was important in imputing missing values. 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.

10.5.2 Stage One Imputation: Pair Relationships

Missing pair relationships were imputed 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 10.10 presents these 11 age group pairs, as well as the pair relationships prevalent within each age group pair.

Table 10.10 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**
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**

Table 10.10 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
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.

10.5.2.1 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 in Westlake, Chen, and Gordek (2013) for technical details about this procedure.

10.5.2.2 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 10.5.1.2. 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.[79] These groups were combined again before the hot-deck

[79] 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.

imputation step. Further details about the variables used in these models can be found in Appendix C.

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.[80] 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 10.5.2.1.

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 10.11 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.

Table 10.11 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.

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

[80] 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.

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.

10.5.2.3 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.

10.5.3 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 10.12. 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 age group of 12 to 20 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 12-to-14, 12-to-17, 15-to-17, and 12-to-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 10.12 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

Table 10.12 Modeled Multiplicities within Domains (continued)

Imputation Set	Domain	Model Type	Base Variables	Imputation-Revised Variables
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

10.5.3.1 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.

10.5.3.1.1 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.

10.5.3.1.2 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 10.5.2.2, 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.

10.5.3.1.3 Hot-Deck Step

Though modeling was only done for the age group of 12 to 20, the multiplicity counts for the 12-to-14, 12-to-17, 15-to-17, and 12-to-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 D.112 and D.113 in Appendix D.

10.5.3.2 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 age group of 12 to 20, 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.

10.5.4 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 10.4: 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 age group of 12 to 20, and the household-level person counts were assigned for 12-to-14, 12-to-17, and 12-to-20 parent-child pair domains from a single donor. The household-level person counts for the 15-to-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.

10.5.4.1 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 10.5.4.1.1 through 10.5.4.1.3.

10.5.4.1.1 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.

10.5.4.1.2 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 10.5.2.2. 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.

10.5.4.1.3 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 D.121 and D.122.

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.

10.5.4.2 Imputation for Sibling-Sibling (15-17/18-25) 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.

10.5.4.3 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.

10.5.4.3.1 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.

10.5.4.3.2 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.

10.5.4.3.3 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.

10.5.4.3.4 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.

10.5.4.3.5 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 D.127. 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 D.121 and D.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.

References

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.

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.

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.

Ault, K., Barnett-Walker, K., Carpenter, L., Cummiskey, C., Dai, L., Edwards, S., Eicheldinger, C., Frechtel, P., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (in press). *National Survey on Drug Use and Health: Predictive mean neighborhood imputation evaluation* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, RTI/0211838.108.006.018). Research Triangle Park, NC: RTI International.

Center for Behavioral Health Statistics and Quality. (2012). *Results from the 2011 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2013). Person-level sampling weight calibration. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Phase II, Deliverable No. 39, RTI/0211838.207.004). Research Triangle Park, NC: RTI International.

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.

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. (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.

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.

Kroutil, L. A., Handley, W., Bradshaw, M. R., Chien, C., & Felts, B. J. (2013). Procedures for editing supplementary self-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.

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

Morton, K. B., Martin, P. C., Shook-Sa, B. E., Chromy, J. R., & Hirsch, E. L. (2012). Sample design report. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Phase II, Deliverable No. 8, RTI/0211838.203.004). Research Triangle Park, NC: RTI International.

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). Rockville, MD: Substance Abuse and Mental Health Services Administration.

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). Rockville, MD: Substance Abuse and Mental Health Services Administration.

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). Rockville, MD: Substance Abuse and Mental Health Services Administration.

RTI International. (2008). *SUDAAN® language manual, Release 10.0*. 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., 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.

SAS Institute Inc. (2004). *SAS/STAT 9.1 user's guide.* 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.

Scott, V., Archambault, H., Carpenter, L., Edwards, S., Frechtel, P., & Moore, A. (2012). *2011 National Survey on Drug Use and Health: Editing and imputation evaluation report* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Deliverable No. 15, RTI/0211838.208.003). Research Triangle Park, NC: RTI International.

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/]

Westlake, M., Chen, P., & Gordek, H. (2013). Questionnaire dwelling unit-level and person pair-level sampling weight calibration. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 13, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.005). Research Triangle Park, NC: RTI International.

Appendix A: Percentage Imputed or Logically Assigned, 2011 Survey

Table A.1 Percentage of Cases Imputed for General Cigarette Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Cigarette Recency	Lifetime Cigarette Users	35,579	469	3	472	0.44%	0.00%	0.44%
Cigarette Frequency Past Month	Past Month Cigarette Users	15,977	34	139	173	0.10%	0.70%	0.79%
Cigarette Age at First Use	Lifetime Cigarette Users	35,579	461	25	486	0.86%	0.08%	0.94%
Cigarette Day of First Use	Lifetime Cigarette Users	35,579	35,579	0	35,579	100.00%	0.00%	100.00%
Cigarette Month of First Use	Lifetime Cigarette Users	35,579	32,982	31	33,013	97.95%	0.02%	97.97%
Cigarette Year of First Use	Lifetime Cigarette Users	35,579	32,818	8	32,826	97.82%	0.00%	97.83%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.2 Percentage of Cases Imputed for Daily Cigarette Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Cigarette Daily Use	Lifetime Cigarette Users	35,579	21	34	55	0.04%	0.07%	0.11%
Cigarette Age at First Daily Use	Daily Cigarette Users	17,442	121	0	121	0.84%	0.00%	0.84%
Cigarette Day of First Daily Use	Daily Cigarette Users	17,442	17,442	0	17,442	100.00%	0.00%	100.00%
Cigarette Month of First Daily Use	Daily Cigarette Users	17,442	16,454	3	16,457	98.52%	0.00%	98.52%
Cigarette Year of First Daily Use	Daily Cigarette Users	17,442	16,406	0	16,406	98.43%	0.00%	98.43%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.3 Percentage of Cases Imputed for Cigarette Dependency 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	15,977	1	2	3	0.01%	0.02%	0.03%
Smoke Same Number of Cigarettes from Day to Day	Past Month Cigarette Users	15,977	3	2	5	0.01%	0.02%	0.03%
Number Cigarettes Smoked per Day Often Changes	Past Month Cigarette Users	15,977	11	2	13	0.02%	0.02%	0.04%
Number Cigarettes Smoked per Day Influenced by Other Things	Past Month Cigarette Users	15,977	6	2	8	0.01%	0.02%	0.03%
Smoke Same Number of Cigarettes on Weekends As on Weekdays	Past Month Cigarette Users	15,977	5	2	7	0.02%	0.02%	0.04%
Smoke Lots of Cigarettes in an Hour, Then No Cigarettes Until Hours Later	Past Month Cigarette Users	15,977	21	2	23	0.05%	0.02%	0.07%
Amount of Smoking Has Increased Since Started Smoking	Past Month Cigarette Users	15,977	6	2	8	0.02%	0.02%	0.04%
Must Smoke Much More Now Before Start To Feel Anything	Past Month Cigarette Users	15,977	33	2	35	0.23%	0.02%	0.25%
Need To Smoke a Lot More To Be Satisfied	Past Month Cigarette Users	15,977	5	2	7	0.05%	0.02%	0.07%
Crave Cigarettes when Haven't Smoked for a Few Hours	Past Month Cigarette Users	15,977	1	2	3	0.00%	0.02%	0.03%
Need To Smoke To Feel Less Irritable	Past Month Cigarette Users	15,977	14	2	16	0.12%	0.02%	0.14%
Feel in Control of Smoking	Past Month Cigarette Users	15,977	10	2	12	0.03%	0.02%	0.05%
Cravings for Cigarettes like Force Can't Control	Past Month Cigarette Users	15,977	4	2	6	0.03%	0.02%	0.05%
Worry about Running Out of Cigarettes	Past Month Cigarette Users	15,977	1	2	3	0.00%	0.02%	0.02%

Table A.3 Percentage of Cases Imputed for Cigarette Dependency 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
Smoking Not Affected by Other Things	Past Month Cigarette Users	15,977	18	2	20	0.10%	0.02%	0.12%
Tend To Avoid Places That Don't Allow Smoking	Past Month Cigarette Users	15,977	20	2	22	0.07%	0.02%	0.09%
No Travel by Airplane Because No Smoking Allowed	Past Month Cigarette Users	15,977	31	2	33	0.20%	0.02%	0.22%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.4 Percentage of Cases Imputed for Cigar 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
Cigar Lifetime Use	All Respondents	70,109	11	0	11	0.01%	0.00%	0.01%
Cigar Recency	Lifetime Cigar Users	20,826	344	2	346	0.70%	0.00%	0.70%
Cigar Frequency Past Month	Past Month Cigar Users	4,477	20	41	61	0.44%	0.94%	1.39%
Cigar Age at First Use	Lifetime Cigar Users	20,826	472	0	472	2.45%	0.00%	2.45%
Cigar Day of First Use	Lifetime Cigar Users	20,826	20,826	0	20,826	100.00%	0.00%	100.00%
Cigar Month of First Use	Lifetime Cigar Users	20,826	18,235	10	18,245	95.46%	0.01%	95.47%
Cigar Year of First Use	Lifetime Cigar Users	20,826	18,049	0	18,049	95.17%	0.00%	95.17%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	7	0	7	0.01%	0.00%	0.01%
Chewing Tobacco Recency	Lifetime Chewing Tobacco Users	7,727	163	2	165	0.74%	0.00%	0.74%
Chewing Tobacco Frequency Past Month	Past Month Chewing Tobacco Users	1,131	5	9	14	0.14%	0.66%	0.81%
Chewing Tobacco Age at First Use	Lifetime Chewing Tobacco Users	7,727	167	0	167	1.42%	0.00%	1.42%
Chewing Tobacco Day of First Use	Lifetime Chewing Tobacco Users	7,727	7,727	0	7,727	100.00%	0.00%	100.00%
Chewing Tobacco Month of First Use	Lifetime Chewing Tobacco Users	7,727	7,014	4	7,018	97.18%	0.01%	97.19%
Chewing Tobacco Year of First Use	Lifetime Chewing Tobacco Users	7,727	6,961	0	6,961	96.98%	0.00%	96.98%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	39	0	39	0.05%	0.00%	0.05%
Snuff Recency	Lifetime Snuff Users	9,819	272	0	272	1.24%	0.00%	1.24%
Snuff Frequency Past Month	Past Month Snuff Users	2,459	7	19	26	0.09%	0.32%	0.41%
Snuff Age at First Use	Lifetime Snuff Users	9,819	250	0	250	1.69%	0.00%	1.69%
Snuff Day of First Use	Lifetime Snuff Users	9,819	9,819	0	9,819	100.00%	0.00%	100.00%
Snuff Month of First Use	Lifetime Snuff Users	9,819	8,638	5	8,643	95.41%	0.02%	95.42%
Snuff Year of First Use	Lifetime Snuff Users	9,819	8,533	0	8,533	95.05%	0.00%	95.05%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.7 Percentage of Cases Imputed for Smokeless 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
Smokeless Tobacco Lifetime Use	All Respondents	70,109	39	0	39	0.05%	0.00%	0.05%
Smokeless Tobacco Recency	Lifetime Smokeless Tobacco Users	11,933	331	2	333	1.18%	0.00%	1.18%
Smokeless Tobacco Age at First Use	Lifetime Smokeless Tobacco Users	11,933	368	0	368	2.00%	0.00%	2.00%
Smokeless Tobacco Day of First Use	Lifetime Smokeless Tobacco Users	11,933	11,933	0	11,933	100.00%	0.00%	100.00%
Smokeless Tobacco Month of First Use	Lifetime Smokeless Tobacco Users	11,933	10,622	5	10,627	96.11%	0.01%	96.11%
Smokeless Tobacco Year of First Use	Lifetime Smokeless Tobacco Users	11,933	10,522	0	10,522	95.83%	0.00%	95.83%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.8 Percentage of Cases Imputed for Pipe Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Pipe Lifetime Use	All Respondents	70,109	14	0	14	0.03%	0.00%	0.03%
Pipe Past Month Use	Lifetime Pipe Users	6,034	4	0	4	0.08%	0.00%	0.08%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	18	0	18	0.01%	0.00%	0.01%
Alcohol Recency	Lifetime Alcohol Users	48,845	864	11	875	1.13%	0.01%	1.14%
Alcohol Frequency Past Year	Past Year Alcohol Users	41,472	796	1,429	2,225	1.38%	1.60%	2.98%
Alcohol Frequency Past Month	Past Month Alcohol Users	30,691	630	184	814	1.73%	0.52%	2.25%
Alcohol 5+ Drinks Past Month	Past Month Alcohol Users	30,691	1,047	79	1,126	2.88%	0.22%	3.10%
Alcohol Age at First Use	Lifetime Alcohol Users	48,845	652	0	652	1.24%	0.00%	1.24%
Alcohol Day of First Use	Lifetime Alcohol Users	48,845	48,845	0	48,845	100.00%	0.00%	100.00%
Alcohol Month of First Use	Lifetime Alcohol Users	48,845	43,699	11	43,710	96.86%	0.01%	96.86%
Alcohol Year of First Use	Lifetime Alcohol Users	48,845	43,371	0	43,371	96.68%	0.00%	96.68%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	55	0	55	0.08%	0.00%	0.08%
Marijuana Recency	Lifetime Marijuana Users	27,389	496	9	505	0.99%	0.01%	1.01%
Marijuana Frequency Past Year	Past Year Marijuana Users	12,679	449	947	1,396	2.72%	5.06%	7.79%
Marijuana Frequency Past Month	Past Month Marijuana Users	7,515	214	63	277	2.30%	0.54%	2.84%
Marijuana Age at First Use	Lifetime Marijuana Users	27,389	245	0	245	0.74%	0.00%	0.74%
Marijuana Day of First Use	Lifetime Marijuana Users	27,389	27,389	0	27,389	100.00%	0.00%	100.00%
Marijuana Month of First Use	Lifetime Marijuana Users	27,389	24,621	2	24,623	96.65%	0.00%	96.65%
Marijuana Year of First Use	Lifetime Marijuana Users	27,389	24,456	0	24,456	96.45%	0.00%	96.45%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	122	0	122	0.08%	0.00%	0.08%
Inhalant Recency	Lifetime Inhalant Users	6,010	257	1	258	1.79%	0.01%	1.80%
Inhalant Frequency Past Year	Past Year Inhalant Users	1,125	142	135	277	9.73%	7.84%	17.57%
Inhalant Frequency Past Month	Past Month Inhalant Users	317	39	24	63	8.12%	8.65%	16.77%
Inhalant Age at First Use	Lifetime Inhalant Users	6,010	343	0	343	3.22%	0.00%	3.22%
Inhalant Day of First Use	Lifetime Inhalant Users	6,010	6,010	0	6,010	100.00%	0.00%	100.00%
Inhalant Month of First Use	Lifetime Inhalant Users	6,010	5,321	6	5,327	95.51%	0.05%	95.56%
Inhalant Year of First Use	Lifetime Inhalant Users	6,010	5,241	0	5,241	95.13%	0.00%	95.13%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	39	0	39	0.04%	0.00%	0.04%
Heroin Recency	Lifetime Heroin Users	934	28	1	29	1.42%	0.07%	1.49%
Heroin Frequency Past Year	Past Year Heroin Users	241	20	26	46	8.28%	10.43%	18.71%
Heroin Frequency Past Month	Past Month Heroin Users	89	10	2	12	9.04%	3.95%	12.99%
Heroin Age at First Use	Lifetime Heroin Users	934	10	0	10	1.49%	0.00%	1.49%
Heroin Day of First Use	Lifetime Heroin Users	934	934	0	934	100.00%	0.00%	100.00%
Heroin Month of First Use	Lifetime Heroin Users	934	801	0	801	93.94%	0.00%	93.94%
Heroin Year of First Use	Lifetime Heroin Users	934	793	0	793	93.78%	0.00%	93.78%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	257	0	257	0.24%	0.00%	0.24%
Hallucinogen Recency	Lifetime Hallucinogen Users	8,967	232	61	293	1.67%	0.37%	2.04%
Hallucinogen Frequency Past Year	Past Year Hallucinogen Users	2,303	109	158	267	4.59%	5.23%	9.83%
Hallucinogen Frequency Past Month	Past Month Hallucinogen Users	591	46	8	54	10.72%	0.43%	11.14%
Hallucinogen Age at First Use	Lifetime Hallucinogen Users	8,967	124	88	212	1.24%	0.85%	2.09%
Hallucinogen Day of First Use	Lifetime Hallucinogen Users	8,967	8,967	0	8,967	100.00%	0.00%	100.00%
Hallucinogen Month of First Use	Lifetime Hallucinogen Users	8,967	7,838	60	7,898	95.61%	0.20%	95.81%
Hallucinogen Year of First Use	Lifetime Hallucinogen Users	8,967	7,777	11	7,788	95.37%	0.04%	95.41%
Hallucinogen Use Other than LSD, PCP, or Ecstasy	All Respondents	70,109	232	0	232	0.22%	0.00%	0.22%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	72	0	72	0.10%	0.00%	0.10%
LSD Recency	Lifetime LSD Users	4,439	74	0	74	1.10%	0.00%	1.10%
LSD Age at First Use	Lifetime LSD Users	4,439	74	5	79	1.61%	0.05%	1.67%
LSD Day of First Use	Lifetime LSD Users	4,439	4,439	0	4,439	100.00%	0.00%	100.00%
LSD Month of First Use	Lifetime LSD Users	4,439	4,044	0	4,044	97.87%	0.00%	97.87%
LSD Year of First Use	Lifetime LSD Users	4,439	4,017	1	4,018	97.72%	0.00%	97.73%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.15 Percentage of Cases Imputed for PCP 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
PCP Lifetime Use	All Respondents	70,109	65	0	65	0.08%	0.00%	0.08%
PCP Recency	Lifetime PCP Users	990	24	0	24	1.35%	0.00%	1.35%
PCP Age at First Use	Lifetime PCP Users	990	38	4	42	3.82%	0.33%	4.15%
PCP Day of First Use	Lifetime PCP Users	990	990	0	990	100.00%	0.00%	100.00%
PCP Month of First Use	Lifetime PCP Users	990	933	1	934	98.81%	0.01%	98.82%
PCP Year of First Use	Lifetime PCP Users	990	928	1	929	98.78%	0.01%	98.79%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.16 Percentage of Cases Imputed for Ecstasy 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
Ecstasy Lifetime Use	All Respondents	70,109	86	0	86	0.10%	0.00%	0.10%
Ecstasy Recency	Lifetime Ecstasy Users	4,778	102	1	103	1.55%	0.00%	1.55%
Ecstasy Age at First Use	Lifetime Ecstasy Users	4,778	54	3	57	1.19%	0.02%	1.20%
Ecstasy Day of First Use	Lifetime Ecstasy Users	4,778	4,778	0	4,778	100.00%	0.00%	100.00%
Ecstasy Month of First Use	Lifetime Ecstasy Users	4,778	3,956	0	3,956	91.34%	0.00%	91.34%
Ecstasy Year of First Use	Lifetime Ecstasy Users	4,778	3,910	1	3,911	90.96%	0.00%	90.96%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	33	0	33	0.05%	0.00%	0.05%
Cocaine Recency	Lifetime Cocaine Users	7,330	93	24	117	0.73%	0.30%	1.03%
Cocaine Frequency Past Year	Past Year Cocaine Users	1,570	72	136	208	5.61%	7.79%	13.40%
Cocaine Frequency Past Month	Past Month Cocaine Users	482	30	13	43	7.93%	3.75%	11.68%
Cocaine Age at First Use	Lifetime Cocaine Users	7,330	64	100	164	0.87%	1.09%	1.96%
Cocaine Day of First Use	Lifetime Cocaine Users	7,330	7,330	0	7,330	100.00%	0.00%	100.00%
Cocaine Month of First Use	Lifetime Cocaine Users	7,330	6,704	3	6,707	97.37%	0.02%	97.39%
Cocaine Year of First Use	Lifetime Cocaine Users	7,330	6,663	1	6,664	97.18%	0.00%	97.19%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	35	0	35	0.06%	0.00%	0.06%
Crack Recency	Lifetime Crack Users	1,624	18	0	18	0.44%	0.00%	0.44%
Crack Frequency Past Year	Past Year Crack Users	199	8	12	20	1.82%	3.23%	5.05%
Crack Frequency Past Month	Past Month Crack Users	61	1	1	2	0.93%	2.40%	3.33%
Crack Age at First Use	Lifetime Crack Users	1,624	23	0	23	1.09%	0.00%	1.09%
Crack Day of First Use	Lifetime Crack Users	1,624	1,624	0	1,624	100.00%	0.00%	100.00%
Crack Month of First Use	Lifetime Crack Users	1,624	1,542	0	1,542	98.36%	0.00%	98.36%
Crack Year of First Use	Lifetime Crack Users	1,624	1,534	0	1,534	98.23%	0.00%	98.23%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	143	0	143	0.09%	0.00%	0.09%
Tranquilizer Recency	Lifetime Tranquilizer Users	5,706	113	0	113	1.43%	0.00%	1.43%
Tranquilizer Frequency Past Year	Past Year Tranquilizer Users	1,915	86	139	225	4.04%	5.36%	9.40%
Tranquilizer Age at First Use	Lifetime Tranquilizer Users	5,706	163	0	163	3.48%	0.00%	3.48%
Tranquilizer Day of First Use	Lifetime Tranquilizer Users	5,706	5,706	0	5,706	100.00%	0.00%	100.00%
Tranquilizer Month of First Use	Lifetime Tranquilizer Users	5,706	4,893	1	4,894	92.98%	0.00%	92.99%
Tranquilizer Year of First Use	Lifetime Tranquilizer Users	5,706	4,832	0	4,832	92.46%	0.00%	92.46%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	162	0	162	0.15%	0.00%	0.15%
Sedative Recency	Lifetime Sedative Users	1,188	45	0	45	1.86%	0.00%	1.86%
Sedative Frequency Past Year	Past Year Sedative Users	223	23	25	48	12.84%	5.48%	18.32%
Sedative Age at First Use	Lifetime Sedative Users	1,188	67	0	67	4.38%	0.00%	4.38%
Sedative Day of First Use	Lifetime Sedative Users	1,188	1,188	0	1,188	100.00%	0.00%	100.00%
Sedative Month of First Use	Lifetime Sedative Users	1,188	1,060	0	1,060	97.18%	0.00%	97.18%
Sedative Year of First Use	Lifetime Sedative Users	1,188	1,052	0	1,052	97.10%	0.00%	97.10%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.21 Percentage of Cases Imputed for Pain Reliever 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
Pain Reliever Lifetime Use	All Respondents	70,109	241	0	241	0.16%	0.00%	0.16%
Pain Reliever Recency	Lifetime Pain Reliever Users	10,714	413	10	423	3.12%	0.03%	3.15%
Pain Reliever Frequency Past Year	Past Year Pain Reliever Users	4,684	330	388	718	6.57%	6.05%	12.62%
Pain Reliever Age at First Use	Lifetime Pain Reliever Users	10,714	628	13	641	6.47%	0.07%	6.54%
Pain Reliever Day of First Use	Lifetime Pain Reliever Users	10,714	10,714	0	10,714	100.00%	0.00%	100.00%
Pain Reliever Month of First Use	Lifetime Pain Reliever Users	10,714	9,184	26	9,210	93.14%	0.12%	93.26%
Pain Reliever Year of First Use	Lifetime Pain Reliever Users	10,714	9,045	3	9,048	92.51%	0.01%	92.53%
Pain Reliever Use Other than OxyContin	All Respondents	70,109	267	0	267	0.17%	0.00%	0.17%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.22 Percentage of Cases Imputed for OxyContin 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
OxyContin Lifetime Use	All Respondents	70,109	188	0	188	0.17%	0.00%	0.17%
OxyContin Recency	Lifetime OxyContin Users	2,423	88	0	88	3.30%	0.00%	3.30%
OxyContin Frequency Past Year	Past Year OxyContin Users	773	50	69	119	5.94%	6.02%	11.97%
OxyContin Age at First Use	Lifetime OxyContin Users	2,423	78	1	79	3.91%	0.00%	3.91%
OxyContin Day of First Use	Lifetime OxyContin Users	2,423	2,423	0	2,423	100.00%	0.00%	100.00%
OxyContin Month of First Use	Lifetime OxyContin Users	2,423	2,024	1	2,025	89.33%	0.05%	89.37%
OxyContin Year of First Use	Lifetime OxyContin Users	2,423	1,979	1	1,980	88.05%	0.02%	88.07%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	70,109	155	0	155	0.12%	0.00%	0.12%
Core Plus Noncore Stimulant Recency	Lifetime Core Plus Noncore Stimulant Users	4,924	144	11	155	2.25%	0.07%	2.32%
Stimulant Lifetime Use	All Respondents	70,109	155	0	155	0.12%	0.00%	0.12%
Stimulant Recency	Lifetime Stimulant Users	4,489	131	10	141	2.07%	0.06%	2.13%
Stimulant Frequency Past Year	Past Year Stimulant Users	1,198	66	142	208	4.70%	12.18%	16.88%
Stimulant Age at First Use	Lifetime Stimulant Users	4,489	126	23	149	2.30%	0.54%	2.84%
Stimulant Day of First Use	Lifetime Stimulant Users	4,489	4,489	0	4,489	100.00%	0.00%	100.00%
Stimulant Month of First Use	Lifetime Stimulant Users	4,489	3,959	9	3,968	95.27%	0.03%	95.30%
Stimulant Year of First Use	Lifetime Stimulant Users	4,489	3,935	3	3,938	95.13%	0.01%	95.14%
Stimulant Use Other than Methamphetamine	All Respondents	70,109	146	0	146	0.12%	0.00%	0.12%

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, 2011.

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	70,109	55	0	55	0.03%	0.00%	0.03%
Core Plus Noncore Methamphetamine Recency	Lifetime Core Plus Noncore Methamphetamine Users	2,461	67	0	67	2.15%	0.00%	2.15%
Methamphetamine Lifetime Use	All Respondents	70,109	55	0	55	0.03%	0.00%	0.03%
Methamphetamine Recency	Lifetime Methamphetamine Users	1,852	53	0	53	2.32%	0.00%	2.32%
Methamphetamine Frequency Past Year	Past Year Methamphetamine Users	263	24	34	58	4.96%	6.89%	11.85%
Methamphetamine Age at First Use	Lifetime Methamphetamine Users	1,852	60	1	61	3.30%	0.00%	3.30%
Methamphetamine Day of First Use	Lifetime Methamphetamine Users	1,852	1,852	0	1,852	100.00%	0.00%	100.00%
Methamphetamine Month of First Use	Lifetime Methamphetamine Users	1,852	1,737	0	1,737	98.11%	0.00%	98.11%
Methamphetamine Year of First Use	Lifetime Methamphetamine Users	1,852	1,731	0	1,731	98.02%	0.00%	98.02%

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, 2011.

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	70,109	5	0	5	0.01%	0.00%	0.01%
Marital Status	Respondents at Least 15 Years of Age	58,584	14	0	14	0.03%	0.00%	0.03%
Employment Status	Respondents at Least 15 Years of Age	58,584	48	0	48	0.11%	0.00%	0.11%
Employment Status: 18+	Respondents at Least 18 Years of Age	46,599	42	0	42	0.11%	0.00%	0.11%
Immigrant Status								
Born in the United States	All Respondents	70,109	39	2	41	0.07%	0.00%	0.08%
Immigrant Age of Entry in the United States	Respondents Not Born in the United States	7,277	36	0	36	0.52%	0.00%	0.52%
Hispanic or Latino Origin								
Hispanic or Latino Origin	All Respondents	70,109	93	6	99	0.06%	0.00%	0.06%
Hispanic or Latino Origin Group	Hispanic or Latino Respondents	11,303	54	84	138	0.48%	0.71%	1.19%
Single/Multiple Hispanic or Latino Origin Group	Hispanic or Latino Respondents	11,303	56	88	144	0.49%	0.73%	1.22%
Race								
Native Hawaiian	All Respondents	70,109	2,275	0	2,275	3.32%	0.00%	3.32%
Other Pacific Islander	All Respondents	70,109	2,276	2	2,278	3.32%	0.00%	3.33%
Asian	All Respondents	70,109	2,280	307	2,587	3.33%	0.52%	3.85%
Black or African American	All Respondents	70,109	2,310	50	2,360	3.37%	0.04%	3.40%
American Indian/Alaska Native	All Respondents	70,109	2,358	29	2,387	3.37%	0.02%	3.38%
White	All Respondents	70,109	2,378	166	2,544	3.41%	0.19%	3.60%
Detailed Race: 15 Levels	All Respondents	70,109	2,420	509	2,929	3.46%	0.75%	4.20%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.26 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	70,109	855	34	889	1.94%	0.03%	1.97%
Respondent's Total Income (Finer Categories)	All Respondents	70,109	1,882	0	1,882	4.32%	0.00%	4.32%
Total Family Income > or < $20,000	All Respondents	70,109	3,031	0	3,031	3.95%	0.00%	3.95%
Total Family Income (Finer Categories)	All Respondents	70,109	6,994	1,238	8,232	9.47%	4.89%	14.36%
Source of Income								
Family Received Income from Job	All Respondents	70,109	218	0	218	0.24%	0.00%	0.24%
Family Received Social Security or Railroad Retirement Payments	All Respondents	70,109	715	0	715	0.70%	0.00%	0.70%
Family Received Public Assistance	All Respondents	70,109	566	0	566	0.51%	0.00%	0.51%
Family Received Supplemental Security Income	All Respondents	70,109	1,019	0	1,019	0.89%	0.00%	0.89%
Respondent/Other Family Member Received Food Stamps	All Respondents	70,109	299	0	299	0.34%	0.00%	0.34%
Family Received Welfare/Job Placement/Child Care	All Respondents	70,109	425	0	425	0.40%	0.00%	0.40%
Number of Months on Welfare	Family Receives Public Assistance or Welfare/Job Placement/Child Care	5,146	232	0	232	3.60%	0.00%	3.60%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Table A.27 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 1999 Survey Method	All Respondents	70,109	516	4	520	0.37%	0.00%	0.37%
As Defined by the 2001 Survey Method	All Respondents	70,109	529	4	533	0.38%	0.00%	0.38%
As Defined by the Constituent Variables Method	All Respondents	70,109	530	0	530	0.38%	0.00%	0.38%
Type of Insurance								
Private – Consistent with Pre-1999 Surveys	All Respondents	70,109	437	0	437	0.31%	0.00%	0.31%
Private – As Defined by the Constituent Variables Method	All Respondents	70,109	437	0	437	0.31%	0.00%	0.31%
Medicare	All Respondents	70,109	243	0	243	0.23%	0.00%	0.23%
Military Health Care (CHAMPUS, TRICARE, CHAMPVA, VA)	All Respondents	70,109	242	0	242	0.23%	0.00%	0.23%
Medicaid/CHIP	All Respondents	70,109	545	0	545	0.44%	0.00%	0.44%
Other	Respondent Does Not Have Private Health Insurance, Medicare, Medicaid/CHIP, or CHAMPUS	12,668	183	0	183	0.83%	0.00%	0.83%

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, 2011.

Table A.28 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 Persons in Household	All Respondents	70,109	37	0	37	0.09%	0.00%	0.09%
Number of Children in Household Aged < 18	All Respondents	70,109	212	0	212	0.25%	0.00%	0.25%
Number of Persons in Household Aged ≥ 65	All Respondents	70,109	382	37	419	0.32%	0.03%	0.35%
Family Roster								
Presence of Family Members in Household	All Respondents	70,109	45	0	45	0.10%	0.00%	0.10%
Number of Respondent's Family Members in Household excluding Foster Relationships	All Respondents	70,109	60	4	64	0.12%	0.01%	0.12%
Number of Respondent's Family Members in Household including Foster Relationships	All Respondents	70,109	64	0	64	0.12%	0.00%	0.12%
Number of Respondent's Family Members in Household Aged < 18 excluding Foster Relationships	All Respondents	70,109	141	0	141	0.19%	0.00%	0.19%
Number of Respondent's Family Members in Household Aged < 18 including Foster Relationships	All Respondents	70,109	151	0	151	0.20%	0.00%	0.20%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

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	39,952	482	0	482	1.03%	0.00%	1.03%
Multiplicities								
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-14	Parent-Child Pair, Child is 12-14	4,098	0	0	0	0.00%	0.00%	0.00%
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-17	Parent-Child Pair, Child is 12-17	7,684	0	0	0	0.00%	0.00%	0.00%
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-20	Parent-Child Pair, Child is 12-20	9,164	2	0	2	0.02%	0.00%	0.02%
Multiplicity Count: Child-Parent, Parent Focus, Child is 15-17	Parent-Child Pair, Child is 15-17	3,586	0	0	0	0.00%	0.00%	0.00%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-14	Parent-Child Pair, Child is 12-14	4,098	76	0	76	1.41%	0.00%	1.41%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-17	Parent-Child Pair, Child is 12-17	7,684	152	0	152	1.75%	0.00%	1.75%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-20	Parent-Child Pair, Child is 12-20	9,164	184	0	184	1.63%	0.00%	1.63%
Multiplicity Count: Child-Parent, Child Focus, Child is 15-17	Parent-Child Pair, Child is 15-17	3,586	76	0	76	2.09%	0.00%	2.09%
Multiplicity Count: Sibling-Sibling (12-14/15-17), 12-14 Focus	Sibling-Sibling Pair: Younger is 12-14, Older is 15-17	4,700	2	0	2	0.09%	0.00%	0.09%
Multiplicity Count: Sibling-Sibling (12-14/15-17), 15-17 Focus	Sibling-Sibling Pair: Younger is 12-14, Older is 15-17	4,700	2	0	2	0.00%	0.00%	0.00%

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	5,236	12	0	12	0.23%	0.00%	0.23%
Multiplicity Count: Sibling-Sibling (12-17/18-25), 18-25 Focus	Sibling-Sibling Pair: Younger is 12-17, Older is 18-25	5,236	8	0	8	0.08%	0.00%	0.08%
Household Person-Level Count								
Household Count: Number of Spouse-Spouse Pairs in Household	All Respondents	70,109	100	0	100	0.15%	0.00%	0.15%
Household Count: Number of Spouse-Spouse Pairs with Children	All Respondents	70,109	49	0	49	0.05%	0.00%	0.05%
Household Count: Child-Parent, Parent Focus, Child is 12-14	All Respondents	70,109	233	0	233	0.25%	0.00%	0.25%
Household Count: Child-Parent, Parent Focus, Child is 12-17	All Respondents	70,109	376	0	376	0.40%	0.00%	0.40%
Household Count: Child-Parent, Parent Focus, Child is 12-20	All Respondents	70,109	450	0	450	0.48%	0.00%	0.48%
Household Count: Child-Parent, Child Focus, Child is 12-14	All Respondents	70,109	67	1	68	0.07%	0.00%	0.07%
Household Count: Child-Parent, Child Focus, Child is 12-17	All Respondents	70,109	107	0	107	0.08%	0.00%	0.08%
Household Count: Child-Parent, Child Focus, Child is 12-20	All Respondents	70,109	170	0	170	0.12%	0.00%	0.12%
Household Count: Sibling-Sibling (12-14/15-17), 15-17 Focus	All Respondents	70,109	45	0	45	0.08%	0.00%	0.08%
Household Count: Sibling-Sibling (12-17/18-25), 18-25 Focus	All Respondents	70,109	91	0	91	0.09%	0.00%	0.09%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2011.

Appendix B: Race and Hispanic/Latino Group Alpha Codes

Appendix B: Race and Hispanic/Latino Group Alpha Codes

B.1 Introduction

For the 2011 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 3, 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.

B.2 Race

In the 2011 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. The race questions used in the 2011 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	Other Pacific Islander
6	Asian (for example: Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese)
7	Other (Specify)

QD05ASIA: (Asked only if level 6 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 B.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.

B.2.1 Race Alpha Responses

The four types of race other-specify responses are described in Chapter 3. 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 "Jamaican" would have a 76.3 percent chance of being allocated to the black/African-American category, because the latest Jamaican census indicated that 76.3 percent of Jamaicans were black. Thus, even though black Jamaicans would not consider themselves African Americans, they were allocated to the black/African-American category specified in the questionnaire. 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 B.1 lists all the race codes used in the 2011 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 3 for more details on the edited Hispanic/Latino indicator. Note that EDHOIND also could be edited to a "no." This is discussed in Section B.3.1.

B.2.1.1 Handling of Directly Mapped Codes

For codes that were directly mapped, the final column of Table B.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 B.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 persons in the Polish population.

B.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 B.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 B.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 B.2. Note that the racial categories for each country listed in Table B.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 B.2. Instead, this information must be obtained from Table B.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 3 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 Indian/Alaska Native was not one of their component races. See Chapter 3 for the explanation of level 19 of EDRACE.

[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 B.2 for an example.

- Multiple: Imputation was restricted to respondents who were multiple race. See Chapter 3 for the explanation of level 16 of EDRACE.

B.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 B.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.

B.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.

B.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
4 Cuban/Cuban American
5 Dominican (from Dominican Republic)

[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?"

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 B.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.

B.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 B.4 lists all the Hispanic/Latino group codes used in the 2011 survey and the Hispanic/Latino groups to which they mapped. Note that these mappings utilized the arbitrary priority rule provided in Chapter 3. 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 3, 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 3 for a discussion of EDQD04*xx*.

Table B.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 B.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 B.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 B.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: Other Asian
132	Iraq	Directly mapped (geographic category)	QD05: White QD05ASIA: Other Asian
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: Other Asian
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 B.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)	Asian Indian
151	India	Directly mapped (geographic category)	Asian Indian
152	Other South Asia (includes Pakistan, Bangladesh, Himalayan countries)	Directly mapped (geographic category)	Asian Indian
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 B.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.
201	Biracial (nonspecific)	Codes informative for formal imputation procedures	Not a Direct Map
202	White & Black/African American	Directly mapped (racial category)	White & Black/African American

Table B.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
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
207	White & Chinese	Directly mapped (racial category)	White & Asian
208	White & Filipino	Directly mapped (racial category)	White & Asian
209	White & Japanese	Directly mapped (racial category)	White & Asian
210	White & Korean	Directly mapped (racial category)	White & Asian
211	White & Vietnamese	Directly mapped (racial category)	White & Asian
212	White & Other Asian	Directly mapped (racial category)	White & Asian
213	White & Asian (nonspecific)	Directly mapped (racial category)	White & Asian
214	White & Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White & American Indian/Alaska Native QD05ASIA: White & Asian
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
227	Black/African American & Chinese	Directly mapped (racial category)	Black/African American & Asian
228	Black/African American & Filipino	Directly mapped (racial category)	Black/African American & Asian
229	Black/African American & Japanese	Directly mapped (racial category)	Black/African American & Asian

Table B.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
230	Black/African American & Korean	Directly mapped (racial category)	Black/African American & Asian
231	Black/African American & Vietnamese	Directly mapped (racial category)	Black/African American & Asian
232	Black/African American & Other Asian	Directly mapped (racial category)	Black/African American & Asian
233	Black/African American & Asian (nonspecific)	Directly mapped (racial category)	Black/African American & Asian
234	Black/African American & Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: Black/African American & American Indian QD05ASIA: Black/African American & Asian
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
247	American Indian/Alaska Native & Chinese	Directly mapped (racial category)	American Indian/Alaska Native & Asian
248	American Indian/Alaska Native & Filipino	Directly mapped (racial category)	American Indian/Alaska Native & Asian
249	American Indian/Alaska Native & Japanese	Directly mapped (racial category)	American Indian/Alaska Native & Asian
250	American Indian/Alaska Native & Korean	Directly mapped (racial category)	American Indian/Alaska Native & Asian
251	American Indian/Alaska Native & Vietnamese	Directly mapped (racial category)	American Indian/Alaska Native & Asian

Table B.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
252	American Indian/Alaska Native & Other Asian	Directly mapped (racial category)	American Indian/Alaska Native & Asian
253	American Indian/Alaska Native & Asian (nonspecific)	Directly mapped (racial category)	American Indian/Alaska Native & Asian
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
267	Native Hawaiian & Chinese	Directly mapped (racial category)	Native Hawaiian & Asian
268	Native Hawaiian & Filipino	Directly mapped (racial category)	Native Hawaiian & Asian
269	Native Hawaiian & Japanese	Directly mapped (racial category)	Native Hawaiian & Asian
270	Native Hawaiian & Korean	Directly mapped (racial category)	Native Hawaiian & Asian
271	Native Hawaiian & Vietnamese	Directly mapped (racial category)	Native Hawaiian & Asian
272	Native Hawaiian & Other Asian	Directly mapped (racial category)	Native Hawaiian & Asian
273	Native Hawaiian & Asian (nonspecific)	Directly mapped (racial category)	Native Hawaiian & Asian
286	Other Pacific Islander & Asian Indian	Directly mapped (racial category)	Other Pacific Islander & Asian
287	Other Pacific Islander & Chinese	Directly mapped (racial category)	Other Pacific Islander & Asian
288	Other Pacific Islander & Filipino	Directly mapped (racial category)	Other Pacific Islander & Asian
289	Other Pacific Islander & Japanese	Directly mapped (racial category)	Other Pacific Islander & Asian
290	Other Pacific Islander & Korean	Directly mapped (racial category)	Other Pacific Islander & Asian
291	Other Pacific Islander & Vietnamese	Directly mapped (racial category)	Other Pacific Islander & Asian
292	Other Pacific Islander & Other Asian	Directly mapped (racial category)	Other Pacific Islander & Asian
293	Other Pacific Islander & Asian (nonspecific)	Directly mapped (racial category)	Other Pacific Islander & Asian

Table B.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
307	Asian Indian & Chinese	Directly mapped (racial category)	Multiple Asian
308	Asian Indian & Filipino	Directly mapped (racial category)	Multiple Asian
309	Asian Indian & Japanese	Directly mapped (racial category)	Multiple Asian
310	Asian Indian & Korean	Directly mapped (racial category)	Multiple Asian
311	Asian Indian & Vietnamese	Directly mapped (racial category)	Multiple Asian
312	Asian Indian & Other Asian	Directly mapped (racial category)	Multiple Asian
328	Chinese & Filipino	Directly mapped (racial category)	Multiple Asian
329	Chinese & Japanese	Directly mapped (racial category)	Multiple Asian
330	Chinese & Korean	Directly mapped (racial category)	Multiple Asian
331	Chinese & Vietnamese	Directly mapped (racial category)	Multiple Asian
332	Chinese & Other Asian	Directly mapped (racial category)	Multiple Asian
349	Filipino & Japanese	Directly mapped (racial category)	Multiple Asian
350	Filipino & Korean	Directly mapped (racial category)	Multiple Asian
351	Filipino & Vietnamese	Directly mapped (racial category)	Multiple Asian
352	Filipino & Other Asian	Directly mapped (racial category)	Multiple Asian
360	Japanese & Korean	Directly mapped (racial category)	Multiple Asian
361	Japanese & Vietnamese	Directly mapped (racial category)	Multiple Asian
362	Japanese & Other Asian	Directly mapped (racial category)	Multiple Asian
371	Korean & Vietnamese	Directly mapped (racial category)	Multiple Asian
372	Korean & Other Asian	Directly mapped (racial category)	Multiple Asian
382	Vietnamese & Other Asian	Directly mapped (racial category)	Multiple Asian

Table B.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 & 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 & Other Pacific Islander
385	Indian (Asian or American unclear) & Chinese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
386	Indian (Asian or American unclear) & Filipino	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
387	Indian (Asian or American unclear) & Japanese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
388	Indian (Asian or American unclear) & Korean	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
389	Indian (Asian or American unclear) & Vietnamese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
390	Indian (Asian or American unclear) & Other Asian	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
401	White, Black/African American, American Indian/Alaska Native	Directly mapped (racial category)	White, Black/African American, American Indian

Table B.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
405	White, Black/African American, Chinese	Directly mapped (racial category)	White, Black/African American, Asian
406	White, Black/African American, Filipino	Directly mapped (racial category)	White, Black/African American, Asian
407	White, Black/African American, Japanese	Directly mapped (racial category)	White, Black/African American, Asian
408	White, Black/African American, Korean	Directly mapped (racial category)	White, Black/African American, Asian
409	White, Black/African American, Vietnamese	Directly mapped (racial category)	White, Black/African American, Asian
410	White, Black/African American, Other Asian	Directly mapped (racial category)	White, Black/African American, Asian
411	White, Black/African American, Asian (nonspecific)	Directly mapped (racial category)	White, Black/African American, Asian
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
415	White, American Indian/Alaska Native, Chinese	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
416	White, American Indian/Alaska Native, Filipino	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian

Table B.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, Asian
418	White, American Indian/Alaska Native, Korean	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
419	White, American Indian/Alaska Native, Vietnamese	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
420	White, American Indian/Alaska Native, Other Asian	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
421	White, American Indian/Alaska Native, Asian (nonspecific)	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
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
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
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 B.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 B.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 B.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 B.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 B.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 B.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 B.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 B.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 B.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 B.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 B.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 B.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 B.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
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)[I]	QD05ASIA: O.A.
873	Panamanian & Jamaican	Indirectly mapped (QD05) (double census)I	QD05ASIA: O.A.

Table B.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
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.
877	Puerto Rican & Virgin Islander	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.
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

Table B.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
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: Other Asian
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.
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

Table B.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
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 & Asian
923	Arab/African	Directly mapped (geographic category)	QD05: White & Black/African American QD05ASIA: Asian & Black/African American
924	Arab/Chaldean	Directly mapped (racial category)	QD05: White & Asian QD05ASIA: Other Asian
925	European & Asian Indian	Directly mapped (racial category)	White & Asian
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
935	Cuban & Spanish (from Spain)	Codes informative for formal imputation procedures	Not a Direct Map
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)

Table B.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
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 B.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 B.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 B.2.1.2 for details.

2 The abbreviation "O.A." is equivalent to "Other Asian."

Table B.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
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: 0.6% White, 39.5% Black/African American, 40.3% Asian Indian, 1.2% Chinese, 18.4% Black/African American and Asian Indian QD05ASIA: 69% Asian Indian, 31% Black/African American and Asian Indian
55	Jamaica	QD05: 3.2% White, 76.3% Black/African American, 1.5% Asian Indian, .6% Chinese, 15.1% White and Black/African American, 1.5% Black/African American and Asian Indian, .6% Black/African American and Chinese, 1.2% Unrestricted Imputation QD05ASIA: 36% Asian Indian, 36% Black/African American and Asian Indian, 14% Chinese, 14% Black/African American and Chinese
56	Virgin Is (St Thomas, St Croix)	QD05: 12% White, 85% Black/African American, 3% Asian Nonspecific QD05ASIA: Impute among Asians
57	Bahamas	QD05: 12% White, 85% Black/African American, 3% Asian Nonspecific QD05ASIA: Impute among Asians
58	Barbados	4% White, 90% Black/African American, 6% Unrestricted Imputation
59	Grenada	82% Black/African American, 13% White and Black/African American, 5% Unrestricted Imputation
60	St. Lucia	QD05: 1% White, 90% Black/African American, 3% Asian Indian, 3% White and Black/African American, 3% Black/African American and Asian Indian QD05ASIA: 50% Asian Indian 50% 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
64	Brazil	55.3% White, 6% Black/African American, .3% American Indian/Alaska Native, .3% Japanese, 38% White and Black/African American
65	Canada	QD05: 66% White, 2% American Indian/Alaska Native, 32% Unrestricted Imputation QD05ASIA: Impute among Asians

Table B.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
70	Mexico	9.3% White, 30.3% American Indian/Alaska Native, 60.3% White and American Indian/Alaska Native
71	Puerto Rico	QD05: 82.7% White, 10.2% Black/African American, .4% American Indian/Alaska Native, .01% Native Hawaiian, .02% Other Pacific Islander, .13% Asian Indian, .05% Chinese, .01% Filipino, .01% Japanese, .01% Korean, .01% Vietnamese, 6.4% White and Black/African American QD05ASIA: 59% Asian Indian, 23% Chinese, 4.5% each Filipino, Japanese, Korean, Vietnamese
72	Cuba	37% White, 11% Black/African American, 1% Chinese, 51% White and Black/African American
73	Dominican Republic	16% White, 11% Black/African American, 73% White and Black/African American
74	Guatemala	43% American Indian/Alaska Native, 55% White and American Indian/Alaska Native, 2% Unrestricted Imputation
75	Honduras	1% White, 2% Black/African American, 7% American Indian/Alaska Native, 90% White and American Indian/Alaska Native
76	El Salvador	9% White, 1% American Indian/Alaska Native, 90% 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
78	Costa Rica	QD05: 3% Black/African American, 1% American Indian/Alaska Native, 1% Chinese, 94% White or Mestizo, 1% Unrestricted Imputation QD05 when in combination with another race: 47.2% White, 3.2% Black/African American, 1.2% American Indian/Alaska Native, 1.2% Chinese, 47.2% 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	7% White, 3% Black/African American, 25% American Indian/Alaska Native, 65% White and American Indian/Alaska Native
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

Table B.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
85	Chile	3% American Indian/Alaska Native, 95% White or Mestizo, 2% Unrestricted Imputation
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	97.2% White, 1.4% Black/African American, 1.4% Asian Indian
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
130	Bahrain	QD05: 73% White, 8% Other Asian, 19% Asian Nonspecific QD05ASIA: 81% Other Asian, 19% Impute among Asian Groups
133	Kuwait	QD05ASIA: 9% Asian Indian, 91% Other Asian
140	Kazakhstan	36.1% White, 57.3% Other Asian, 6.6% 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	8% Asian Indian, 24% Chinese, 58% Other Asian, 10% Asian Nonspecific
159	Singapore	7.9% Asian Indian, 76.7% Chinese, 14% Other Asian, 1.4% Not American Indian/Alaska Native
165	Djibouti	2.5% White, 97.5% Black/African American
166	Sudan	39% White, 58% Black/African American, 3% Not American Indian/Alaska Native
168	South Africa	13.6% White, 75.2% Black/African American, 2.6% Asian Indian, 8.6% White and Black/African American
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

Table B.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
171	Zambia	1.1% White, 98.7% Black/African American, .2% 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% White, 91% Black/African American
186	Australia	QD05: 92% White, 7% Asian Nonspecific, 1% Not American Indian/Alaska Native QD05ASIA: Impute among Asians
187	New Zealand	QD05: 79.1% White, 13.5% Other Pacific Islander, 7.4% Asian Nonspecific 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 B.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 B.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 B.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 B.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
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
209	Cuban/West Indies	Cuban
210	Dominican/West Indies	Dominican

Table B.4 Mapping of Hispanic/Latino Group Codes (continued)

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
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 B.1.

Appendix C: Model Summaries

Appendix C: Model Summaries

C.1 Introduction

The tables in this appendix list the covariates used in all the imputation models that were run in the 2011 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 C.3 and the models for the drug variables are presented in Section C.4. With the exception of the lifetime usage models, separate tables are provided in Section C.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 C.5. Section C.6 presents the models for the income variables and Section C.7 presents the models for the health insurance variables. Section C.7 also presents the models for both the "Old Method" and the "Constituent Variables Method," used to create the final imputation-revised health insurance variables. Chapter 9 provides a more detailed description of these two methods. Section C.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 C.3 (Chapters 3 & 4), Section C.4 (Chapters 5 & 6), Section C.5 (Chapter 7), Section C.6 (Chapter 8), Section C.7 (Chapter 9), and Section C.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.

C.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.

C.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.

C.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.

C.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 persons; segment in CBSA with 250,000 to 999,999 persons; segment in CBSA with fewer than 250,000 persons; segment not in CBSA and not in rural area; and segment not in CBSA and in rural area.

C.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.

C.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.

---

[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.

C.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 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.

C.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.

C.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.

C.3 Demographic Variables

For justifications of the aggregation of age groups for certain imputation steps, see Chapters 3 and 4.

Exhibit C.1 Definitions of Levels for Variables

Census Region
N: Northeast, M: Midwest, S: South, W: West[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]
Percentage Hispanic/Latino in Segment
H: ≥ 70%, M: [20,70)%, L: < 20%[1]
Percentage Black/African American in Segment
H: ≥ 40%, M: [10,40)%, L: < 10%[1]
Percentage American Indian/Alaska Native in Segment
H: ≥ 3%, M: [1,3)%, L: < 1%[1]
Percentage Asian in Segment
H: ≥ 10%, M: [5,10)%, L: < 5%[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 Only, MR: Multiple Race

Exhibit C.1 Definitions of Levels for Variables (continued)

Race/Hispanic Recode (8 Levels)
W: Not Hispanic/Latino and White,[1] B: Not Hispanic/Latino and Black/African American, A: Not Hispanic/Latino and Asian, AI: Not Hispanic/Latino and American Indian/Alaska Native, MR: Not Hispanic/Latino and Multiple Race, PR: Hispanic/Latino and Puerto Rican, M: Hispanic/Latino and Mexican, O: Not Hispanic/Latino and Other

Population Density
LC: Segment in a CBSA with 1 Million or More Persons, MC: Segment in a CBSA from 250,000 to 999,999 Persons, SC: Segment in a CBSA with Fewer than 250,000 Persons, NC: Segment Not in a CBSA and Not in a Rural Area, R: Segment Not in a CBSA and in a Rural Area[1]

MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]

Gender
M: Male, F: Female[1]

Age Category
Y: 12-17, T: 18-25, A1: 26-34, A2: 35-49, A3: 50+[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, 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,[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]

Imputation-Revised Education Level
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]

Imputation-Revised Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]

Imputation-Revised Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]

Imputation-Revised Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]

Imputation-Revised Hispanic/Latino Origin Indicator
H: Hispanic/Latino, NH: Not Hispanic/Latino[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 C.1 Summaries for Response Propensity Models

Imputation Step	Variables Included in Response Propensity Model
Marital Status	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Population Density (LC, MC, SC, NC); Gender (M); Age Category (Y, T, A1, A2); Age Category * Gender (MY, MT, MA1, MA2)
Race 12-17	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M)
Race 18-25	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Imputation-Revised Marital Status (3 Levels) (M, WM)
Race 26+	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Age Category (A1, A2); Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Hispanic/Latino Origin 12-17	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR)
Hispanic/Latino Origin 18-25	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR)
Hispanic/Latino Origin 26+	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Age Category (A1, A2)
Hispanic/Latino Group	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Gender (M); Age Category (Y, T, A1, A2); Age Category * Gender (MY, MT, MA1, MA2)
Education Level 12-17	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M)
Education Level 18+	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age Category (T, A1, A2); Age Category * Gender (MT, MA1, MA2)

Table C.1 Summaries for Response Propensity Models (continued)

Imputation Step	Variables Included in Response Propensity Model
Employment Status 12-25	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age Category (Y); Age Category * Gender (MY); Imputation-Revised Education Level (L, HS, SC)
Employment Status 26+	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age Category (A1, A2); Age Category * Gender (MA1, MA2); Imputation-Revised Education Level (L, HS, SC)
Born in US 12-17	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); **Imputation-Revised Education Level (L, HS, SC)**; **Imputation-Revised Employment Status (FT, PT, UN)**
Born in US 18-25	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Imputation-Revised Education Level (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)
Born in US 26+	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age Category (A1, A2); Age Category * Gender (MA1, MA2); Imputation-Revised Education Level (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)
Age of Entry	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age Category (Y, T, A1, A2); Age Category * Gender (MY, MT, MA1, MA2); Race/Hispanic Recode (8 Levels) * Gender (MB, MA, MAI, MMR, MPR, MM, MO); Imputation-Revised Education Level (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.1 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.2 Summaries for Predictive Mean Models

Imputation Step	Variables Included in Predictive Mean Model
Marital Status	Census Region (N, M, S); Age; **Age Squared**; **Age Cubed**; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Population Density (LC, MC, SC, NC); Gender (M); Age * Gender
Race 12-17	Census Region (N, M, S); **Race/Ethnicity of Householder (B, H)**; Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M)
Race 18-25	Census Region (N, M, S); **Race/Ethnicity of Householder (B, H)**; **Age**; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Imputation-Revised Marital Status (3 Levels) (M, WM)
Race 26+	Census Region (N, M, S); **Race/Ethnicity of Householder (B, H)**; **Age**; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); **Percentage Black/African American in Segment (H, M)**; Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); **Imputation-Revised Marital Status (4 Levels) (M, W, DS)**
Hispanic/Latino Origin 12-17	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR)
Hispanic/Latino Origin 18-25	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Age; Age Squared; Age Cubed; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Imputation-Revised Marital Status (3 Levels) (M, WM)
Hispanic/Latino Origin 26+	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Age; Age Squared; Age Cubed; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Hispanic/Latino Group	Census Region (N, M, S); **Age**; **Age Squared**; **Age Cubed**; **Percentage Owner Occupied in Segment (H, M)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); **Percentage American Indian/Alaska Native in Segment (H, M)**; **Percentage Asian in Segment (H, M)**; Race/Hispanic Recode (5 Levels) (B, AI, A, MR); **Gender (M)**; **Age * Gender**; **Age Squared * Gender**
Education Level 12-17	Census Region (N, M, S); **Age**; **Age Squared**; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M)

Table C.2 Summaries for Predictive Mean Models (continued)

Imputation Step	Variables Included in Predictive Mean Model
Education Level 18+	Census Region (N, M, S); Age; Age Squared; Age Cubed; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Employment Status 12-25	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Imputation-Revised Education Level (L, HS, SC)
Employment Status 26+	Census Region (N, M, S); Age; **Age Squared**; **Age Cubed**; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; **Age Squared * Gender**; Imputation-Revised Education Level (L, HS, SC); Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Born in US 12-17	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age * Gender; Age Squared * Gender; **Imputation-Revised Education Level (L, HS, SC)**; **Imputation-Revised Employment Status (FT, PT, UN)**
Born in US 18-25	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age * Gender; Age Squared * Gender; Imputation-Revised Education Level (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)
Born in US 26+	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age * Gender; Age Squared * Gender; Imputation-Revised Education Level (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)
Age of Entry	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age * Gender; Age Squared * Gender; Race/Hispanic Recode (8 Levels) * Gender (MB, MA, MAI, MMR, MPR, MM, MO); Imputation-Revised Education Level (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.1 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

C.4 Drug Variables

Exhibit C.2 Definitions of Levels for Variables

Age Category
A1: 26-34, A2: 35-49, A3: 50+[1]
Gender
M: Male, F: Female[1]
Race/Hispanic Recode (2 Levels)[2]
W: Not Hispanic/Latino and White Only, O: Hispanic/Latino or Not White Only[1]
Race/Hispanic Recode (4 Levels)[2]
H: Hispanic/Latino,[1] O: Not Hispanic/Latino and Other, B: Not Hispanic/Latino and Black/African American Only, W: Not Hispanic/Latino and White Only[1]
Gender * Race/Hispanic Recode (4 Levels)[2]
MH: Male Hispanic/Latino,[1] MO: Male Not Hispanic/Latino and Other, MB: Male Not Hispanic/Latino and Black/African American Only, MW: Male Not Hispanic/Latino and White Only,[1] FH: Female Hispanic/Latino,[1] FO: Female Not Hispanic/Latino and Other,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FW: Female Not Hispanic/Latino and White Only[1]
Age * Race/Hispanic Recode (4 Levels)[2]
H: Age * Hispanic/Latino,[1] O: Age * Not Hispanic/Latino and Other, B: Age * Not Hispanic/Latino and Black/African American Only, W: Age * Not Hispanic/Latino and White Only[1]
Age Squared * Race/Hispanic Recode (4 Levels)
H: Age Squared * Hispanic/Latino, O: Age Squared * Not Hispanic/Latino and Other, B: Age Squared * Not Hispanic/Latino and Black/African American Only, W: Age Squared * Not Hispanic/Latino and White Only[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]
Education Level
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]
Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]
State Rank
L: Low State Rank (lowest tertile), M: Middle State Rank (middle tertile), H: High State Rank (highest tertile)[1]
Cigarette Lifetime Indicator
Y: Yes, N: No[1]
Intermediate Drug-Specific Past Month Indicator
Y: Yes, N: No[1]
Intermediate Drug-Specific Lifetime Indicator
Y: Yes, N: No[1]
Imputation-Revised Drug-Specific Lifetime Indicator
Y: Yes, N: No[1]
Imputation-Revised Pipe Recency
PM: Past Month, LF: Lifetime but Not Past Month, NU: Lifetime Nonuser[1]
Imputation-Revised 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]
Imputation-Revised 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]

Exhibit C.2 Definitions of Levels for Variables (continued)

Imputation-Revised 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]
Imputation-Revised 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]

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 C.3 Lifetime Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12-17	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); Cigarette Lifetime Indicator (Y)
18-25	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Cigarette Lifetime Indicator (Y)
26+	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Cigarette Lifetime Indicator (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.4 Cigarettes: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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)
12-Month Frequency	N/A	N/A
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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)	Daily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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) Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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)

Table C.4 Cigarettes: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)
Ever Daily Used	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Age at First Use for Cigarettes; Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.5 Cigarettes: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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)
12-Month Frequency	N/A	N/A
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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)	Daily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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)

Table C.5 Cigarettes: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

Table C.5 Cigarettes: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Age at First Use for Cigarettes; Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.6 Cigarettes: Persons 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); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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)
12-Month Frequency	N/A	N/A
30-Day Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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)	Daily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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)

Table C.6 Cigarettes: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

Table C.6 Cigarettes: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Age at First Use for Cigarettes; Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.7 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender
Recency	Smokeless Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Smokeless Tobacco: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
	Chewing Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Chewing Tobacco: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.7 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	Chewing Tobacco: Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; Census Region (N, M, S); MSA (R, SM); **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes (**PM, PY, **P3Y**, **LF)**	Chewing Tobacco Daily Users: **Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; Census Region (N, M, S); **MSA (R**, **SM)**; **State Rank (**L, **M)**; **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes** (PM, **PY**, **P3Y**, **LF)** Chewing Tobacco Nondaily Users: Age; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); Imputation-Revised Drug-Specific Lifetime Indicator for Alcohol (Y), Heroin (Y), and Hallucinogens (Y)

Table C.7 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)** for **Cigarettes (PM, PY, P3Y, LF)**	Snuff Daily Users: **Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Census Region (N, M,** S); **MSA (R, SM)**; **State Rank** (L, **M)**; **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF) Snuff Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.7 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, and Daily Cigarette Use; Imputation-Revised Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Chewing Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.8 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender
Recency	Smokeless Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Smokeless Tobacco: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
	Chewing Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Chewing Tobacco: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW**)**; **Age * Race/Hispanic Recode (4 Levels) (O**, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.8 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	Chewing Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Chewing Tobacco Daily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, MW); **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.8 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Chewing Tobacco Nondaily Users: Age; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, MW**)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, W); Census Region (N, M, S); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y)
	Snuff: Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff Daily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**H, O, **B)**; **Gender * Race/Hispanic Recode (4 Levels) (**MH, MO, **MB)**; **Age * Race/Hispanic Recode (4 Levels) (**H, O, **B)**; Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
		Snuff Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW**)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.8 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, and Daily Cigarette Use; Imputation-Revised Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Chewing Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.9 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender
Recency	Smokeless Tobacco: Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Smokeless Tobacco: Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Marital Status (4 Levels)** (M, **W, DS)**; Education Level (L, HS, SC); **Employment Status (**FT, PT, **UN)**; **Census Region (N, M, S)**; MSA (R, SM); State Rank (L, M); **Age * Gender**; **Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
	Chewing Tobacco: Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels)** (O, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels)** (MO, MB, **MW)**; **Marital Status (4 Levels)** (M, **W**, **DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**	Chewing Tobacco: Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, P3Y, **LF)**

Table C.9 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff: Age; Age Squared; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; **Education Level (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	Chewing Tobacco: **Age Category (**A1, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, M, S); **MSA (R, SM)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Chewing Tobacco Daily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**O, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (**O, B, **W)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF) Chewing Tobacco Nondaily Users: Age; Race/Hispanic Recode (4 Levels) (O, B, W); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.9 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: **Age Category (A1**, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, W**)**; Census Region (N, M, S); **MSA (**R, **SM)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Snuff Daily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**O, **B**, W**)**; **Gender * Race/Hispanic Recode (4 Levels) (**MO, **MB**, MW**)**; **Age * Race/Hispanic Recode (4 Levels) (**O, **B**, W**)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)** Snuff Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**O, **B**, W**)**; **Gender * Race/Hispanic Recode (4 Levels) (**MO, **MB**, MW**)**; **Age * Race/Hispanic Recode (4 Levels) (**O, **B**, W**)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**

Table C.9 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, and Daily Cigarette Use; Imputation-Revised Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Chewing Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.10 Cigars: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF)**
12-Month Frequency	N/A	N/A
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised 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 Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.10 Cigars: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cigars; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.11 Cigars: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A

Table C.11 Cigars: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised 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 Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised 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 at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cigars; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table C.12 Cigars: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y)
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A

Table C.12 Cigars: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised 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 Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised 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 at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cigars; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table C.13 Pipes: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y)
Recency	Gender (M); **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **State Rank (L, M)**; **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
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 C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.14 Pipes: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y)
Recency	Gender (M); **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Marital Status (3 Levels) (M, WM)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; **MSA (R, SM)**; State Rank (L, M); **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
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 C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.15 Pipes: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y)
Recency	**Age Category (A1, A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels)** (M, **W, DS**); **Education Level** (L, HS, SC); **Employment Status (FT**, PT, **UN)**; **Census Region (N, M**, S); **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
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 C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.16 Alcohol: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.16 Alcohol: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Binge Drinking Frequency	N/A	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Intermediate Alcohol 30-Day Frequency; Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.16 Alcohol: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.17 Alcohol: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.17 Alcohol: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.17 Alcohol: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Binge Drinking Frequency	N/A	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Intermediate Alcohol 30-Day Frequency; Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.17 Alcohol: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.18 Alcohol: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.18 Alcohol: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.18 Alcohol: Persons 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	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Intermediate Alcohol 30-Day Frequency; Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.18 Alcohol: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Alcohol; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

.

Table C.19 Inhalants: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.19 Inhalants: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants**; Imputation-Revised 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); **Imputation-Revised Pipe Recency (PM, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.19 Inhalants: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Inhalants; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.20 Inhalants: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.20 Inhalants: Persons 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	Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW); **Marital Status (3 Levels) (M**, **WM)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r **Alcohol (**PM, PY, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**H, **O**, B**)**; **Gender * Race/Hispanic Recode (4 Levels) (**MH, **MO**, MB**)**; **Age * Race/Hispanic Recode (4 Levels) (**H, **O**, B**)**; **Marital Status (3 Levels) (**M, **WM)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r **Alcohol (**PM, PY, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); **Imputation-Revised Drug-Specific Lifetime Indicator fo**r Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, Heroin (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r Cigarettes (PM, PY, P3Y, LF), **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, PY, P3Y, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.20 Inhalants: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Hallucinogens (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r **Alcohol (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Race/Hispanic Recode (4 Levels) (O, B, W); Education Level (L, HS, SC); Census Region (N, M, S); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants

Table C.20 Inhalants: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Inhalants; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.21 Inhalants: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); **Marital Status (4 Levels) (M, W, DS)**; **Education Level (L, HS, SC)**; Employment Status (FT, PT, UN); **Census Region (N, M, S)**; **MSA (R, SM)**; State Rank (L, M); **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; **Gender (M)**; **Race/Hispanic Recode (4 Levels)** (O, **B**, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels)** (O, **B**, **W)**; **Marital Status (4 Levels) (M, W, DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; **Age * Gender**; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, P3Y, **LF)**

Table C.21 Inhalants: Persons 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	**Age Category (A1, A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (O, B,** W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **State Rank (L, M)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **State Rank (L, M)**; **Age * Gender**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**
12-Month Frequency	**Age Category (A1, A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (O, B,** W); **Census Region (**N, **M**, S); **MSA (R, SM)**; Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); **Imputation-Revised Drug-Specific Lifetime Indicator fo**r Marijuana (Y), **Pain Relievers (Y)**, Tranquilizers (Y), **Stimulants (Y)**, Sedatives (Y), **Cocaine (Y)**, Crack (Y), **Heroin (Y)**, and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**, **Cigars (PM, PY, P3Y, LF)**, and **Smokeless Tobacco (PM, PY, P3Y, LF)**	Age; Race/Hispanic Recode (2 Levels) (W); Education Level (L); **Employment Status (**FT, **PT**, UN**)**; Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), and Pain Relievers (Y)

Table C.21 Inhalants: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Age Category (A1**, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, **M**, S); **MSA (R**, **SM)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r Marijuana (Y), **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), **Heroin (Y)**, and Hallucinogens (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; **Employment Status (**FT, **PT**, UN**)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), and Crack (Y)

Table C.21 Inhalants: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Inhalants; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.22 Marijuana: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.22 Marijuana: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.22 Marijuana: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Marijuana; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.23 Marijuana: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.23 Marijuana: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.23 Marijuana: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Marijuana; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.24 Marijuana: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.24 Marijuana: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r Alcohol (PM, PY, LF), and **Inhalants (**PM, **PY**, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.24 Marijuana: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Marijuana; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.25 Hallucinogens: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.25 Hallucinogens: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); **Race/Hispanic Recode (4 Levels)** (O, **B**, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; **Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y)**, **Stimulants (Y)**, Sedatives (Y), **Cocaine (Y)**, Crack (Y), Heroin (Y), and **Tranquilizers (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r Alcohol (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, and **Inhalants** (PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (**PM, PY, P3Y, **LF)**	**Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; Census Region (N, M, S); **MSA (**R, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r Pain Relievers (Y), **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, Heroin (Y), and **Tranquilizers (Y)**; **Imputation-Revised Pipe Recency** (PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r Alcohol (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**

Table C.25 Hallucinogens: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Hallucinogens; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.26 Hallucinogens: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.26 Hallucinogens: Persons 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	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.26 Hallucinogens: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (**O, B, **W)**; Census Region (N, M, S); **MSA (**R, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users fo**r **Hallucinogens**; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r Pain Relievers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), Crack (Y), Heroin (Y), and **Tranquilizers (Y)**; **Imputation-Revised Pipe Recency (**PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r **Alcohol (PM**, **PY**, **LF)**, **Marijuana (**PM, PY, **LF)**, and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r Cigarettes (PM, PY, P3Y, LF), **Cigars (**PM, **PY**, **P3Y**, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.26 Hallucinogens: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Hallucinogens; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.27 Hallucinogens: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); **Marital Status (4 Levels) (M, W, DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Cigarette Lifetime Indicator (Y)**; Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, **Chewing Tobacco (Y)**, Snuff (Y), and **Pipes (Y)**
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; **Age Squared**; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Marital Status (4 Levels)** (M, **W**, **DS**); **Education Level (L, HS, SC)**; **Employment Status (FT, PT**, UN); **Census Region (N, M, S)**; **MSA (R, SM)**; **State Rank (L, M)**; **Age * Gender**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Marijuana** (PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**

Table C.27 Hallucinogens: Persons 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	**Age Category (A1**, **A2)**; **Gender (M)**; Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; Census Region (N, M, S); **MSA (**R, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW); **Age * Race/Hispanic Recode (4 Levels) (O**, B, W); **Marital Status (4 Levels)** (M, **W**, DS); Education Level (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; **Census Region (**N, M, **S)**; MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**
12-Month Frequency	**Age Category (A1**, **A2)**; **Gender (M)**; Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); **MSA (**R, **SM)**; Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for** Pain Relievers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, Heroin (Y), and **Tranquilizers (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Race/Hispanic Recode (4 Levels) (O, B, W); Education Level (L, HS, SC); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), and Stimulants (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY, LF)

Table C.27 Hallucinogens: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Age Category (**A1, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens**; **Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), **Heroin (Y)**, and **Tranquilizers (Y)**; **Imputation-Revised Pipe Recency (PM**, LF); **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Race/Hispanic Recode (2 Levels) (W); Marital Status (3 Levels) (M); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; **Imputation-Revised Pipe Recency (PM)**

Table C.27 Hallucinogens: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Hallucinogens; Imputation-Revised 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); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.28 Pain Relievers: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.28 Pain Relievers: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.28 Pain Relievers: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.29 Pain Relievers: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.29 Pain Relievers: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.29 Pain Relievers: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.30 Pain Relievers: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.30 Pain Relievers: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.30 Pain Relievers: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.31 Tranquilizers: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; Cigarette Lifetime Indicator (Y); Age * Gender; **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)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.31 Tranquilizers: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.31 Tranquilizers: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.32 Tranquilizers: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.32 Tranquilizers: Persons 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	Gender (M); **Race/Hispanic Recode (4 Levels) (**O, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (**MO, **MB**, **MW)**; **Marital Status (3 Levels) (**M, **WM)**; Education Level (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; Census Region (N, M, S); **MSA (R**, **SM)**; State Rank (L, M); **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**, Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.32 Tranquilizers: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.33 Tranquilizers: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.33 Tranquilizers: Persons 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	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (**MO, MB, **MW)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R, SM)**; State Rank (L, M); **Imputation-Revised 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)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); **MSA (R, SM)**; Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); **Imputation-Revised Pipe Recency (PM, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Marijuana (**PM, **PY**, **LF)**, **Hallucinogens (PM, PY, LF)**, Pain Relievers (PM, PY, LF), and **Inhalants (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM, PY, P3Y, LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.33 Tranquilizers: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.34 Stimulants: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator fo**r Cigars (Y), Alcohol (Y), **Inhalants (Y)**, Marijuana (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.34 Stimulants: Persons 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	Gender (M); **Race/Hispanic Recode (4 Levels) (**O, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes (**PM, PY, P3Y, **LF)**
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.34 Stimulants: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.35 Stimulants: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.35 Stimulants: Persons 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	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, **W)**; Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); **Imputation-Revised Drug-Specific Lifetime Indicator fo**r Sedatives (Y), Crack (Y), Heroin (Y), and **Cocaine (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r Alcohol (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), **Pain Relievers (**PM, PY, **LF)**, Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**, Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.35 Stimulants: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.36 Stimulants: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.36 Stimulants: Persons 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	Age Category (A1, A2); **Gender (M)**; Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level** (L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; Census Region (N, M, S); **MSA (R**, **SM)**; State Rank (L, M); **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Marital Status (4 Levels)** (M, **W**, **DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	**Age Category (A1**, **A2)**; **Gender (M)**; Race/Hispanic Recode (4 Levels) (O, B, W); **Census Region** (N, M, **S)**; **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for** Sedatives (Y), **Crack (Y)**, Heroin (Y), and **Cocaine (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, Pain Relievers (PM, PY, LF), **Tranquilizers (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars** (PM, PY, **P3Y**, **LF)**, and **Smokeless Tobacco** (PM, PY, **P3Y**, **LF)**	Age; Age Squared; Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Heroin (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM)
30-Day Frequency	N/A	N/A

Table C.36 Stimulants: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.37 Sedatives: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; Cigarette Lifetime Indicator (Y); Age * Gender; **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)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); **MSA (R, SM)**; State Rank (L, M); **Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.37 Sedatives: Persons 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	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB**, **MW)**; Census Region (N, M, S); **MSA (R**, **SM)**; **State Rank (L, M)**; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r Alcohol (PM, PY, LF), and **Marijuana (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes (**PM, PY, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator fo**r 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r **Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes (**PM, PY, **P3Y**, **LF)**
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Census Region (N**, M, **S)**; **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Imputation-Revised Drug-Specific Lifetime Indicator fo**r **Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r Alcohol (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (**PM, **PY**, **LF)**, **Tranquilizers (**PM, PY, **LF)**, **Stimulants (PM**, **PY**, **LF)**, and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r **Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**O, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (**MO, **MB**, MW); **Age * Race/Hispanic Recode (4 Levels) (**O, **B**, W); Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y)**, Heroin (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) fo**r Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), **Pain Relievers (**PM, PY, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, and **Inhalants (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) fo**r Cigarettes (PM, PY, P3Y, LF), **Cigars (**PM, PY, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**
30-Day Frequency	N/A	N/A

Table C.37 Sedatives: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); **MSA (R, SM)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cocaine (Y), Heroin (Y), and **Crack (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Marijuana (PM, PY, LF)**, Hallucinogens (PM, PY, LF), **Pain Relievers (PM, PY, LF)**, Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives** (PM, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco** (PM, PY, **P3Y, LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), Daily Cigarette Use (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.38 Sedatives: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)
Recency: Past Year vs. Not Past Year	**Gender (M)**; Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Marital Status (3 Levels) (M, WM)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; **MSA (R, SM)**; State Rank (L, M); **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, P3Y, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.38 Sedatives: Persons 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	**Gender (M)**; **Race/Hispanic Recode (4 Levels)** (O, **B**, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Marital Status (3 Levels) (M**, **WM)**; Education Level (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; State Rank (L, M); **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)** for **Alcohol (**PM, **PY**, **LF)**, and **Marijuana** (PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)** for **Cigarettes (**PM, PY, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Marital Status (3 Levels) (M**, **WM)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)** for **Alcohol (PM**, **PY**, **LF)**, and **Marijuana** (PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)** for **Cigarettes (**PM, **PY**, **P3Y**, **LF)**
12-Month Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels)** (O, **B**, **W)**; **Census Region (N**, M, S**)**; **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Imputation-Revised Drug-Specific Lifetime Indicator** for **Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)** for **Alcohol (PM**, **PY**, **LF)**, Marijuana (PM, PY, LF), **Hallucinogens (PM**, **PY**, **LF)**, Pain Relievers (PM, PY, LF), **Tranquilizers** (PM, PY, **LF)**, **Stimulants (**PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)** for **Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Imputation-Revised Drug-Specific Lifetime Indicator** for **Cocaine (Y)**, Heroin (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)** for **Alcohol (PM**, **PY**, **LF)**, Marijuana (PM, PY, LF), **Hallucinogens (PM**, **PY**, **LF)**, Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), **Stimulants** (PM, PY, **LF)**, and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)** for Cigarettes (PM, PY, P3Y, LF), **Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**
30-Day Frequency	N/A	N/A

Table C.38 Sedatives: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); **Race/Hispanic Recode (4 Levels)** (O, B, **W**); **Census Region (N, M, S)**; **MSA (R, SM)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; **Imputation-Revised Pipe Recency (PM, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF**), **Marijuana (PM, PY, LF)**, **Hallucinogens** (PM, PY, **LF**), **Pain Relievers** (PM, **PY**, **LF**), **Tranquilizers** (PM, **PY**, **LF**), **Stimulants** (PM, **PY**, **LF**), and **Inhalants** (PM, PY, **LF**); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF**), Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (PM, PY, P3Y, LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), Daily Cigarette Use (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.39 Sedatives: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, MW); **Marital Status (4 Levels)** (M, **W**, **DS)**; Education Level (L, HS, SC); **Employment Status** (FT, PT, **UN)**; Census Region (N, M, S); **MSA (R**, **SM)**; State Rank (L, M); **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana** (PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.39 Sedatives: Persons 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	**Age Category** (A1, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Marital Status (4 Levels)** (M, W, DS); **Education Level (L**, HS, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Marital Status (4 Levels)** (M, **W**, **DS)**; **Education Level** (L, HS, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; Age * Gender; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	**Age Category** (A1, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y)**, Heroin (Y), and Crack (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers** (PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Gender (M); Marital Status (3 Levels) (M); Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table C.39 Sedatives: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels)** (O, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Age Squared * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Marital Status (4 Levels) (M**, W, DS); **Education Level** (L, HS, **SC**); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; State Rank (L, M); **Age * Gender**; **Age Squared * Gender**; **Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives**; **Drug-Specific Imputation-Revised Age at First Use for Cigarettes**, **Cigars**, **Alcohol**, **Inhalants**, **Marijuana**, **Hallucinogens**, **Pain Relievers**, **Tranquilizers**, **Stimulants**, **Daily Cigarette Use**, and **Smokeless Tobacco**; **Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y)**, **Heroin (Y)**, **Daily Cigarette Use (Y)**, and Crack (Y); **Imputation-Revised Pipe Recency** (PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers** (PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM**, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, P3Y, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.40 Cocaine (and Crack): Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Cocaine: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; **Cigarette Lifetime Indicator (Y)**; Age * Gender; **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)** Crack: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW); **Age * Race/Hispanic Recode (4 Levels) (O**, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)
Recency	**Gender (M)**; Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Census Region** (N, M, **S)**; MSA (R, SM); **State Rank** (L, **M)**; **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**	Age; Age Squared; Age Cubed; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Marijuana (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**

Table C.40 Cocaine (and Crack): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); **Race/Hispanic Recode (4 Levels)** (O, **B**, W); **Census Region** (N, M, **S**); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and Heroin (Y); **Imputation-Revised Pipe Recency** (PM, **LF**); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Marijuana (PM, PY, LF)**, Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), **Tranquilizers (PM, PY, LF)**, Stimulants (PM, PY, LF), **Sedatives** (PM, **PY**, **LF**), and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**, Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (PM, PY, P3Y, LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Intermediate Drug-Specific Past Month Indicator for Cocaine (Y)**; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); **Imputation-Revised Pipe Recency (PM, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), **Tranquilizers** (PM, **PY**, **LF**), Stimulants (PM, PY, LF), **Sedatives (PM, PY, LF)**, and **Inhalants** (PM, PY, **LF**); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF**), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Census Region (N, M, S)**; **MSA (R, SM)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and **Heroin (Y)**; **Imputation-Revised Pipe Recency (PM, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, Marijuana (PM, PY, LF), **Hallucinogens (PM, PY, LF)**, **Pain Relievers** (PM, **PY**, **LF**), **Tranquilizers (PM, PY, LF)**, **Stimulants (PM, PY, LF)**, **Sedatives (PM, PY, LF)**, and **Inhalants (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**, **Cigars** (PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM, PY, P3Y, LF)**	Age; **Census Region (N**, M, S); **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars** (PM, **PY**, **P3Y**, **LF)**

Table C.40 Cocaine (and Crack): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); **Imputation-Revised Drug-Specific Lifetime Indicator for** Crack (Y), and **Heroin (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF**), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), **Pain Relievers** (PM, PY, **LF**), **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF)**, **Cigars** (PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Age Squared * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cocaine; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM**, **PY)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.41 Cocaine (and Crack): Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Cocaine: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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) Crack: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)

Table C.41 Cocaine (and Crack): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.41 Cocaine (and Crack): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Census Region (**N, M, **S)**; **MSA (**R, **SM)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Marijuana (**PM, PY, **LF)**, **Hallucinogens (**PM, PY, **LF)**, Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), **Sedatives (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table C.41 Cocaine (and Crack): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cocaine; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.42 Cocaine (and Crack): Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Cocaine: Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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) Crack: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)

Table C.42 Cocaine (and Crack): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (**PM, PY, **LF)**, Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), **Stimulants (**PM, **PY**, **LF)**, **Sedatives (**PM, **PY**, **LF)**, and **Inhalants (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	**Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (**N, M, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Intermediate Drug-Specific Past Month Indicator for Cocaine (Y)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and Heroin (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (**PM, PY, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants** (PM, **PY**, **LF)**, **Sedatives** (PM, **PY**, **LF)**, and **Inhalants (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**

Table C.42 Cocaine (and Crack): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Age Category (A1**, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, B, **W)**; Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and **Heroin (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (**PM, PY, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, Tranquilizers (PM, PY, LF), **Stimulants (PM**, **PY**, **LF)**, **Sedatives (**PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**, **Cigars (**PM, PY, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)

Table C.42 Cocaine (and Crack): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cocaine; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.43 Heroin: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, B, W); **Census Region (N**, **M**, **S)**; MSA (R, SM); **State Rank (L, M)**; **Cigarette Lifetime Indicator (Y)**; Age * Gender; **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)**
Recency: Past Year vs. Not Past Year	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Census Region (N**, M, S); **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region** (N, **M**, **S)**; **MSA** (R, **SM)**; State Rank (L, M); **Age * Gender**; **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana** (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**

Table C.43 Heroin: Persons 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	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, PY, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana** (PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Imputation-Revised Pipe Recency** (PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens** (PM, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants** (PM, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; **Race/Hispanic Recode (4 Levels) (O**, **B**, W); **Intermediate Drug-Specific Past Month Indicator for Heroin (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana** (PM, PY, **LF)**

Table C.43 Heroin: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)** for **Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)** for **Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco** (PM, PY, **P3Y**, **LF)**	Age; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM), Crack (PM), and Sedatives (PM)

Table C.43 Heroin: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **Imputation-Revised Pipe Recency** (PM, **LF**); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF**), **Marijuana (PM, PY, LF)**, **Hallucinogens** (PM, **PY**, **LF**), Pain Relievers (PM, PY, LF), **Tranquilizers (PM, PY, LF)**, **Stimulants** (PM, **PY**, **LF**), **Sedatives (PM, PY, LF)**, **Cocaine (PM, PY, LF)**, **Crack (PM, PY, LF)**, and **Inhalants (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**, **Cigars (PM, PY, P3Y, LF)**, and **Smokeless Tobacco (PM, PY, P3Y, LF)**	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Age Squared * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **State Rank (L, M)**; **Age * Gender**; **Age Squared * Gender**; **Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin**; **Drug-Specific Imputation-Revised Age at First Use for Cigarettes**, **Cigars**, **Alcohol**, Inhalants, **Marijuana**, Hallucinogens, **Pain Relievers**, Tranquilizers, **Stimulants**, Sedatives, **Cocaine**, Crack, **Daily Cigarette Use**, and Smokeless Tobacco; **Imputation-Revised Drug-Specific 30-Day Frequency for Heroin**; **Imputation-Revised Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y)**; **Imputation-Revised Pipe Recency** (PM, **LF**); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, **Marijuana (PM, PY, LF)**, Hallucinogens (PM, PY, LF), **Pain Relievers (PM, PY, LF)**, Tranquilizers (PM, PY, LF), **Stimulants (PM, PY, LF)**, Sedatives (PM, PY, LF), **Cocaine** (PM, PY, **LF**), Crack (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF**), **Cigars** (PM, PY, P3Y, **LF**), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.44 Heroin: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Marital Status (3 Levels)** (M, **WM)**; **Education Level** (L, HS, **SC)**; **Employment Status** (FT, PT, **UN)**; **Census Region (N, M, S)**; **MSA (R, SM)**; State Rank (L, M); Cigarette Lifetime Indicator (Y); **Age * Gender**; **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)**
Recency: Past Year vs. Not Past Year	Gender (M); **Race/Hispanic Recode (4 Levels) (O, B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB**, MW); **Marital Status (3 Levels) (M, WM)**; **Education Level (L, HS, SC)**; Employment Status (FT, PT, UN); **Census Region (N**, M, **S)**; MSA (R, SM); State Rank (L, M); **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana** (PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW); **Age * Race/Hispanic Recode (4 Levels) (O**, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM, PY, LF)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.44 Heroin: Persons 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	Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, MW); **Marital Status (3 Levels) (M**, **WM)**; **Education Level (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); **Census Region (N**, M, S); **MSA (R**, **SM)**; State Rank (L, M); **Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, B, **W)**; Marital Status (3 Levels) (M, WM); **Education Level (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; **State Rank (L**, **M)**; Age * Gender; Imputation-Revised 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); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana** (PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, W); **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens** (PM, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers** (PM, **PY**, **LF)**, **Stimulants** (PM, PY, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Marital Status (3 Levels) (M, WM); Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana** (PM, PY, **LF)**, Crack (PM, PY, LF), and Tranquilizers (PM, PY, LF)

Table C.44 Heroin: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (**N, M, **S)**; **MSA (R**, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (**PM, **PY**, **LF)**, **Tranquilizers (**PM, **PY**, **LF)**, **Stimulants (**PM, PY, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Race/Hispanic Recode (2 Levels) (W); **MSA (R)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM), and Crack (PM)

Table C.44 Heroin: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Census Region (N**, M, S); **MSA** (R, **SM)**; **Imputation-Revised Pipe Recency (PM, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, **Marijuana** (PM, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM**, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco** (PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW); **Age * Race/Hispanic Recode (4 Levels) (O**, B, W); **Age Squared * Race/Hispanic Recode (4 Levels) (O**, **B**, W); Marital Status (3 Levels) (M, WM); **Education Level (**L, HS, **SC)**; **Employment Status (**FT, PT, **UN)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin; **Drug-Specific Imputation-Revised Age at First Use for** Cigarettes, Cigars, **Alcohol**, Inhalants, Marijuana, **Hallucinogens**, Pain Relievers, Tranquilizers, **Stimulants**, Sedatives, **Cocaine**, **Crack**, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Heroin; Imputation-Revised Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Marijuana (PM, PY, LF), **Hallucinogens** (PM, **PY**, **LF)**, Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), **Stimulants (PM**, **PY**, **LF)**, Sedatives (PM, PY, LF), **Cocaine (**PM, PY, **LF)**, **Crack (**PM, PY, **LF)**, and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM**, **PY)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

Table C.45 Heroin: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW); **Age * Race/Hispanic Recode (4 Levels) (O**, B, W); **Marital Status (4 Levels) (M**, W, DS); **Education Level (L, HS, SC)**; **Employment Status (FT**, PT, UN**)**; Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; Cigarette Lifetime Indicator (Y); Age * Gender; **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)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (**O, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (L, HS, SC); **Employment Status (**FT, PT, **UN)**; Census Region (N, M, S); **MSA (R, SM)**; State Rank (L, M); **Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**	Age; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels)** (O, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Marital Status (4 Levels)** (M, **W**, **DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; **Age * Gender**; **Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**

Table C.45 Heroin: Persons 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	Age Category (A1, A2); **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (L, HS, SC); **Employment Status (**FT, PT, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, SM); **State Rank (L**, M); **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (L, HS, SC); **Employment Status (**FT, PT, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	Age Category (A1, A2); **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, B, **W)**; Census Region (N, M, S); **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (**PM, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	Race/Hispanic Recode (2 Levels) (W); Marital Status (3 Levels) (M); **Intermediate Drug-Specific Past Month Indicator for Heroin (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana** (PM, PY, **LF)**, and Cocaine (PM, PY, LF)

Table C.45 Heroin: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Age Category (A1**, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users** for **Heroin**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)** for **Alcohol (PM**, **PY**, **LF)**, **Marijuana** (PM, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers** (PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)** for **Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Gender (M); Employment Status (FT); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM)

Table C.45 Heroin: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, **W)**; Census Region (N, M, S); **MSA (R**, SM); **Imputation-Revised Pipe Recency (PM, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers** (PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine** (PM, **PY**, **LF)**, **Crack** (PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM**, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, **MB**, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Age Squared * Race/Hispanic Recode (4 Levels) (O**, **B**, **W)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Age Squared * Gender**; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin; **Drug-Specific Imputation-Revised Age at First Use for Cigarettes**, **Cigars**, **Alcohol**, **Inhalants**, **Marijuana**, **Hallucinogens**, **Pain Relievers**, Tranquilizers, Stimulants, **Sedatives**, **Cocaine**, **Crack**, **Daily Cigarette Use**, and **Smokeless Tobacco**; Imputation-Revised Drug-Specific 30-Day Frequency for Heroin; Imputation-Revised Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF)**, **Marijuana** (PM, **PY**, **LF)**, **Hallucinogens** (PM, **PY**, **LF)**, **Pain Relievers** (PM, **PY**, **LF)**, Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), **Sedatives** (PM, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, Crack (PM, PY, LF), and **Inhalants** (PM, **PY**, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.46 Stimulants (Core plus Noncore): Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	**Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B**, **W)**; **Census Region (N, M, S)**; **MSA (R, SM)**; State Rank (L, M); Cigarette Lifetime Indicator (Y); **Age * Gender**; **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)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.46 Stimulants (Core plus Noncore): Persons 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	Gender (M); **Race/Hispanic Recode (4 Levels) (O, B, W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Census Region (N, M, S)**; MSA (R, SM); **State Rank (L, M)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol** (PM, **PY**, **LF**), and **Marijuana** (PM, **PY**, **LF**); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes** (PM, PY, P3Y, **LF)**
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 C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.47 Stimulants (Core plus Noncore): Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; 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)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table C.47 Stimulants (Core plus Noncore): Persons 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	Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Age * Race/Hispanic Recode (4 Levels) (O, B, W); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
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 C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table C.48 Stimulants (Core plus Noncore): Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table C.3.	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, MW); **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Marital Status (4 Levels)** (M, **W**, **DS)**; Education Level (L, HS, SC); **Employment Status** (FT, **PT**, **UN)**; Census Region (N, M, S); **MSA** (R, **SM)**; State Rank (L, M); **Cigarette Lifetime Indicator (Y)**; **Age * Gender**; **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)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (O, B, W); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised 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); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (O**, B, **W)**; **Gender * Race/Hispanic Recode (4 Levels) (MO**, MB, **MW)**; **Age * Race/Hispanic Recode (4 Levels) (O**, B, **W)**; **Marital Status (4 Levels)** (M, **W**, **DS)**; **Education Level (L, HS, SC)**; Employment Status (FT, PT, UN); **Census Region (N**, M, **S)**; **MSA** (R, **SM)**; State Rank (L, M); Age * Gender; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana** (PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**

Table C.48 Stimulants (Core plus Noncore): Persons 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	**Age Category (A1, A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **State Rank (L, M)**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Marijuana (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)**	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (O**, B, W); **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; **Age * Race/Hispanic Recode (4 Levels) (O, B, W)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level (L, HS, SC)**; **Employment Status** (FT, PT, **UN)**; **Census Region (N**, M, **S)**; **MSA** (R, **SM)**; **State Rank** (L, **M)**; **Age * Gender**; **Imputation-Revised 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)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
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 C.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

C.5 Household Composition Variables

Exhibit C.3 Definitions of Levels for Variables

Gender
M: Male, F: Female[1]
Race/Hispanic Recode (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]
Gender * Race/Hispanic Recode (4 Levels)
MH: Male Hispanic/Latino, MO: Male Not Hispanic/Latino and Other, MB: Male Not Hispanic/Latino and Black/African American Only, MW: Male Not Hispanic/Latino and White Only,[1] FH: Female Hispanic/Latino,[1] FO: Female Not Hispanic/Latino and Other,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FW: Female Not Hispanic/Latino and White Only[1]
Age * Race/Hispanic Recode (4 Levels)
H: Age * Hispanic/Latino, O: Age * Not Hispanic/Latino and Other, B: Age * Not Hispanic/Latino and Black/African American Only, W: Age * Not Hispanic/Latino and White Only[1]
Age Squared * Race/Hispanic Recode (4 Levels)
H: Age Squared * Hispanic/Latino, O: Age Squared * Not Hispanic/Latino and Other, B: Age Squared * Not Hispanic/Latino and Black/African American Only, W: Age Squared * Not Hispanic/Latino and White Only[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]
Education Level
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]
Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]
Percentage Hispanic/Latino in Segment
H: ≥ 70%, M: [20,70)%, L: < 20%[1]
Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]
Imputation-Revised Family in Household
Y: Other Family Members in Household,[1] N: No Other Family Members in Household

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 C.49 Household Composition: Persons 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**; Gender (M); **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **Percentage Hispanic/Latino in Segment (H, M)**; **Percentage Owner Occupied in Segment** (H, **M)**; Total People in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total People in Household (Screener)
Number of Persons Younger than 18 in Household (KID17)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)
Number of Persons Older than 64 in Household (HH65)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household

Table C.49 Household Composition: Persons 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; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Census Region (N, M, S); MSA (R, SM); **Percentage Hispanic/Latino in Segment (H, M)**; **Percentage Owner Occupied in Segment (H, M)**; **Imputation-Revised Household Size**; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household

Table C.49 Household Composition: Persons 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 Younger than 18 including Foster Relationships (KIDFMLY)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); **Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)**	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)

Table C.49 Household Composition: Persons 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 Younger than 18 excluding Foster Relationships (KIDFAMSZ)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; **Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.3 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.50 Household Composition: Persons 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**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (**MH, **MO**, **MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (3 Levels) (M**, **WM)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (**H, **M)**; Percentage Owner Occupied in Segment (H, M); Total People in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total People in Household (Screener)
Number of Persons Younger than 18 in Household (KID17)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)
Number of Persons Older than 64 in Household (HH65)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household

Table C.50 Household Composition: Persons 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; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (3 Levels) (M, WM)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); **Percentage Owner Occupied in Segment (H, M)**; **Imputation-Revised Household Size**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; Age * Gender; Age Squared * Gender; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (3 Levels) (M, WM)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R, SM)**; **Percentage Hispanic/Latino in Segment (H, M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household

Table C.50 Household Composition: Persons 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; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (3 Levels)** (M, **WM)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R, SM)**; **Percentage Hispanic/Latino in Segment** (H, **M)**; Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**; **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)

Table C.50 Household Composition: Persons 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; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); **Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.3 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.51 Household Composition: Persons 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; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (**MH, MO, **MB)**; Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total People in Household (Screener)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total People in Household (Screener)
Number of Persons Younger than 18 in Household (KID17)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)
Number of Persons Older than 64 in Household (HH65)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household

Table C.51 Household Composition: Persons 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; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M, W, DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household

Table C.51 Household Composition: Persons 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; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, **MB)**; Age * Gender; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, **MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, **MB)**; Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); **Census Region (**N, **M**, **S)**; MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)

Table C.51 Household Composition: Persons 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; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**; **Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships)**; **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; **Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.3 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.52 Household Composition: Persons 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**; Gender (M); **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels) (**M, **W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (**N, **M, S)**; **MSA (R, SM)**; **Percentage Hispanic/Latino in Segment (H, M)**; **Percentage Owner Occupied in Segment** (H, **M)**; Total People in Household (Screener)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; Age Squared * Gender; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels) (M, W, DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total People in Household (Screener)
Number of Persons Younger than 18 in Household (KID17)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels) (**M, **W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (**N, **M, S)**; MSA (R, SM); **Percentage Hispanic/Latino in Segment (H, M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Number of Eligible 12 to 17 in Household (Screener)**	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)
Number of Persons Older than 64 in Household (HH65)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels) (**M, **W, DS)**; **Education Level (**L, **HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; MSA (R, SM); **Percentage Hispanic/Latino in Segment (H, M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household

Table C.52 Household Composition: Persons 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**; Gender (M); **Race/Hispanic Recode (4 Levels) (H**, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, **MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H**, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, **MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; MSA (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**

Table C.52 Household Composition: Persons 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**; Gender (M); **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels)** (**M, W, DS)**; **Education Level** (L, **HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (**N, **M**, **S)**; MSA (R, SM); **Percentage Hispanic/Latino in Segment (H, M)**; **Percentage Owner Occupied in Segment** (H, **M)**; **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; Age * Gender; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels)** (M, W, **DS)**; **Education Level** (L, **HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; MSA (R, SM); **Percentage Hispanic/Latino in Segment (H, M)**; Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); **Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)**	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels)** (M, W, **DS)**; Education Level (L, HS, SC); **Employment Status (**FT, PT, **UN)**; Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)

Table C.52 Household Composition: Persons 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**; Gender (M); **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (4 Levels)** (M, **W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; MSA (R, SM); **Percentage Hispanic/Latino in Segment (H, M)**; **Percentage Owner Occupied in Segment** (H, **M)**; **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household**; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; **Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)**	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; Imputation-Revised Number of Respondent's Family Members Younger Than 18 in Household (Includes Foster Relationships)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.3 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

C.6 Income Variables

Exhibit C.4 Definitions of Levels for Variables

Gender
M: Male, F: Female[1]

Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only, B: Not Hispanic/Latino and Black/African American Only, H: Hispanic/Latino, O: Not Hispanic/Latino and Other[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, MH: Male Hispanic/Latino, MO: Male Not Hispanic/Latino and Other,[1] FW: Female Not Hispanic/Latino and White Only,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FH: Female Hispanic/Latino,[1] FO: Female Not Hispanic/Latino and Other[1]

Age * Race/Hispanic Recode (4 Levels)
W: Age * Not Hispanic/Latino and White Only, B: Age * Not Hispanic/Latino and Black/African American Only, H: Age * Hispanic/Latino, O: Age * Not Hispanic/Latino and Other[1]

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, H: Age Squared * Hispanic/Latino, O: Age Squared * Not Hispanic/Latino and Other[1]

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, H: Age Cubed * Hispanic/Latino, O: Age Cubed * Not Hispanic/Latino and Other[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]

Education Level
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]

Census Region
N: Northeast, M: Midwest, S: South, W: West[1]

MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[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]

Percentage Hispanic/Latino in Segment
H: ≥ 70%, M: [20,70)%, L: < 20%[1]

Percentage Non-Hispanic/Latino Black/African American in Segment
H: ≥ 70%, M: [20,70)%, L: < 20%[1]

Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]

Imputation-Revised Family Received Social Security or Railroad Payments
Y: Yes, N: No[1]

Imputation-Revised Family Received Supplemental Security Income
Y: Yes, N: No[1]

Imputation-Revised Family Received Public Assistance
Y: Yes, N: No[1]

Imputation-Revised Family Received Welfare/Job Placement/Childcare Services
Y: Yes, N: No[1]

Exhibit C.4 Definitions of Levels for Variables (continued)

Imputation-Revised Family Received Income from a Job
Y: Yes, N: No[1]
Imputation-Revised Family Received Food Stamps
Y: Yes, N: No[1]
Imputation-Revised Total Family Income > or < $20,000
GT: ≥ $20,000, LT: < $20,000[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 Public Assistance
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]
Intermediate Family Received Food Stamps
Y: Yes, N: No[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 C.53 Dichotomous Income Indicators in Response Propensity Models

Age Group	Variables Included in Response Propensity (Dichotomous Income Indicators)
12-17	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age Cubed * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)
18-25	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age Cubed * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)**
26-64	Age; Age Squared; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; **Age Squared * Gender**; **Age Cubed * Gender**; Age * Race/Hispanic Recode (4 Levels) (W, B, H); **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W, B, H)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Number of Adults in Household**; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)
65+	Age; Age Squared; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; **Age Cubed * Gender**; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); **Age Cubed * Race/Hispanic Recode (4 Levels) (W, B, H)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); **Employment Status (**FT, PT, **UN)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.54 Dichotomous Income Indicators in Predictive Mean Models: Persons 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)
Supplemental Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y)
Wages	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y)
Food Stamps	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y)

Table C.54 Dichotomous Income Indicators in Predictive Mean Models: Persons 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y)
Welfare Services	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y)
# Welfare Months	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

Table C.54 Dichotomous Income Indicators in Predictive Mean Models: Persons 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.55 Dichotomous Income Indicators in Predictive Mean Models: Persons 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)**
Supplemental Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y)
Wages	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y)
Food Stamps	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y)

Table C.55 Dichotomous Income Indicators in Predictive Mean Models: Persons 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y)
Welfare Services	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y)
# Welfare Months	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

Table C.55 Dichotomous Income Indicators in Predictive Mean Models: Persons 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.56 Dichotomous Income Indicators in Predictive Mean Models: Persons 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R, SM)**; Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Number of Adults in Household**; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)**
Supplemental Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y)
Wages	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y)
Food Stamps	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y)

Table C.56 Dichotomous Income Indicators in Predictive Mean Models: Persons 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y)
Welfare Services	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y)
# Welfare Months	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

Table C.56 Dichotomous Income Indicators in Predictive Mean Models: Persons 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.57 Dichotomous Income Indicators in Predictive Mean Models: Persons 65 Years or Older

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)
Supplemental Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); **Employment Status (**FT, PT, **UN)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y)
Wages	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); **Employment Status (**FT, **PT**, **UN)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y)
Food Stamps	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); **Intermediate Family Received Supplemental Security Income (Y)**; **Intermediate Family Received Public Assistance (Y)**; Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y)

Table C.57 Dichotomous Income Indicators in Predictive Mean Models: Persons 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); **Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; Census Region (N, M, S); MSA (R, SM); **Income State Rank (L, M)**; **Percentage Hispanic/Latino in Segment (H, M)**; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H, M)**; Imputation-Revised Number of Adults in Household; **Imputation-Revised Number of Persons Younger Than 18 Years Old in Household**; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)**; Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y)
Welfare Services	Age; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (W, B**, H); **Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)**; Age * Gender; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **Income State Rank (L, M)**; Percentage Hispanic/Latino in Segment (H, M); **Percentage Non-Hispanic/Latino Black/African American in Segment (H, M)**; **Percentage Owner Occupied in Segment (H, M)**; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)**; Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); **Intermediate Family Received Public Assistance (Y)**
# Welfare Months	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; **Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Marital Status (4 Levels)** (M, **W, DS)**; **Education Level (L, HS, SC)**; **Employment Status (FT, PT, UN)**; **Census Region (**N, **M, S)**; **MSA** (R, **SM)**; **Income State Rank (L, M)**; **Percentage Hispanic/Latino in Segment (H, M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment** (H, **M)**; **Percentage Owner Occupied in Segment (H, M)**; **Imputation-Revised Number of Adults in Household**; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; **Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous)**; Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); **Intermediate Family Received Public Assistance (Y)**; Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); **Intermediate Family Received Income from a Job (Y)**; Intermediate Family Received Food Stamps (Y)

Table C.57 Dichotomous Income Indicators in Predictive Mean Models: Persons 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.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; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)
18-25	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)
26-64	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)

Table C.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; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.59 Income Finer Categories in Predictive Mean Models

Age Group	Variables Included in Income Models (Finer Categorization)
12-17	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)
18-25	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)
26-64	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)

Table C.59 Income Finer Categories in Predictive Mean Models (continued)

Age Group	Variables Included in Income Models (Finer Categorization)
65+	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

C.7 Health Insurance Variables

Exhibit C.5 Definitions of Levels for Variables

Gender
M: Male, F: Female[1]
Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only, B: Not Hispanic/Latino and Black/African American Only, H: Hispanic/Latino, O: Not Hispanic/Latino and Other[1]
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, H: Age * Hispanic/Latino, O: Age * Not Hispanic/Latino and Other[1]
Age Squared * Race/Hispanic Recode (4 Levels)[2]
W: Age Squared * Not Hispanic/Latino and White Only,[1] B: Age Squared * Not Hispanic/Latino and Black/African American Only, H: Age Squared * Hispanic/Latino, O: Age Squared * Not Hispanic/Latino and Other[1]
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, H: Age Cubed * Hispanic/Latino, O: Age Cubed * Not Hispanic/Latino and Other[1]
Gender * Race/Hispanic Recode (4 Levels)[2]
MH: Male Hispanic/Latino,[1] MO: Male Not Hispanic/Latino and Other, MB: Male Not Hispanic/Latino and Black/African American Only, MW: Male Not Hispanic/Latino and White Only,[1] FH: Female Hispanic/Latino,[1] FO: Female Not Hispanic/Latino and Other,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FW: Female Not Hispanic/Latino and White Only[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]
Education Level
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]
MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[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]
Percentage Hispanic/Latino in Segment
H: ≥ 70%, M: [20,70)%, L: < 20%[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]
Household Size
L: At Least Four People in Household, S: Fewer than Four People in Household[1]
Other Family Members In Household
N: No, Y: Yes[1]
Lifetime Military Service
N: No,[1] Y: Yes
Family Wages
N: No,[1] Y: Yes
Family Participation in Government Assistance Programs
N: No,[1] Y: Yes
Family Social Security
N: No,[1] Y: Yes

Exhibit C.5 Definitions of Levels for Variables (continued)

Intermediate MEDICAID/CHIP Coverage
N: No,[1] Y: Yes
Intermediate MEDICARE Coverage
N: No,[1] Y: Yes
Intermediate CHAMPUS Coverage
N: No,[1] Y: Yes

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] Under the old method, the reference cell for the race/Hispanic recode was "Not Hispanic/Latino and Other." Under the constituent variables method, the reference cell for the race/Hispanic recode was "Not Hispanic/Latino and White Only." The latter choice 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. See Chapter 9 for details on the processing of the health insurance variables, as well as a fuller description of the differences between the old and constituent variables methods applied to them.

Table C.60 Health Insurance, Constituent Variables Method: 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; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O)**; Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
	Other Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
18-25	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (2 Levels) (M); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
	Other Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (2 Levels) (M); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)

Table C.60 Health Insurance, Constituent Variables Method: 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; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
	Other Health Insurance[1]	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); **MSA (R, SM)**; Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
65+	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, **O)**; Marital Status (2 Levels) (M); Education Level (L, HS, SC); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M)
	Other Health Insurance[1]	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O)**; **Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW)**; Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); **MSA (R, SM)**; Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] The 26-64 and 65+ age groups were included in the same response propensity model for other health insurance.

Table C.61 Health Insurance, Constituent Variables Method: Predictive Mean Models, Persons 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)
Medicare	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); MSA (R, SM); Percentage Owner Occupied in Segment (H, M); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y)
CHAMPUS	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y)
Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Intermediate CHAMPUS Coverage (Y)
Other Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.62 Health Insurance, Constituent Variables Method: Predictive Mean Models, Persons 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (2 Levels) (M); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)
Medicare[1]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Percentage Owner Occupied in Segment (H, M); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y)
CHAMPUS	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (2 Levels) (M); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); **Lifetime Military Service (Y)**; Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y)
Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (2 Levels) (M); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Intermediate CHAMPUS Coverage (Y)
Other Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (2 Levels) (M); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] The 18-25 and 26-64 age groups were included in the same predictive mean model for Medicare.

Table C.63 Health Insurance, Constituent Variables Method: Predictive Mean Models, Persons 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); **Family Participation in Government Assistance Programs (Y)**; Family Social Security (Y)
Medicare[1]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Percentage Owner Occupied in Segment (H, M); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y)
CHAMPUS	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Lifetime Military Service (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y)
Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Intermediate CHAMPUS Coverage (Y)
Other Health Insurance[2]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[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 C.64 Health Insurance, Constituent Variables Method: Predictive Mean Models, Persons 65 Years or Older

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (L, HS, SC); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Household Size (L); Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)
Medicare	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (L, HS, SC); MSA (R, SM); Percentage Owner Occupied in Segment (H, M); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y)
CHAMPUS	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (L, HS, SC); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Lifetime Military Service (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y)
Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (L, HS, SC); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Household Size (L); Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Intermediate CHAMPUS Coverage (Y)
Other Health Insurance[1]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MO, MB, MW); Marital Status (4 Levels) (M, W, DS); Education Level (L, HS, SC); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] The 26-64 and 65+ age groups were included in the same predictive mean model for other health insurance.

Table C.65 Old Method Health Insurance: Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12-17	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
18-25	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
26-64	Age; Age Squared; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **MSA (R, SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
65+	**Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (**W, **B**, H**)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Gender * Race/Hispanic Recode (4 Levels) (**MH, **MO**, **MB)**; MSA (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.66 Old Method Health Insurance: Predictive Mean Models, Persons 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
Private Health Insurance	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.67 Old Method Health Insurance: Predictive Mean Models, Persons 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
Private Health Insurance	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.68 Old Method Health Insurance: Predictive Mean Models, Persons 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
Private Health Insurance	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table C.69 Old Method Health Insurance: Predictive Mean Models, Persons 65 Years or Older

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (**W, B, **H)**; Age * Gender; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (**W, B, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Gender * Race/Hispanic Recode (4 Levels) (**MH, MO, **MB)**; MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
Private Health Insurance	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.5 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

C.8 Roster Pair Variables

Exhibit C.6 Definitions of Levels for 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 (older) - 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]
Gender
M: Male,[1] F: Female
Gender (older)
M: Male,[1] F: Female
Gender (younger)
M: Male,[1] F: Female
Race/Hispanic Recode (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
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]
Marital Status (older) (2 Levels)
MWD: Married, Widowed or Divorced, NM: Never Been Married[1]
Marital Status (older) (2 Levels) - Non-Pair
MWD: Married, Widowed or Divorced, NM: Never Been Married[1]
Marital Status (older) (4 Levels)
M: Married, W: Widowed, D: Divorced, NM: Never Been Married[1]
Marital Status (younger)
M: Married, W: Widowed, D: Divorced, NM: Never Been Married[1]
Education Level
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Education Level (older)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Education Level (younger)
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]
Employment Status (older)
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Employment Status (younger)
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
Population Density
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA
Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,7[1])%, L: < 20%
Percentage Non-Hispanic/Latino Black/African American in Segment
H: ≥ 40%, M: [10,40)%, L: < 10%[1]
Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10,50)%, L: < 10%[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 C.70 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)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; **Gender (older) (F)**; Gender (younger) (F); **Race/Hispanic Recode (4 Levels) (H**, O, B**)**; **Census Region (N**, M, S**)**; **Population Density (**R, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Census Region (N**, M, S**)**; Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
1 (12-14, 15-17)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); **Race/Hispanic Recode (4 Levels) (**H, O, **B)**; Census Region (N, M, S); **Population Density (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); **Percentage Non-Hispanic/Latino Black/African American in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H**, **M)**	Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H**, **M)**; 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+

Table C.70 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)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (**M, **W**, **D)**; Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (**M, **W**, **D)**; Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
3 (15-17, 15-17)	Household Size (Continuous); **Household Size (5 Categories)**; **Gender (older) (F)**; **Gender (younger) (F)**; **Race/Hispanic Recode (4 Levels) (**H, O, **B)**; **Census Region (**N, **M**, **S)**; **Population Density (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.70 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)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (M**, **W**, **D)**; Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (M**, **W**, **D)**; Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
5 (18-20, 18-25)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels)** (M, **W**, D); **Marital Status (younger)** (M, **W**, D); Education Level (older) (L, HS, SC); Education Level (younger) (L, HS, SC); Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3); Gender (older) (F); **Gender (younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels)** (M, **W**, D); **Marital Status (younger) (M**, **W**, **D)**; Education Level (older) (L, HS, SC); **Education Level (younger)** (L, HS, **SC)**; Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T3); Gender (older) (F); **Gender (younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (M**, **W**, D); **Marital Status (younger) (M**, **W**, **D)**; **Education Level (older) (L**, HS, SC); **Education Level (younger)** (L, HS, **SC)**; **Employment Status (older) (FT**, PT, UN); **Employment Status (younger)** (FT, PT, **UN)**; Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **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+**

Table C.70 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)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); **Marital Status (younger) (**M, **W**, D**)**; Education Level (older) (L, HS, SC); Education Level (younger) (L, HS, SC); Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); **Gender (younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (**M, **W**, **D)**; **Marital Status (younger) (**M, **W**, D); **Education Level (older) (**L, **HS**, **SC)**; Education Level (younger) (L, HS, SC); Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Gender (older) (F)**; **Gender (younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (M**, **W**, **D)**; **Marital Status (younger) (M**, **W**, **D)**; **Education Level (older) (**L, **HS**, **SC)**; Education Level (younger) (L, HS, SC); Employment Status (older) (FT, PT, UN); **Employment Status (younger) (FT**, **PT**, **UN)**; Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **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+**
7 (12-14, 26+)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); **Race/Hispanic Recode (4 Levels) (H**, **O**, B**)**; **Census Region (N**, **M**, S**)**; **Population Density (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.70 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
8 (15-17, 26+)	Household Size (Continuous); **Household Size (5 Categories)**; **Age Category (older) (**A2, **A3)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); **Education Level (older) (L**, **HS**, **SC)**; Employment Status (older) (FT, PT, UN); **Census Region (**N, M, **S)**; **Population Density (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (M**, **W**, **D)**; Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.70 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+)	**Household Size (Continuous)**; Household Size (5 Categories); **Age Category (older) (A2, A3)**; Gender (older) (F); **Gender (younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (M, W, D)**; **Marital Status (younger) (**M, **W**, **D)**; Education Level (older) (L, HS, SC); **Education Level (younger) (L**, **HS**, **SC)**; Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H, M)**	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); **Marital Status (younger) (**M, **W**, **D)**; Education Level (older) (L, HS, SC); **Education Level (younger) (L**, **HS**, **SC)**; Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); **Marital Status (younger) (**M, **W**, **D)**; Education Level (older) (L, HS, SC); **Education Level (younger) (L**, **HS**, **SC)**; Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.70 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
10 (21+, 26+)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (4 Levels) (M, W, D); Marital Status (younger) (M, W, D); Education Level (older) (L, HS, SC); Education Level (younger) (L, HS, SC); Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); **Age Category (older)** (A2, **A3**); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (older) (4 Levels) (M, W, D)**; **Marital Status (younger) (M, W, D)**; Education Level (older) (L, HS, SC); Education Level (younger) (L, HS, SC); Employment Status (older) (FT, PT, UN); Employment Status (younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Age Category (older) (A2, A3)**; **Gender (older) (F)**; **Gender (younger) (F)**; **Race/Hispanic Recode (4 Levels) (H, O, B)**; **Marital Status (older) (4 Levels) (M, W, D)**; **Marital Status (younger) (M, W, D)**; **Education Level (older)** (L, **HS**, **SC)**; Education Level (younger) (L, HS, SC); **Employment Status (older) (FT, PT, UN)**; **Employment Status (younger) (FT, PT, UN)**; **Census Region** (N, **M**, S); **Population Density** (R, **SM)**; **Percentage Hispanic/Latino in Segment (H, M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (H, M)**; **Percentage Owner Occupied in Segment (H, M)**; **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+**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.6 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table C.71 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	Household Size (Continuous); **Household Size (5 Categories)**; **Age Category (older) (**T3, A1, A2, **A3)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; **Age Category (older) (T2**, T3, A1, A2**)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Age Category (older) (T2**, T3, A1, **A2)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**

Table C.71 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; **Age Category (older) (T2**, T3, A1, A2**)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Age Category (older) (T2**, T3, A1, A2**)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
Sibling (12-14) Sibling (15-17) Older Sibling Focus	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.71 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	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
Sibling (12-17) Sibling (18-25) Older Sibling Focus	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T2); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.71 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T2); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T2); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T2); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.6 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table C.72 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
Parent-Child (12-20) Child Focus, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.72 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (Y); Gender (older) (F); **Gender (younger) (F)**; **Race/Hispanic Recode (4 Levels)** (H, **O**, B); **Census Region (N**, M, **S)**; Population Density (R, SM); **Percentage Hispanic/Latino in Segment (H, M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (**H, **M)**; Percentage Owner Occupied in Segment (H, M)	Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**
Parent-Child (12-20) Parent Focus, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**

Table C.72 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	**Household Size (Continuous)**; Household Size (5 Categories); Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); **Race/Hispanic Recode (4 Levels) (**H, O, **B)**; **Census Region (N**, M, **S)**; Population Density (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (**H, **M)**	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
Sibling (12-14) Sibling (15-17), Older Sibling Focus, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**

Table C.72 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
Sibling (12-17) Sibling (18-25), Older Sibling Focus, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.72 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category (older) (Y); **Gender (older) (F)**; Gender (younger) (F); **Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Census Region (N, M, S); **Population Density (**R, **SM)**; **Percentage Hispanic/Latino in Segment (**H, **M)**; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H**, **M)**	Age Category (older) (Y); Gender (older) (F); Gender (younger) (F); **Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Census Region (N, M, S); **Population Density (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (H**, **M)**; 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+**
Spouse-Spouse, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**

Table C.72 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T1, T2, T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); Age Category (older) (T1, T2, T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T1, T2, T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **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+**

Table C.72 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) (T1, T2, T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); Age Category (older) (T1, T2, T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) (T1, T2, T3, A1, A2, A3); Gender (older) (F); Gender (younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (older) (2 Levels) (MWD); Education Level (older) (L, HS, SC); Employment Status (older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **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+**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.6 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table C.73 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
Parent-Child (12-20) Child Focus, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.73 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**
Parent-Child (12-20) Parent Focus, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; **Age Category (older) - Non-Pair (T3**, **A1**, **A2**, **A3)**; Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**

Table C.73 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	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); **Marital Status (older) (2 Levels) - Non-Pair (MWD)**; **Census Region** (N, **M**, S); **Population Density (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Non-Hispanic/Latino Black/African American in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.73 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	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+

Table C.73 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, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**
Spouse-Spouse, Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**

Table C.73 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, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); **Marital Status (older) (2 Levels) - Non-Pair (MWD)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); **Marital Status (older) (2 Levels) - Non-Pair (MWD)**; Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); 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+**
Spouse-Spouse with Children	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (older) - Non-Pair (T1, T2, T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); **Age Category (older) - Non-Pair (**T1, T2, T3, A1, A2, **A3)**; Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (older) - Non-Pair (T1, T2, T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (older) (2 Levels) - Non-Pair (MWD); Education Level (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **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+**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit C.6 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Appendix D: Hot-Deck Procedure Summaries

Appendix D: Hot-Deck Procedure Summaries

D.1 Introduction

For the majority of variables that had missing values imputed in the 2011 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 2 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.

D.2 Demographics

Tables D.1 through D.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 D.12 through D.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 3 for details). The segment-level variable, SEGID (Segment ID), was used only in the likeness constraints for demographic imputation.

As described in the 2011 NSDUH sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2012) 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 3 and 4. For more information regarding segments, see the 2011 NSDUH sample design report (Morton et al., 2012).

D.2.1 Marital Status Variables

Table D.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 D.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	14	None	1. M1 2. M2 3. M3	1,2: 2	1,2: 4 1: 1	1,2: 6 1: 1

[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)

D.2.2 Race Variables

Table D.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 D.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[1]
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[1]
LikC5	Donor must be Mexican (Hispanic/Latino or non-Hispanic/Latino)[1]
LikC6	Donor must be Cuban (Hispanic/Latino or non-Hispanic/Latino)[1]
LikC7	Donor must be Central or South American (Hispanic/Latino or non-Hispanic/Latino)[1]
LikC8	Donor must be Dominican (Hispanic/Latino or non-Hispanic/Latino)[1]
LikC9	Donor must be Spanish (Hispanic/Latino or non-Hispanic/Latino)[1]

[1] These likeness constraints are never loosened.

Table D.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	2,254	None	1. R1 2. R2 3. R3 4. R4	1-4: 193 2-4: 426 3,4: 163	1-4: 290 2-4: 421 3,4: 160	1-4: 211 2-4: 344 3,4: 46
2	Known to be Asian	10	1	1. R1 2. R2 3. R3 4. R4	1-4: 0 2-4: 2	1-4: 1 2-4: 1 3,4: 3	1-4: 0 2-4: 2 3,4: 0 3,4: 1
3	Known to be multiple race, but no other information	6	2	1. R1 2. R2 3. R3 4. R4	No Cases	1-4: 0 2-4: 3	1-4: 0 2-4: 2 3,4: 1
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: 0 3,4: 3	1-4: 0 2-4: 0 3,4: 2	1-4: 1 2-4: 1 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	1-4: 0 2-4: 1	No Cases
6	Known not to be American Indian/ Alaska Native, in part or in full	0	5	1. R1/(1-R3) 2. R2/(1-R3) 3. R4/(1-R3)	No Cases	No Cases	No Cases
7	Known to be non-Hispanic Mexican	4	None	1. R1 2. R2 3. R3 4. R4	1-5: 0 2-5: 1	1-5: 0 2-5: 1	1-5: 1 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 D.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

[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)

D.2.3 Hispanic/Latino Origin Variables

Table D.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 D.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	93	None	1. H1	1,2: 25 2: 57	1,2: 1 2: 3	1,2: 3 2: 4

[1] The predictive mean vector components are defined by the following:
1. H1 = P(Hispanic/Latino origin)

D.2.4 Hispanic/Latino Group Variables

Table D.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 IRDETAILEDRACE value must indicate a subset of the racial categories mentioned by recipient[1]

[1] This likeness constraint is never loosened.

Table D.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	54	None	1. H1 2. H2 3. H3	1-3: 32 2,3: 21 3: 1

[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.

D.2.5 Education Variables

Table D.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 D.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	5	None	1. E1 2. E2 3. E3 4. E4	1. E1 2. E2 3. E3	1-3: 0 2,3: 1	1-3: 0 2,3: 2	1,2: 0 2: 1 None: 1

[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)

D.2.6 Employment Variables

Table D.12 Logical Constraints for Employment Status

Constraint #	Logical Constraint
LogC1	Donor must be employed

Table D.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 D.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	33	None	1. E1 2. E2 3. E3	1-3: 1 2,3: 4 3: 1	1-3: 1 2,3: 9	1-3: 0 2,3: 10 3: 7
2	Known to be employed; part-time vs. full-time status unknown	15	1	1. E1/ (E1+E2)	No Cases	1-3: 0 2,3: 9	1-3: 0 2,3: 6

[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)

D.2.7 Immigrant Variables

Table D.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 D.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	39	None	1. B1	1,2: 4 2: 9	1,2: 2 2: 9	1,2: 2 2: 13

[1] The predictive mean vector components are defined by the following:
1. B1 = P(Born in the United States)

Table D.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 D.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 D.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	36	1,2	1. PrAgeEntry	1,2: 6 2: 29 None: 0 None*: 1

*A donor could not be found who met all the logical constraints after all the likeness constraints had been loosened. Therefore, age of entry was assigned to the most likely value based on the respondent's date of birth and interview date.

[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.

D.3 Household Composition (Roster) Variables

Tables D.20 through D.41 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 persons, number of children younger than 18, number of persons 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 D.20 Likeness Constraints for Number of Rostered Persons

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means

Table D.21 Constraints and Portion of the Predictive Mean Vector for Number of Rostered Persons

#	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	37	None	1. C1	1: 7	1: 12	1: 16 None: 1	1: 1

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table D.22 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 D.23 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 D.24 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	212	1	1. C1	1,2: 63 2: 1	1,2: 85 2: 2 None: 1	1,2: 56 2: 1	1,2: 3

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table D.25 Logical Constraints for Number of Persons 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 D.26 Likeness Constraints for Number of Persons 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 D.27 Constraints and Portion of the Predictive Mean Vector for Number of Persons 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	382	1	1. C1	1,2: 188	1,2: 142 2: 1	1,2: 47 2: 1	1,2: 3

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table D.28 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 D.29 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	45	None	1. F1	1-4: 7	1-4: 15	1-4: 21	1-4: 2

[1] The predictive mean vector components are defined by the following:
1. F1 = P(No other family members in the household)

Table D.30 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 D.31 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 D.32 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	64	None	1. C1	1-3: 15	1-3: 21	1-3: 24 2,3: 1	1-3: 3

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table D.33 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 D.34 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 D.35 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	83	None	1. C1	1-4: 38 2-4: 2	1-4: 21	1-4: 20 2-4: 1 3,4: 0 4: 1	No Cases

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table D.36 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 D.37 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 D.38 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	60	None	1. C1	1-4: 14	1-4: 20	1-4: 21 2-4: 2	1-4: 3

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table D.39 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 D.40 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 D.41 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	72	None	1. C1	1-6: 32 2-6: 2	1-6: 17 2-6: 1	1-6: 15 2-6: 5	No Cases

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

D.4 Drug Variables

Tables D.42 through D.97 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.

D.4.1 Lifetime Drug Use

There were a large number of missingness patterns for lifetime drug use. There were 15 "gate questions" in the questionnaire that corresponded to lifetime use variables, plus several "subgate" questions (multiple gate questions). To be considered a completed case for purposes of analysis, a respondent had to provide "yes" or "no" answers to the cigarette gate question and at least 9 of the other 14 gate questions. Aside from these restrictions, any combination of the lifetime drug variables may have been missing.

The probabilities associated with the 14 gate questions, other than cigarettes (Table D.48), formed the full predictive mean vector. Only the probabilities associated with the gate questions for which the responses 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, as explained in the preceding paragraph.

Note that if only a subgate question was missing, the predicted mean associated with the gate question was used. No predicted means were calculated for subgate questions. 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.[3]

Table D.42 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)

[3] Crack is an exception to this rule. The question on crack lifetime use would be considered a subgate question except that crack is 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 D.42 Elements of Full Predictive Mean Vector for Lifetime Drug Use (continued)

Lifetime Drug Use	Predicted Mean
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 D.43 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 D.44 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 D.45 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	936	1	1. LF for gate question associated with each missing variable	1-3: 524 1,2,4: 70 1,4: 40	1-3: 97 1,2,4: 52 1,4: 14	1-3: 71 1,2,4: 51 1,4: 17

[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 D.46 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	If recipient was a past year cigarette ever daily user, then donor was a past year cigarette ever daily user; if recipient was a past 3 years but not past year cigarette ever daily user, then donor was the same; if recipient was a lifetime but not past 3 years cigarette ever daily user, then donor was the same
LikC4	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC5	Age of donor must be less than or equal to age of recipient

Table D.47 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	17	None	1. CIG	1-4: 8 1,3,4: 1	1-4: 2	1-4: 4 1,3,4: 1 1,3: 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)

D.4.2 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 D.48 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 D.49 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 D.50 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	10	4,6	1. R1/(R1+R2) 2. (R1*D)/(R1+R2) 3. R1*(1-D)*PM/(R1+R2)	1,2: 4 1: 4	1,2: 2	No Cases
2	Recency and 30-day frequency missing (lifetime use known)	9	6	1. R1 2. R2 3. R3 4. R1*D 5. R1*(1-D)*PM	1,2: 0 1: 6	1,2: 2	1,2: 0 1: 1
3	Past month, 30-day frequency missing	24	5-7	1. D 2. PM	1,2: 7 1: 6	1,2: 6	1,2: 4 1: 1
4	Not past year	198	3,6	1. R3/(R3+R4)	1,2: 111 1: 1	1,2: 84	1,2: 2
5	Not past month	92	2,6	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1,2: 59 1: 3	1,2: 19 1: 2	1,2: 7 1: 2
6	Past year but not past month, or past 3 years but not past year	158	6,8	1. R2/(R2+R3)	1,2: 107	1,2: 49	1,2: 2
7	Past 3 years, 30-day frequency missing	2	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: 2	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 D.51 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	If recipient was never a daily user of cigarettes (CIGDLYMO = 2), donor's 30-day cigarette frequency cannot equal 30
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 D.52 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 D.53 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	6	3	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	1,2: 5	1,2: 1	No Cases
2	Recency and 30-day frequency missing (lifetime use known)	7	None	1. R1 2. R2 3. R3 4. R1*PM	1,2: 0 1: 2	1,2: 0 1: 4	1,2: 1 1: 1
2	Recency and 30-day frequency missing (lifetime use imputed)	1					
3	Past month, 30-day frequency missing	12	4,6	1. PM	1,2: 6	1,2: 3	1,2: 3
4	Not past year	133	3	1. R3/(R3+R4)	1,2: 52	1,2: 68	1,2: 13
5	Not past month	77	2	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1,2: 39 1: 3	1,2: 30	1,2: 5
6	Past year but not past month, or past 3 years but not past year	119	7	1. R2/(R2+R3)	1,2: 65 1: 2	1,2: 49	1,2: 3
7	Past 3 years, 30-day frequency missing	1	1	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. (R1*PM)/(R1+R2+R3)	No Cases	1,2: 1	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 D.54 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 D.55 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 D.56 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		3	5,7	1. DC 2. PMC	1-3: 1 1,2,4: 0 1,4: 2	No Cases	No Cases
3		(Past month)		Missing	3	6,8	1. DS 2. PMS	1-3: 0 1,2,4: 0 1,4: 3	No Cases	No Cases
4		Missing (lifetime use known)		Missing	5	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: 2	1-3: 0 1,2,4: 0 1,4: 4	1-3: 0 1,2,4: 0 1,4: 1
4		Missing (lifetime use imputed)		Missing	2					
5	(Past month)	Missing (lifetime use known)	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
5	(Past month)	Missing (lifetime use imputed)	Missing	Missing	0					
6	Missing (lifetime use known)		Missing		4	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: 3	1-3: 0 1,2,4: 0 1,4: 1	No Cases
6	Missing (lifetime use imputed)		Missing		0					

Table D.56 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+
7	Missing (lifetime use known)	(Past month)	Missing	Missing	0	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	No Cases
7	Missing (lifetime use imputed)	(Past month)	Missing	Missing	0					
8		Past year		Missing	5	6,12	1. R1/(R1+R2) 2. RS1*DS/ (RS1+RS2) 3. RS1*(1-DS)*PMS/ (RS1+RS2)	1-3: 0 1,2,4: 1 1,4: 3	1-3: 0 1,2,4: 0 1,4: 1	No Cases
9	Past year		Missing		2	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	No Cases	No Cases
10	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	3-6	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. RS1*DS 7. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases
10	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
10	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
10	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.56 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+
11	Not past year				53	13	1. R3/(R3+R4)	1-3: 7 1,2,4: 4 1,4: 5	1-3: 35 1,2,4: 0 1,4: 1	1-3: 1
12		Not past year			93	14	1. R3/(R3+R4)	1-3: 33 1,2,4: 4 1,4: 6	1-3: 44 1,2,4: 2	1-3: 4
13	Not past year	Not past year			18	13,14	1. R3/(R3+R4)	1-3: 6 1,2,4: 0 1,4: 2	1-3: 10	No Cases
14	Not past month				24	15	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 6 1,2,4: 0 1,4: 9	1-3: 4 1,2,4: 0 1,4: 1	1-3: 1 1,2,4: 1 1,4: 2
15		Not past month			54	16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 22 1,2,4: 0 1,4: 9	1-3: 14 1,2,4: 1 1,4: 3	1-3: 4 1,2,4: 0 1,4: 1
16	Not past month	Not past month			11	15,16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 6 1,2,4: 0 1,4: 1	1-3: 2 1,2,4: 0 1,4: 1	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

Table D.56 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+
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 (lifetime use known)		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
19	Not past month	Missing (lifetime use imputed)		Missing	0					
20	Missing (lifetime use known)	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
20	Missing (lifetime use imputed)	Not past month	Missing		0					
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)	No Cases	1-3: 1	No Cases
23	Not past year	Missing (lifetime use known)		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
23	Not past year	Missing (lifetime use imputed)		Missing	0					

Table D.56 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+
24	Missing (lifetime use known)	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
24	Missing (lifetime use imputed)	Not past year	Missing		0					
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
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		0	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)	No Cases	No Cases	No Cases

Table D.56 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+
28	Past year but not past month, or past 3 years but not past year				29	17	1. R2/(R2+R3)	1-3: 11 1,2,4: 0 1,4: 3	1-3: 13	1-3: 2
29		Past year but not past month, or past 3 years but not past year			62	18	1. R2/(R2+R3)	1-3: 31 1,2,4: 1 1,4: 2	1-3: 25	1-3: 3
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			15	17,18	1. R2/(R2+R3)	1-3: 6	1-3: 9	No Cases
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		0	5,17,18	1. R2/(R2+R3) 2. DC 3. PMC	No Cases	No Cases	No Cases

Table D.56 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+
33	Past year but not past month, or past 3 years but not past year	Missing (lifetime use known)		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
33	Past year but not past month, or past 3 years but not past year	Missing (lifetime use imputed)		Missing	0					
34	Missing (lifetime use known)	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
34	Missing (lifetime use imputed)	Past year but not past month, or past 3 years but not past year	Missing		0					
35	Past year but not past month, or past 3 years but not past year	Not past year			1	14,17	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	1-3: 1	No Cases

Table D.56 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			1	13,18	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 0 1,2,4: 0 1,4: 1	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			1	15,18	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	1-3: 1	No Cases
39	Missing (lifetime use known)	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
39	Missing (lifetime use imputed)	Past 3 years	Missing	Missing	0					

Table D.56 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+
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			2	16,17	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 1	1-3: 0 1,2,4: 0 1,4: 1	No Cases

Table D.56 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		1	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)	1-3: 0 1,2,4: 0 1,4: 1	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 D.57 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 D.58 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 (lifetime use known)	1	None	1. R1	1,2: 1	1,2: 2	1,2: 1
	Missing (lifetime use imputed)	3					

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month pipe use | lifetime pipe use)

Table D.59 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 D.59 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 D.59 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 D.60 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 D.61 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		11	1,2,4-11,17,19	1. PM 2. PY	1-3: 0 1,3: 5	1-3: 4	1-3: 2
2	(Past month)		Missing		196	4-6,8,9,11	1. PM	1-3: 30	1-3: 92 1,3: 1	1-3: 73
3	(Past month)	Missing			185	1-3,8,10,18	1. PY	1-3: 57 1,3: 2	1-3: 60	1-3: 66
4	(Past year but not past month)	Missing			142	1,2,12,13	1. PY	1-3: 74	1-3: 40	1-3: 28
5	(Past month)			Missing	560	8,14,15	1. PMB	1-3: 95	1-3: 275	1-3: 190
6	(Past month)		Missing	Missing	19	4,5,8,9,11	1. PM 2. PMB	1-3: 10 1,3: 1	1-3: 4	1-3: 3 1,3: 1
7	(Past month)	Missing		Missing	64	1-3,8,10,14, 15,18	1. PY 2. PMB	1-3: 9 1,3: 6	1-3: 26 1,3: 1	1-3: 20 1,3: 2
8	(Past month)	Missing	Missing	Missing	19	1,2,4,5,7-11, 17,20	1. PM 2. PY 3. PMB	1-3: 3 1,3: 8	1-3: 2	1-3: 1 1,3: 5
9	Past year		Missing	Missing	352	4,5,9,11,16	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. R1*PMB/(R1+R2)	1-3: 76 1,3: 46	1-3: 72 1,3: 48	1-3: 62 1,3: 48
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: 5 1,3: 17	1-3: 4 1,3: 24	1-3: 2 1,3: 28

Table D.61 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	422	2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY 5. R1*PMB	1-3: 99 1,3: 229	1-3: 2 1,3: 85	1-3: 1 1,3: 16
11	Lifetime (imputed)	Missing	Missing	Missing	10					

[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 D.62 Constraints and Portion of the Predictive Mean Vector for Inhalants 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	6	1,2,4,5,8-11, 17,20	1. PM 2. PY	1-3: 1 1,3: 5	No Cases	No Cases
2	(Past month)		Missing	2	4,5,8,9,11	1. PM	1-3: 1 1,3: 1	No Cases	No Cases
3	(Past month)	Missing		18	1-3,8,10,18	1. PY	1-3: 12 1,3: 3	1-3: 0 1,3: 2	1-3: 0 1,3: 1
4	(Past year but not past month)	Missing		23	1,2,12,13	1. PY	1-3: 18 1,3: 1	1-3: 1 1,3: 1	1-3: 1 1,3: 1
5	Past year		Missing	17	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	1-3: 5 1,3: 7	1-3: 0 1,3: 4	1-3: 0 1,3: 1
6	Past year	Missing	Missing	9	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: 8	1-3: 0 1,3: 1	No Cases
7	Missing (lifetime use known)	Missing	Missing	222	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 14 1,3: 182	1-3: 0 1,3: 24	1-3: 0 1,3: 11
7	Missing (lifetime use imputed)	Missing	Missing	9					

[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 D.63 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	9	1,2,4,5,8-11, 17,20	1. PM 2. PY	1-3: 1 1,3: 2	1-3: 3 1,3: 1	1-3: 2
2	(Past month)		Missing	15	4,5,8,9,11	1. PM	1-3: 10	1-3: 4	1-3: 1
3	(Past month)	Missing		84	1,2,3,8,10,18	1. PY	1-3: 47 1,3: 1	1-3: 28	1-3: 8
4	(Past year but not past month)	Missing		55	1,2,12,13	1. PY	1-3: 26	1-3: 21	1-3: 8
5	Past year		Missing	87	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-3: 26 1,3: 6	1-3: 32 1,3: 1	1-3: 14 1,3: 8
6	Past year	Missing	Missing	145	1,2,4,5,7,9, 10-12,16,17,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-3: 4 1,3: 45	1-3: 15 1,3: 62	1-3: 0 1,3: 19
7	Missing (lifetime use known)	Missing	Missing	236	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 12 1,3: 135	1-3: 6 1,3: 81	1-3: 0 1,3: 30
7	Missing (lifetime use imputed)	Missing	Missing	28					

[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 D.64 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	1	4,5,8,9,11	1. PM	No Cases	1-3: 0 1,3: 1	No Cases
3	(Past month)	Missing		1	1-3,8,10,18	1. PY	No Cases	No Cases	1-3: 0 1,3: 1
4	(Past year but not past month)	Missing		2	1,2,12,13	1. PY	No Cases	1-3: 1	1-3: 0 1,3: 0 3: 1
5	Past year		Missing	1	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	1-3: 0 1,3: 1
6	Past year	Missing	Missing	15	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: 1 1,3: 9	1-3: 0 1,3: 4
7	Missing (lifetime use known)	Missing	Missing	11	1,2,4,5,7,9, 10-12,17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 0 1,3: 5	1-3: 0 1,3: 6	1-3: 0 1,3: 1
7	Missing (lifetime use imputed)	Missing	Missing	1					

[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 D.65 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	12	1,2,10,21	1. PY	1-3: 4 1,3: 3	1-3: 4	1-3: 1
2	(Past year but not past month)	Missing	19	1,2,10	1. PY	1-3: 4 1,3: 2	1-3: 4	1-3: 8 1,3: 1
3	Past year		3	16	1. R1/(R1+R2)	1-3: 2 1,3: 1	No Cases	No Cases
4	Past year	Missing	13	1,2,10,12,22	1. R1/(R1+R2) 2. PY	1-3: 0 1,3: 1	1-3: 2 1,3: 8	1-3: 0 1,3: 2
5	Missing (lifetime use known)	Missing	92	1,2,10,12	1. R1 2. R2 3. (R1+R2)*PY	1-3: 0 1,3: 34	1-3: 1 1,3: 47	1-3: 0 1,3: 15
5	Missing (lifetime use imputed)	Missing	5					

[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 D.66 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	3	1,2,10,21	1. PY	No Cases	1-3: 0 1,3: 1	1-3: 1 1,3: 1
2	(Past year but not past month)	Missing	2	1,2,10	1. PY	1-3: 0 1,3: 2	No Cases	No Cases
3	Past year		0	16	1. R1/(R1+R2)	No Cases	No Cases	No Cases
4	Past year	Missing	2	1,2,10,12,22	1. R1/(R1+R2) 2. PY	1-3: 0 1,3: 1	No Cases	1-3: 0 1,3: 1
5	Missing (lifetime use known)	Missing	37	1,2,10,12	1. R1 2. R2 3. (R1+R2)*PY	1-3: 0 1,3: 23	1-3: 0 1,3: 12	1-3: 0 1,3: 8
5	Missing (lifetime use imputed)	Missing	6					

[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 D.67 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	Recipient's proportion of past year cocaine use * 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)
LogC5	Donor's proportion of past year cocaine use * recipient's number of days could have used cocaine in the past year must be greater than or equal to 30-day cocaine use
LogC6	Donor's 30-day cocaine use must be less than number of days between recipient's interview date and birthday (inclusive)
LogC7	Donor's 30-day cocaine use must be less than recipient's maximum number of days could have used cocaine in past 30 days
LogC8	Donor's 30-day cocaine use must be greater than recipient's minimum number of days could have used cocaine in past 30 days
LogC9	If recipient's age at first cocaine use equals his or her current age, donor's 30-day cocaine frequency (1) cannot be greater than recipient's days between his or her interview date and date of first cocaine 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 cocaine use equals his or her current age, (1) donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine 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 cocaine use * recipient's maximum number of days could have used cocaine in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)

Table D.67 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC11	Recipient's estimated cocaine 30-day frequency is not given/legitimately skipped (estimated cocaine frequency ≠ 1-6)
LogC12	Donor's crack recency equals recipient's crack recency
LogC13	Donor must be a past year but not past month cocaine user (cocaine recency = 2)
LogC14	If recipient's age at first cocaine use equals his or her current age, donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year cannot be greater than recipient's days between his or her interview date and date of first cocaine use (minus 29)
LogC15	Donor must be a past month or past year but not past month cocaine user (cocaine recency = 1 or 2)
LogC16	Donor must be a past month, past year but not past month, or a lifetime but not past year cocaine user (cocaine recency = 1, 2, or 3)
LogC17	If recipient's age at first cocaine use equals his or her current age, donor cannot be a lifetime but not past year cocaine user (cocaine recency ≠ 3)
LogC18	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)
LogC19	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
LogC20	Recipient's proportion of past year crack use * maximum number of days could have used crack in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC21	Donor's proportion of past year crack use * recipient's number of days could have used crack in the past year must be greater than or equal to 30-day crack use
LogC22	Donor's 30-day crack use must be less than number of days between recipient's interview date and birthday (inclusive)
LogC23	Donor's 30-day crack use must be less than recipient's maximum number of days could have used crack in past 30 days

Table D.67 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC24	Donor's 30-day crack use must be greater than recipient's minimum number of days could have used crack in past 30 days
LogC25	If recipient's age at first crack use equals his or her current age, donor's 30-day crack frequency (1) cannot be greater than recipient's days between his or her interview date and date of first crack use (inclusive) and (2) cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC26	If recipient's age at first crack use equals his or her current age, (1) donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year cannot be greater than recipient's days between his or her interview date and date of first crack use (inclusive) and (2) donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC27	Recipient's estimated 30-day crack frequency is not given/legitimately skipped (estimated crack frequency ≠ 1-6)
LogC28	Donor must be a past month crack user (crack recency = 1)
LogC29	Donor must be a past month or past year not past month crack user (crack recency = 1, 2)
LogC30	Donor must be a past month, past year not past month, or lifetime but not past year crack user (crack recency = 1, 2)
LogC31	Donor's cocaine recency must equal recipient's cocaine recency, or donor's cocaine recency must equal recipient's cocaine recency (minus 10)
LogC32	If recipient's age at first crack use equals his or her current age, donor cannot be a lifetime but not past year crack user (crack recency ≠ 3)
LogC33	Donor must be a past year but not past month crack user (crack recency = 2)
LogC34	If recipient's age at first crack use equals his or her current age, donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year cannot be greater than recipient's days between his or her interview date and date of first crack use (minus 29)
LogC35	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), 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
LogC36	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), 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
LogC37	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
LogC38	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), 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 D.67 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC39	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), 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 donor's 30-day crack frequency
LogC40	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

Table D.68 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 drug 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 D.69 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 month)		Missing		Missing		8	1-12,36	1. PM 2. PY	1-4: 0 1-3: 0 1,3: 3	1-4: 0 1-3: 0 1,3: 2	1-4: 1 1-3: 0 1,3: 2
2	(Past month)				Missing		8	1,6-9,11,12	1. PM	1-4: 0 1-3: 2	1-4: 0 1-3: 2 1,3: 1	1-4: 0 1-3: 1 1,3: 1 3: 1
3	(Past month)		Missing				6	2-4,10,12,35	1. PY	1-4: 0 1-3: 0 1,3: 1	1-4: 1 1-3: 0 1,3: 1 3: 1	1-4: 2
4	(Past year not past month)		Missing				17	2-4,12-14	1. PY	1-4: 1	1-4: 0 1-3: 3	1-4: 7 1-3: 3 1,3: 1 3: 0 None: 1
5	Past year				Missing		15	6-9,11,12,15	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-4: 1 1-3: 1 1,3: 2	1-4: 4 1-3: 2 1,3: 3	1-4: 0 1-3: 0 1,3: 2
6	Past year		Missing		Missing		6	2-12,15,37	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: 2	1-4: 0 1-3: 0 1,3: 2
7	Missing (lifetime use known)		Missing		Missing		53	2-12,16,17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 14	1-4: 0 1-3: 0 1,3: 33	1-4: 0 1-3: 1 1,3: 8
7	Missing (lifetime use imputed)		Missing		Missing		3					
8	(Past month)	(Past month)		Missing		Missing	0	1,18-27,39	1. PM 2. PY	No Cases	No Cases	No Cases

Table D.69 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+
9	(Past month)	(Past month)				Missing	0	1,22-25,27,28	1. PM	No Cases	No Cases	No Cases
10	(Past month)	(Past month)		Missing			1	15,18-20,26, 28,38	1. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
11	(Past year not missing)	(Past year not past month)		Missing			0	15,18-20,26,29	1. PY	No Cases	No Cases	No Cases
12	(Past month)	Past year				Missing	0	1,22-25,27,29	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
13	(Past month)	Past year		Missing		Missing	0	1,18-27,29,40	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
14	(Past month)	Missing (lifetime use known)		Missing		Missing	2	16,18-26,30-32	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 0 3: 1	1-4: 0 1-3: 0 1,3: 1	No Cases
14	(Past month)	Missing (lifetime use imputed)		Missing		Missing	0					
15	(Past month)	(Past month)	Missing	Missing			0	1-4,10,18-20, 26,28,35,38	1. PY	No Cases	No Cases	No Cases
16	(Past month)	(Past year but not past month)	Missing	Missing			0	1-4,10,18-20, 26,33,35	1. PY	No Cases	No Cases	No Cases
17	(Past year but not past month)	(Past year but not past month)	Missing	Missing			0	2-4,14,18-20, 33,34	1. PY	No Cases	No Cases	No Cases

Table D.69 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

	Missingness Pattern									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.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	12-17	18-25	26+
18	(Past month)	(Past month)			Missing	Missing	0	1,6-9,11,22-25, 27,28	1. PM	No Cases	No Cases	No Cases
19	(Past month)	(Past month)	Missing	Missing	Missing	Missing	0	1-11,18-28,36, 39	1. PM 2. PY	No Cases	No Cases	No Cases
20	(Past month)	(Past month)	Missing		Missing	Missing	0	1-11,16,22-25, 27,28,36	1. PM 2. PY	No Cases	No Cases	No Cases
21	(Past month)	(Past month)		Missing	Missing	Missing	0	1,6-9,11,18-28, 39	1. PM 2. PY	No Cases	No Cases	No Cases
22	(Past month)	(Past month)	Missing	Missing	Missing		0	1-11,18-21,26, 28,36,38	1. PM 2. PY	No Cases	No Cases	No Cases
23	(Past month)	(Past year not past month)	Missing	Missing	Missing		0	1-11,18-20,33, 34,36	1. PM 2. PY	No Cases	No Cases	No Cases
24	(Past month)	(Past month)	Missing	Missing		Missing	1	1-4,10,18-26, 28,36	1. PM 2. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
25	(Past month)	(Past month)		Missing	Missing		0	1,6-9,18-20,26, 28,38	1. PM 2. PY	No Cases	No Cases	No Cases
26	(Past month)	(Past year not past month)		Missing	Missing		0	1,6-9,11,18-20, 26,33	1. PM 2. PY	No Cases	No Cases	No Cases
27	(Past month)	(Past month)	Missing			Missing	0	1-4,10,22-25, 27,28,35	1. PM 2. PY	No Cases	No Cases	No Cases
28	Past year	Past year			Missing	Missing	0	6-9,11,15, 22-25,27,29	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
29	Past year	Past year	Missing		Missing	Missing	0	2-11,15,21-25, 27,29,37	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
30	Past year	Past year		Missing	Missing	Missing	2	6-9,11,15, 18-27,29,40	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-4: 0 1-3: 0 1,3: 2	No Cases	No Cases

Table D.69 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+
31	Past year	Past year	Missing	Missing	Missing	Missing	1	2-11,15,18-27, 29,37,40	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-4: 0 1-3: 0 1,3: 1	No Cases
32	Past year	Missing (lifetime use known)		Missing	Missing	Missing	4	1,6-9,11,15, 18-27,30	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-4: 0 1-3: 0 1,3: 4	No Cases
32	Past year	Missing (lifetime use imputed)		Missing	Missing	Missing	0					
33	Past year	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	2-11,15,18-27, 30,32,37	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
33	Past year	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0					
34	(Past month)	Missing (lifetime use known)		Missing	Missing	Missing	0	1,6-9,11,18-27, 30,32	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
34	(Past month)	Missing (lifetime use imputed)		Missing	Missing	Missing	0					
35	(Past month)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	1-11,18-27,30, 36	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
35	(Past month)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0					

Table D.69 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+
36	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	8	2-11,16-27,30	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 1	1-4: 0 1-3: 0 1,3: 4	1-4: 0 1-3: 0 1,3: 4
36	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0					
36	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0					
36	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	1					
37	(Past month)	Past year		Missing	Missing	Missing	0	1,6-9,11,18-27, 29,40	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
38	(Past month)	Past year			Missing	Missing	0	1,6-9,22-25, 27,29	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases

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 D.70 Logical Constraints for Hallucinogens Recency and Frequency

Constraint #	Logical Constraint
LogC1	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)
LogC2	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
LogC3	Recipient's proportion of past year hallucinogen use * 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)
LogC4	Donor's 30-day hallucinogen use must be less than the number of days between recipient's interview date and birthday (inclusive)
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's LSD recency must equal recipient's LSD recency
LogC11	Donor's PCP recency must equal recipient's PCP recency
LogC12	Donor's Ecstasy recency must equal recipient's Ecstasy recency
LogC13	Donor must be an LSD and PCP user (LSD and PCP recencies ≠ 91)
LogC14	Donor must be an LSD and Ecstasy user (LSD and Ecstasy recencies ≠ 91)
LogC15	Donor must be a PCP and Ecstasy user (PCP and Ecstasy recencies ≠ 91)
LogC16	Donor must be an LSD and PCP and Ecstasy user (LSD and PCP and Ecstasy recencies ≠ 91)

Table D.70 Logical Constraints for Hallucinogens Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC17	Donor must be a past month hallucinogens user (hallucinogen recency = 1)
LogC18	Donor must be a hallucinogen past year but not past month or past month hallucinogen user (hallucinogen recency = 1 or 2)
LogC19	Donor must be a hallucinogen user (hallucinogen recency = 1, 2, or 3)
LogC20	Donor must be an LSD past year but not past month or past month LSD user (LSD recency = 1 or 2)
LogC21	Donor must be a PCP past year but not past month or past month PCP user (PCP recency = 1 or 2)
LogC22	Donor must be an Ecstasy past year but not past month or past month Ecstasy user (Ecstasy recency = 1 or 2)
LogC23	Donor must not be an LSD past year but not past month or past month LSD user (LSD recency ≠ 1 or 2)
LogC24	Donor must not be a PCP past year but not past month or past month PCP user (PCP recency ≠ 1 or 2)
LogC25	Donor must not be an Ecstasy past year but not past month or past month Ecstasy user (Ecstasy recency ≠ 1 or 2)
LogC26	Donor's hallucinogen recency must equal recipient's hallucinogen recency, or donor's hallucinogen recency must equal recipient's hallucinogen recency (minus 10)
LogC27	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), 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
LogC28	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), 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 donor's 30-day hallucinogen frequency
LogC29	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
LogC30	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 donor's 30-day hallucinogen frequency
LogC31	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 D.71 Likeness Constraints for Hallucinogens Recency and Frequency

Constraint #	Likeness Constraint
LikC1	Donor's 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)

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table D.72 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 (lifetime use known)					0	7,11,12,26	1. R1 2. R2	No Cases	No Cases	No Cases
1		Missing (lifetime use imputed)					0					
2			Missing (lifetime use known)				2	8,10,12,26	1. R1 2. R2	1-5: 0 1-4: 0 1-3: 2	No Cases	No Cases
2			Missing (lifetime use imputed)				0					
3		Missing (lifetime use known)	Missing (lifetime use known)				0	7,8,12,26,31	1. R1 2. R2	No Cases	No Cases	No Cases
3		Missing (lifetime use known)	Missing (lifetime use imputed)				0					
3		Missing (lifetime use imputed)	Missing (lifetime use known)				0					
3		Missing (lifetime use imputed)	Missing (lifetime use imputed)				0					

Table D.72 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+
4	(Past month)				Missing	Missing	1	1-6,17,28	1. PM 2. PY	No Cases	No Cases	1-5: 0 1-4: 0 1-3: 1
5	(Past month)					Missing	5	4-6,17	1. PM	1-5: 0 1-4: 0 1-3: 1	1-5: 2 1-4: 0 1-3: 1	1-5: 0 1-4: 0 1-3: 1
6	(Past year)				Missing		19	1-3,18,30	1. PY	1-5: 2 1-4: 4 1-3: 2	1-5: 5 1-4: 4 1-3: 1	1-5: 0 1-4: 0 1-3: 1
7	(Past month)	Missing (lifetime use known)				Missing	0	4-7,11,12,17	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
7	(Past month)	Missing (lifetime use imputed)				Missing	0					
8	(Past month)		Missing (lifetime use known)			Missing	0	4-6,8,10,12, 17	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
8	(Past month)		Missing (lifetime use imputed)			Missing	0					

Table D.72 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+
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)			Missing	0	4-8,12,17,31	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)			Missing	0					
9	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)			Missing	0					
9	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)			Missing	0					
10	(Past year)	Missing (lifetime use known)			Missing		0	1-3,7,11,12, 18,30	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
10	(Past year)	Missing (lifetime use imputed)			Missing		0					
11	(Past year)		Missing (lifetime use known)		Missing		0	1-3,8,10,12, 18,30	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
11	(Past year)		Missing (lifetime use imputed)		Missing		0					

Table D.72 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+
12	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)		Missing		0	1-3,7,8,12,18, 30,31	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
12	(Past year)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing		0					
12	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing		0					
12	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing		0					
13	(Past month)	Missing (lifetime use known)			Missing	Missing	0	1-6,7,11,12, 17,28	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
13	(Past month)	Missing (lifetime use imputed)			Missing	Missing	0					
14	(Past month)		Missing (lifetime use known)		Missing	Missing	0	1-6,8,10,12, 17,28	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
14	(Past month)		Missing (lifetime use imputed)		Missing	Missing	0					

Table D.72 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+
15	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	Missing	0	1-8,12,17,28, 31	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
15	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	Missing	0					
15	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	Missing	0					
15	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	Missing	0					
16	Past year	(Not past month)	(Not past month)	(Not past month)		Missing	13	4-6,10-12,18	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 3	1-5: 4 1-4: 2 1-3: 3	1-5: 0 1-4: 0 1-3: 1
17	Past year	(Not past month)	(Not past month)	(Not past month)	Missing	Missing	7	1-6,10-12,18, 29	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4: 0 1-3: 6	No Cases	1-5: 0 1-4: 0 1-3: 1
18	Past year	Past year	(Not past month)	(Not past month)		Missing	7	4-6,11,12,18, 20,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 4	1-5: 0 1-4: 0 1-3: 2	1-5: 0 1-4: 0 1-3: 0 1,2: 1

Table D.72 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+
19	Past year	(Not past month)	Past year	(Not past month)		Missing	0	4-6,10,12,18, 21,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
20	Past year	Past year	Past year	(Not past month)		Missing	1	4-6,12,18,20, 21,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 0 1,2: 1	No Cases	No Cases
21	Past year	Missing (lifetime use known)	(Not past month)	(Not past month)		Missing	10	4-7,11,12,18, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 2	1-5: 1 1-4: 2 1-3: 4	1-5: 0 1-4: 0 1-3: 1
21	Past year	Missing (lifetime use imputed)	(Not past month)	(Not past month)		Missing	0					
22	Past year	(Not past month)	Missing (lifetime use known)	(Not past month)		Missing	1	4-6,8,10,12, 18,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-5: 0 1-4: 0 1-3: 1	No Cases
22	Past year	(Not past month)	Missing (lifetime use imputed)	(Not past month)		Missing	0					

Table D.72 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+
23	Past year	Missing (lifetime use known)	Missing (lifetime use known)	(Not past month)		Missing	0	4-8,12,18,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
23	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past month)		Missing	0					
23	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past month)		Missing	0					
23	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past month)		Missing	0					
24	Past year	Past year	(Not past month)	(Not past month)	Missing	Missing	2	1-6,11,12,18, 20,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-5: 0 1-4: 0 1-3: 2	No Cases
25	Past year	(Not past month)	Past year	(Not past month)	Missing	Missing	0	1-6,10,12,18, 21,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
26	Past year	Past year	Past year	(Not past month)	Missing	Missing	0	1-6,12,18,20, 21,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases

Table D.72 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+
27	Past year	Missing (lifetime use known)	(Not past month)	(Not past month)	Missing	Missing	1	1-7,11,12,18, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
27	Past year	Missing (lifetime use imputed)	(Not past month)	(Not past month)	Missing	Missing	0					
28	Past year	(Not past month)	Missing (lifetime use known)	(Not past month)	Missing	Missing	0	1-6,8,11,12, 18,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
28	Past year	(Not past month)	Missing (lifetime use imputed)	(Not past month)	Missing	Missing	0					
29	Past year	Missing (lifetime use known)	Missing (lifetime use known)	(Not past month)	Missing	Missing	0	1-8,12,18,29, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
29	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past month)	Missing	Missing	0					
29	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past month)	Missing	Missing	0					
29	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past month)	Missing	Missing	0					

Table D.72 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+**
30	Missing (lifetime use known)	(Not past year)	(Not past year)	(Not past year)	Missing	Missing	43	1-6,10-12,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 25	1-5: 0 1-4: 0 1-3: 16	1-5: 0 1-4: 0 1-3: 5
30	Missing (lifetime use imputed)	(Not past year)	(Not past year)	(Not past year)	Missing	Missing	3					
31	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing	Missing	37	1-7,11,12,19, 31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 5	1-5: 0 1-4: 0 1-3: 21	1-5: 0 1-4: 0 1-3: 11
31	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing	Missing	0					
31	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing	Missing	0					
31	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing	Missing	0					
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	(Not past year)	Missing	Missing	10	1-6,8,10,12, 19,31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 3	1-5: 0 1-4: 0 1-3: 2	1-5: 0 1-4: 0 1-3: 5
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0					

Table D.72 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern: Hallucinogens Recency	Missingness Pattern: LSD Recency	Missingness Pattern: PCP Recency	Missingness Pattern: Ecstasy Recency	Missingness Pattern: Hallucinogen 12-Mo. Freq.	Missingness Pattern: Hallucinogen 30-Day Freq.	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group: 12-17	Likeness Constraints: Number of Cases, by Age Group: 18-25	Likeness Constraints: Number of Cases, by Age Group: 26+
32	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0					
32	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0					
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing	Missing	3	1-8,12,19,31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 1	No Cases	1-5: 0 1-4: 0 1-3: 2
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0					
33	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0					
33	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0					
33	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0					
33	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0					

Table D.72 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+
33	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0					
33	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0					
34				Missing (lifetime use known)			2	9-11,26	1. R1 2. R2	1-5: 0 1-4: 0 1-3: 1	1-5: 0 1-4: 0 1-3: 2	No Cases
34				Missing (lifetime use imputed)			1					
35		Missing (lifetime use known)		Missing (lifetime use known)			2	7,9,11,26,31	1. R1 2. R2	1-5: 0 1-4: 0 1-3: 1	1-5: 0 1-4: 0 1-3: 1	No Cases
35		Missing (lifetime use known)		Missing (lifetime use imputed)			0					
35		Missing (lifetime use imputed)		Missing (lifetime use known)			0					
35		Missing (lifetime use imputed)		Missing (lifetime use imputed)			0					

Table D.72 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+
36			Missing (lifetime use known)	Missing (lifetime use known)			0	8-10,26,31	1. R1 2. R2	No Cases	No Cases	No Cases
36			Missing (lifetime use known)	Missing (lifetime use imputed)			0					
36			Missing (lifetime use imputed)	Missing (lifetime use known)			0					
36			Missing (lifetime use imputed)	Missing (lifetime use imputed)			0					
37		Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)			0	7-9,26,31	1. R1 2. R2	No Cases	No Cases	No Cases
37		Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)			0					
37		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)			0					
37		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)			0					

Table D.72 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+
37		Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)			0					
37		Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)			0					
37		Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)			0					
37		Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)			0					
38	(Past month)			Missing (lifetime use known)		Missing	0	4-6,9-11,17	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
38	(Past month)			Missing (lifetime use imputed)		Missing	0					
39	(Past month)	Missing (lifetime use known)		Missing (lifetime use known)		Missing	0	4-7,9,11,17, 31	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
39	(Past month)	Missing (lifetime use known)		Missing (lifetime use imputed)		Missing	0					

Table D.72 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+
39	(Past month)	Missing (lifetime use imputed)		Missing (lifetime use known)		Missing	0					
39	(Past month)	Missing (lifetime use imputed)		Missing (lifetime use imputed)		Missing	0					
40	(Past month)		Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	4-6,8-10,17, 31	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
40	(Past month)		Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0					
40	(Past month)		Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0					
40	(Past month)		Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0					
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	4-9,17,31	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0					

Table D.72 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+
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0					
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0					
41	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0					
41	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0					
41	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0					
41	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0					
42	(Past year)			Missing (lifetime use known)	Missing		0	1-3,9-11,18, 30	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
42	(Past year)			Missing (lifetime use imputed)	Missing		0					

Table D.72 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+
43	(Past year)	Missing (lifetime use known)		Missing (lifetime use known)	Missing		0	1-3,7,9,11,18, 30,31	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
43	(Past year)	Missing (lifetime use known)		Missing (lifetime use imputed)	Missing		0					
43	(Past year)	Missing (lifetime use imputed)		Missing (lifetime use known)	Missing		0					
43	(Past year)	Missing (lifetime use imputed)		Missing (lifetime use imputed)	Missing		0					
44	(Past year)		Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	1-3,8-10,18, 30,31	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
44	(Past year)		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing		0					
44	(Past year)		Missing (lifetime use imputed)	Missing (lifetime use known)	Missing		0					
44	(Past year)		Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing		0					

Table D.72 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+
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	1-3,7-9,18, 30,31	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing		0					
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing		0					
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing		0					
45	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing		0					
45	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing		0					
45	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing		0					
45	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing		0					

Table D.72 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern: Hallucinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 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+
46	(Past month)			Missing (lifetime use known)	Missing	Missing	0	1-6,9-11,17, 28	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
46	(Past month)			Missing (lifetime use imputed)	Missing	Missing	0					
47	(Past month)	Missing (lifetime use known)		Missing (lifetime use known)	Missing	Missing	0	1-7,9,11,17, 28,31	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
47	(Past month)	Missing (lifetime use known)		Missing (lifetime use imputed)	Missing	Missing	0					
47	(Past month)	Missing (lifetime use imputed)		Missing (lifetime use known)	Missing	Missing	0					
47	(Past month)	Missing (lifetime use imputed)		Missing (lifetime use imputed)	Missing	Missing	0					
48	(Past month)		Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	1-6,8-10,17, 28,31	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
48	(Past month)		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern: Hallu-cinogens Recency	Missingness Pattern: LSD Recency	Missingness Pattern: PCP Recency	Missingness Pattern: Ecstasy Recency	Missingness Pattern: Hallu-cinogen 12-Mo. Freq.	Missingness Pattern: Hallu-cinogen 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+
48	(Past month)		Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
48	(Past month)		Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	1-9,17,28,31	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					
49	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0					
49	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 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)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
49	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					
50	Past year	(Not past month)	(Not past month)	Past year		Missing	8	4-6,10,11,18, 22,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 1 1-3: 2	1-5: 0 1-4: 0 1-3: 4	1-5: 0 1-4: 0 1-3: 1
51	Past year	Past year	(Not past month)	Past year		Missing	1	4-6,11,18,20, 22,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
52	Past year	(Not past month)	Past year	Past year		Missing	1	4-6,10,18,21, 22,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-5: 0 1-4: 0 1-3: 1	No Cases
53	Past year	Past year	Past year	Past year		Missing	0	4-6,18,20-22, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
54	Past year	(Not past month)	(Not past month)	Missing (lifetime use known)		Missing	5	4-6,9-11,18, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 2	1-5: 0 1-4: 2 1-3: 1	No Cases
54	Past year	(Not past month)	(Not past month)	Missing (lifetime use imputed)		Missing	0					

Table D.72 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+
55	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use known)		Missing	0	4-7,9,11,18, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
55	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use imputed)		Missing	0					
55	Past year	Missing (lifetime use imputed)	(Not past month)	Missing (lifetime use known)		Missing	0					
55	Past year	Missing (lifetime use imputed)	(Not past month)	Missing (lifetime use imputed)		Missing	0					
56	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	4-6,8-10,18, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
56	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0					
56	Past year	(Not past month)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0					
56	Past year	(Not past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0					

Table D.72 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern: 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]	Likeness Constraints: Number of Cases, by Age Group: 12-17	18-25	26+
57	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	4-9,18,31	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	No Cases	No Cases	No Cases
57	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0					
57	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0					
57	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0					
57	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0					
57	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0					
57	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0					
57	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0					

Table D.72 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+
58	Past year	(Not past month)	(Not past month)	Past year	Missing	Missing	5	1-6,10,11,18, 22,29,31	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	1-5: 0 1-4: 0 1-3: 2	1-5: 0 1-4: 1 1-3: 1	1-5: 0 1-4: 0 1-3: 1
59	Past year	Past year	(Not past month)	Past year	Missing	Missing	3	1-6,11,18,20, 22,29,31	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	1-5: 0 1-4: 0 1-3: 2	1-5: 0 1-4: 0 1-3: 1	No Cases
60	Past year	(Not past month)	Past year	Past year	Missing	Missing	1	1-6,10,18,21, 22,29,31	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
61	Past year	Past year	Past year	Past year	Missing	Missing	0	1-6,18,20-22, 29,31	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	No Cases
62	Past year	(Not past month)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0	1-6,9-11,18, 29,31	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	No Cases
62	Past year	(Not past month)	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 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+
63	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0	1-7,9,11,18, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
63	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0					
63	Past year	Missing (lifetime use imputed)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0					
63	Past year	Missing (lifetime use imputed)	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0					
64	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	1-6,8-10,18, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
64	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
64	Past year	(Not past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
64	Past year	(Not past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 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+
65	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	1-9,18,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
65	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
65	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
65	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					
65	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0					
65	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
65	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
65	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 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+
66	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing (lifetime use known)	Missing	Missing	64	1-6,9-11,19, 31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 28	1-5: 0 1-4: 0 1-3: 34	1-5: 0 1-4: 0 1-3: 3
66	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	1					
66	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing (lifetime use known)	Missing	Missing	0					
66	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0					
67	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing	Missing	3	1-7,9,11,19, 31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 1	1-5: 0 1-4: 0 1-3: 2	No Cases
67	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0					
67	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	Missing	Missing	0					
67	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 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+
67	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing	Missing	0					
67	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0					
67	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	Missing	Missing	0					
67	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0					
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	1	1-6,8-10,19, 31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	1-5: 0 1-4: 0 1-3: 1
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 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+
68	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0					
68	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
68	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
68	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	1	1-9,19,31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	1-5: 0 1-4: 0 1-3: 1	No Cases
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 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+
69	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0					
69	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
69	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
69	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					
69	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0					
69	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
69	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
69	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.72 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+
69	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0					
69	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
69	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
69	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					
70				Past year			2	10,11,22,26	1. R1/(R1+R2)	1-5: 1	1-5: 1	No Cases
71		Past year	Past year				1	12,20,21,26, 31	1. R1/(R1+R2)	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
72		Past year		Past year			0	11,20,22,26, 31	1. R1/(R1+R2)	No Cases	No Cases	No Cases
73	(Past month)	Past year			Missing	Missing	0	1-6,11,12,17, 20,28	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
74	(Past month)		Past year		Missing	Missing	0	1-6,10,12,17, 21,28	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases

Table D.72 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+
75	Past year	Missing (lifetime use known)	Past year	Missing (lifetime use known)	Missing	Missing	0	1-7,9,18,21, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
75	Past year	Missing (lifetime use known)	Past year	Missing (lifetime use imputed)	Missing	Missing	0					
75	Past year	Missing (lifetime use imputed)	Past year	Missing (lifetime use known)	Missing	Missing	0					
75	Past year	Missing (lifetime use imputed)	Past year	Missing (lifetime use imputed)	Missing	Missing	0					
76	Past year	(Not past month)	Past year	Missing (lifetime use known)	Missing	Missing	0	1-6,9,10,18, 21,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
76	Past year	(Not past month)	Past year	Missing (lifetime use imputed)	Missing	Missing	0					
77			Past year				2	10,12,21,26	1. R1/(R1+R2)	1-5: 0 1-4: 0 1-3: 2	No Cases	No Cases
78		Past year					0	11,12,20,26	1. R1/(R1+R2)	No Cases	No Cases	No Cases
79	(Past month)	Past year				Missing	1	4-7,11,12,17	1. R1/(R1+R2) 2. PM	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
80	(Past month)		Past year			Missing	0	4-6,8,10,12, 17	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases

Table D.72 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern: Hallucinogens Recency	Missingness Pattern: LSD Recency	Missingness Pattern: PCP Recency	Missingness Pattern: Ecstasy Recency	Missingness Pattern: Hallucinogen 12-Mo. Freq.	Missingness Pattern: Hallucinogen 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+
81	(Past month)			Past year		Missing	0	4-6,9-11,17	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
82	(Past month)	Past year			Missing		0	1-3,7,11,12, 17,27	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
83	(Past month)		Past year		Missing		0	1-3,8,10,12, 17,27	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
84	(Past month)			Past year	Missing		0	1-3,9-11,17, 27	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
85	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Past year		Missing	0	4-6,13,18,22, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
86	Past year	Missing (lifetime use known)	Past year		Missing	Missing	0	1-7,12,18,21, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
87	Past month	Past year	Past year	Past year	Missing		0	1-3,7-12,16, 17,27	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
88			Past year	Past year			0	11,12,21,22, 26,31	1. R1/(R1+R2)	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 D.73 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 D.73 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 D.73 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 D.74 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 D.75 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		2	1-5,23	1. PY	1-4: 2	No Cases	No Cases
2	(Past year but not past month)		Missing		9	1-3,6,7	1. PY	1-4: 3 1-3: 1 1,2: 1	1-4: 2	1-4: 2
3	Past year				1	8,27	1. R1/(R1+R2)	No Cases	1-4: 1	No Cases
4	Past year		Missing		8	1-3,5,7,8,25,27	1. R1/(R1+R2) 2. PY	1-4: 0 1-3: 0 1,2: 1	1-4: 1 1-3: 0 1,2: 3	1-4: 0 1-3: 1 1,2: 2
5	Missing (lifetime use known)		Missing		67	1,2,5,7,22,25,27	1. R1 2. R2 3. (R1+R2)*PY	1-4: 0 1-3: 0 1,2: 42	1-4: 2 1-3: 0 1,2: 16	1-4: 0 1-3: 0 1,2: 11
5	Missing (lifetime use imputed)		Missing		4					
6	(Past month)	(Past month)		Missing	0	4,15-18,20,24	1. PY	No Cases	No Cases	No Cases
7	(Past year not missing)	(Past year not past month)		Missing	0	15-17,19,21	1. PY	No Cases	No Cases	No Cases
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	0	14-19,26	1. R1/(R1+R2) 2. PY	No Cases	No Cases	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 (lifetime use known)		Missing	4	10,15,16,18,19,22,29	1. R1 2. R2 3. (R1+R2)*PY	1-4: 0 1-3: 0 1,2: 2	1-4: 0 1-3: 0 1,2: 2	No Cases
12	(Past year not missing)	Missing (lifetime use imputed)		Missing	0					

Table D.75 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	2	1-5,15-18,20, 23,24,29	1. PY	1-4: 0 1-3: 0 1,2: 2	No Cases	No Cases
14	(Past month)	(Past year not past month)	Missing	Missing	1	1-5,15-17,19, 21,23,29	1. PY	1-4: 0 1-3: 0 1,2: 0 1: 1	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			0	8,14,28	1. R1/(R1+R2)	No Cases	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	1	8,14-19,19, 26,28	1. R1/(R1+R2) 2. PY	1-4: 0 1-3: 0 1,2: 1	No Cases	No Cases
19	Past year	Past year	Missing	Missing	7	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: 2	1-4: 0 1-3: 0 1,2: 4	1-4: 0 1-3: 0 1,2: 1
20	Past year	Missing (lifetime use known)		Missing	2	8,10,15,16,18, 19,28,29	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: 1
20	Past year	Missing (lifetime use imputed)		Missing	0					

Table D.75 Constraints and Portion of the Predictive Mean Vector for Stimulants Recency and Frequency (continued)

#	Missingness Pattern: Stimulant Recency	Missingness Pattern: Meth. Recency	Missingness Pattern: Stimulant 12-Mo. Freq.	Missingness Pattern: Meth. 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+
21	Past year	Missing (lifetime use known)	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
21	Past year	Missing (lifetime use imputed)	Missing	Missing	0					
22	(Past month)	Missing (lifetime use known)		Missing	0	4,10,15,16,18,19,29	1. R1 2. R2 3. (R1+R2)*PY	No Cases	No Cases	No Cases
22	(Past month)	Missing (lifetime use imputed)		Missing	0					
23	(Past month)	Missing (lifetime use known)	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
23	(Past month)	Missing (lifetime use imputed)	Missing	Missing	0					

Table D.75 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+
24	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	39	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: 16	1-4: 0 1-3: 0 1,2: 10	1-4: 0 1-3: 0 1,2: 13
24	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0					
24	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
24	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0					
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 D.76 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency

#	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	Likeness Constraints: Number of Cases, by Age Group: 18-25	Likeness Constraints: Number of Cases, by Age Group: 26+
1	(Past month)		Missing		49	1-5,23	1. PY	1-3,5: 29 1-3: 2 1,2: 1	1-3,5: 8	1-3,5: 6 1-3: 2 1,2: 1
2	(Past year but not past month)		Missing		65	1-3,6,7	1. PY	1-3,5: 40	1-3,5: 14 1-3: 1	1-3,5: 9 1-3: 0 1,2: 1
3	Past year				4	7,8	1. R1/(R1+R2)	1-3,5: 4	No Cases	No Cases
4	Past year		Missing		23	1-3,5,7,8,25	1. R1/(R1+R2) 2. PY	1-3,5: 1 1-3: 3 1,2: 2	1-3,5: 1 1-3: 4 1,2: 6	1-3,5: 2 1-3: 1 1,2: 3
5	Missing (lifetime use known)		Missing		315	1-3,5,7,25	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 33 1-3: 2 1,2: 124	1-3,5: 12 1-3: 4 1,2: 96 1: 1	1-3,5: 5 1-3: 1 1,2: 49
5	Missing (lifetime use imputed)		Missing		12					
6	(Past month)	(Past month)		Missing	1	4,15-20,24	1. PY	No Cases	1-3,5: 0 1-3: 0 1,2: 1	No Cases
7	(Past year not missing)	(Past year not past month)		Missing	6	8,14-20	1. PY	1-3,5: 1 1-3: 0 1,2: 1	1-3,5: 0 1-3: 0 1,2: 1	1-3,5: 2 1-3: 0 1,2: 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	6	7,8,15-17,26	1. R1/(R1+R2) 2. PY	1-3,5: 3	1-3,5: 2	1-3,5: 0 1-3: 0 1,2: 1

Table D.76 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency (continued)

#	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+
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 (lifetime use known)		Missing	16	7,8,15-17	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 0 1-3: 0 1,2: 6	1-3,5: 1 1-3: 0 1,2: 9	1-3,5: 0 1-3: 0 1,2: 2
12	(Past year not missing)	Missing (lifetime use imputed)		Missing	2					
13	(Past month)	(Past month)	Missing	Missing	3	1-4,7,15-17,20, 23,24	1. PY	1-3,5: 0 1-3: 0 1,2: 3	No Cases	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	1	1-3,6,7,15-17,21	1. PY	1-3,5: 1	No Cases	No Cases
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	2	7,8,14-17,19, 26,28	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 2	No Cases	No Cases
19	Past year	Past year	Missing	Missing	4	1-3,7,8,14-17, 19,25,26,28	1. R1/(R1+R2) 2. PY	1-3,5: 1	1-3,5: 1 1-3: 0 1,2: 2	No Cases
20	Past year	Missing (lifetime use known)		Missing	9	7,8,10,15-17, 19,28	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: 4	1-3,5: 0 1-3: 0 1,2: 2
20	Past year	Missing (lifetime use imputed)		Missing	0					

Table D.76 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency (continued)

#	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+
21	Past year	Missing (lifetime use known)	Missing	Missing	2	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: 2	No Cases	No Cases
21	Past year	Missing (lifetime use imputed)	Missing	Missing	0					
22	(Past month)	Missing (lifetime use known)		Missing	0	4,7,10,15-17,19	1. R1 2. R2 3. (R1+R2)*PY	No Cases	No Cases	No Cases
22	(Past month)	Missing (lifetime use imputed)		Missing	0					
23	(Past month)	Missing (lifetime use known)	Missing	Missing	0	1-4,7,10,15-17, 19,23	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 0 1-3: 0 1,2: 1	No Cases	No Cases
23	(Past month)	Missing (lifetime use imputed)	Missing	Missing	1					

Table D.76 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	Likeness Constraints: Number of Cases, by Age Group: 18-25	Likeness Constraints: Number of Cases, by Age Group: 26+
24	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	35	1-3,7,10,15-17, 19,22,28	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 0 1-3: 0 1,2: 18	1-3,5: 0 1-3: 0 1,2: 17	1-3,5: 0 1-3: 0 1,2: 7
24	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	4					
24	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0					
24	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	3					
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	0	1-3,6-8,10, 15-17,19,25	1. R1/(R1+R2) 2. PY	No Cases	No Cases	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 D.77 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 D.78 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 D.79 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		10	8,27	1. R1/(R1+R2)	1-4: 0 1-3: 0 1,2: 1	1-4: 6	1-4: 0 1-3: 1 1,2: 2
2	Missing (lifetime use known)		67	22,27	1. R1 2. R2	1-4: 0 1-3: 0 1,2: 42	1-4: 8 1-3: 0 1,2: 12	1-4: 2 1-3: 2 1,2: 7
2	Missing (lifetime use imputed)		6					
3	(Past month)	Past year	0	14	1. R1/(R1+R2)	No Cases	No Cases	No Cases
4	(Past year not missing)	Missing (lifetime use known)	4	10	1. R1 2. R2	1-4: 0 1-3: 0 1,2: 2	1-4: 0 1-3: 0 1,2: 2	No Cases
4	(Past year not missing)	Missing (lifetime use imputed)	0					
5	Past year	Past year	11	8,14,28	1. R1/(R1+R2)	1-4: 1 1-3: 0 1,2: 2	1-4: 3 1-3: 0 1,2: 1	1-4: 0 1-3: 0 1,2: 4
6	Past year	Missing (lifetime use known)	2	8,10,28	1. R1/(R1+R2)	1-4: 1	No Cases	1-4: 1
6	Past year	Missing (lifetime use imputed)	0					

Table D.79 Constraints and Portion of the Predictive Mean Vector for Core-Plus-Noncore Stimulants Recency (continued)

#	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+
7	Missing (lifetime use known)	Missing (lifetime use known)	45	10,22,28	1. R1 2. R2	1-4: 0 1-3: 0 1,2: 16	1-4: 5 1-3: 0 1,2: 7	1-4: 4 1-3: 0 1,2: 13
7	Missing (lifetime use known)	Missing (lifetime use imputed)	0					
7	Missing (lifetime use imputed)	Missing (lifetime use known)	0					
7	Missing (lifetime use imputed)	Missing (lifetime use imputed)	0					

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)

D.4.3 Age at First Use

Table D.80 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 D.81 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 D.82 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	461	1-3,5,6,8	1. PrAFU	1-4: 229 2-4: 0 3,4: 1	1-4: 132	1-4: 91 2-4: 5 3,4: 3

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigarette age at first use

Table D.83 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	118	1-3,5,6,9	1. PrAFU	1-4: 16 2-4: 2	1-4: 27 2-4: 0 3,4: 1	1-4: 62 2-4: 4 3,4: 6

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigarette age at first daily use

Table D.84 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	472	1-3,5,6	1. PrAFU	1-4: 124 2-4: 5 3,4: 3	1-4: 147 2-4: 2	1-4: 164 2-4: 17 3,4: 7 4,7: 1 5,7: 2

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigar age at first use

Table D.85 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	652	1,2,4,6,7	1. PrAFU	1,2,4,6: 302 2,4,6: 1	1,2,4,6: 128 2,4,6: 1	1,2,4,6: 210 2,4,6: 4 4,6: 6

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted alcohol age at first use

Table D.86 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	343	1,2,4,6,7	1. PrAFU	1,2,4,6: 240 2,4,6: 2 4,6: 3	1,2,4,6: 52 2,4,6: 1 4,6: 1	1,2,4,6: 27 2,4,6: 8 4,6: 8 4,8: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted inhalants age at first use

Table D.87 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	245	1,2,4,6,7	1. PrAFU	1,2,4,6: 109	1,2,4,6: 62	1,2,4,6: 67 2,4,6: 1 4,6: 6

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted marijuana age at first use

Table D.88 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	163	1,2,4,6,7	1. PrAFU	1,2,4,6: 43 2,4,6: 5 4,6: 2	1,2,4,6: 52	1,2,4,6: 35 2,4,6: 12 4,6: 10 4,8: 0 5,8: 4

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted tranquilizers age at first use

Table D.89 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	67	1,2,4,6,7	1. PrAFU	1,2,4,6: 15 2,4,6: 9 4,6: 3	1,2,4,6: 15 2,4,6: 1	1,2,4,6: 16 2,4,6: 3 4,6: 3 4,8: 1 5,8: 0 8: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted sedatives age at first use

Table D.90 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	10	1,2,4,6,7	1. PrAFU	1,2,4,6: 3	1,2,4,6: 2	1,2,4,6: 3 2,4,6: 1 4,6: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted heroin age at first use

Table D.91 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.)
LogC10	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)
LogC11	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 minimum of recipient's child drug(s) age(s) at first use (if existing)
LogC12	For each child drug for which recipient was missing age at first use, donor must be a lifetime user
LogC13	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
LogC14	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 D.91 Logical Constraints for Age at First Use, Multivariate Assignment (continued)

Constraint #	Logical Constraint
LogC15	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.)

Table D.92 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	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
LikC4	Age of donor = Age of recipient
LikC5	Age of donor must be greater than or equal to age of recipient
LikC6	For the parent drug, if recipient was a past year user, then donor was a past year user; if recipient was a lifetime but not past year user, then donor was the same
LikC7	For each child drug, if recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime but not past year user, then donor was either a lifetime but not past year user or a lifetime nonuser[1]
LikC8	Donor agrees with recipient with respect to lifetime use for child drug(s)
LikC9	For the parent drug, if recipient was a lifetime but not past year user, then donor was a lifetime but not past year user
LikC10	For each child drug, if recipient was a lifetime but not past year user, then donor was not a past year user[1]
LikC11	Donor was at least as old as recipient, but no more than 20 years older than recipient
LikC12	Donor was no more than 20 years older than recipient
LikC13	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

[1] These constraints were intended to match recency of use for child drugs. They were not applied exactly as intended, because lifetime nonusers were not correctly considered. However, the practical impact was small because child drug nonusers could not be donors for a missing child drug age at first use.

Table D.93 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	368	1-3,6,7,12	1. PrAFU	1-4: 135 2-4: 12 3,4: 7 4,13: 2 5,13: 1	1-4: 144 2-4: 3 3,4: 1	1-4: 40 2-4: 10 3,4: 12 4,13: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted smokeless tobacco age at first use

Table D.94 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	77	1,2,4,5,7-12	1. PrAFU	1,2,4,6-8: 7 2,4,6-8: 2 2,4,6,7: 0 4,6,7: 3	1,2,4,6-8: 23 2,4,6-8: 0 2,4,6,7: 0 4,6,7: 1	1,2,4,6-8: 30 2,4,6-8: 4 2,4,6,7: 1 4,6,7: 3 4,9,10: 2 9-11: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cocaine age at first use

Table D.95 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	195	1,2,4,7,8, 10-12,14,15	1. PrAFU	1,2,4,6-8: 39 2,4,6-8: 6 2,4,6,7: 6 4,6,7: 5 4,9,10: 0 9-11: 2	1,2,4,6-8: 50 2,4,6-8: 1 2,4,6,7: 4 4,6,7: 3	1,2,4,6-8: 32 2,4,6-8: 15 2,4,6,7: 8 4,6,7: 12 4,9,10: 3 9-11: 5 9,10,12: 4

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted hallucinogens age at first use

Table D.96 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	669	1,2,4,5,7-13	1. PrAFU	1,2,4,6-8: 259 2,4,6-8: 8 2,4,6,7: 0 4,6,7: 3	1,2,4,6-8: 185 2,4,6-8: 1	1,2,4,6-8: 157 2,4,6-8: 27 2,4,6,7: 2 4,6,7: 22 4,9,10: 1 9-11: 2 9,10,12: 2

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted pain relievers age at first use

Table D.97 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	148	1,2,4,5,7-13	1. PrAFU	1,2,4,6-8: 37 2,4,6-8: 9 2,4,6,7: 1 4,6,7: 7 4,9,10: 1 9-11: 1	1,2,4,6-8: 34 2,4,6-8: 3 2,4,6,7: 0 4,6,7: 2	1,2,4,6-8: 25 2,4,6-8: 10 2,4,6,7: 1 4,6,7: 12 4,9,10: 3 9-11: 1 9,10,12: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted stimulants age at first use

D.5 Income Variables

Tables D.98 through D.103 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the income variables.

D.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 D.100 is used to summarize the many multiple variable missingness patterns.

Table D.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 D.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 D.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	176	1,2	1. WMS, and probabilities associated with other missing elements	1-7: 37 1,2,4-7: 35 2,4-8: 1	1-7: 33 1,2,4-7: 44 2,4-8: 3	1-7: 2 1,2,4-7: 16 2,4-8: 4	1-7: 0 1,2,4-7: 0 2,4-8: 0 2,4-7: 1
2	Missing welfare months, not receiving welfare payments, missing welfare services	96	1,2	1. SVC*WMS, SVC, and probabilities associated with other missing elements	1-7: 17 1,2,4-7: 28	1-7: 15 1,2,4-7: 22	1-7: 2 1,2,4-7: 10	1-7: 1 1,2,4-7: 1
3	Missing welfare months, missing welfare payments, not receiving welfare services	216	1,2	1. PMT*WMS, PMT, and probabilities associated with other missing elements	1-7: 51 1,2,4-7: 50	1-7: 21 1,2,4-7: 59	1-7: 5 1,2,4-7: 26 2,4-8: 1	1-7: 1 1,2,4-7: 2
4	Missing welfare months, missing welfare payments, missing welfare services	318	1,2	1. [1-(1-PMT)(1-SVC)]*WMS, PMT, SVC, and probabilities associated with other missing elements	1-7: 32 1,2,4-7: 138	1-7: 6 1,2,4-7: 84	1-7: 0 1,2,4-7: 50	1-7: 0 1,2,4-7: 7 2,4-8: 1

Table D.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]	3,431	1	1. Probabilities associated with missing elements	1-7: 997 1,2,5-7: 82	1-7: 1,341 1,2,5-7: 171	1-7: 466 1,2,5-7: 170 2,5-8: 6	1-7: 110 1,2,5-7: 84 2,5-8: 4

[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.

D.5.2 Specific Category Phase

Table D.101 Logical Constraints for Finer Categories of Income

Constraint #	Logical Constraint
LogC1	IRFAMSKP of donor = IRFAMSKP of recipient
LogC2	Personal (PINC2) and family (FINC2) income of donor must be consistent with personal and family income of recipient

Table D.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, if nonmissing
LikC3	FINC2 of donor = FINC2 of recipient, if nonmissing
LikC4	IRPINC1 of donor = IRPINC1 of recipient
LikC5	IRFINC1 of donor = IRFINC1 of recipient

Table D.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	Completely missing	7,234	1,2	1. Xβ	1-5: 2,250 2-5: 3	1-5: 2,940 2-5: 2	1-5: 1,562 2-5: 1	1-5: 466 2-5: 10

[1] For more details, see Section 8.3.4.

D.6 Health Insurance Variables

Tables D.104 through D.109 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the health insurance variables. Tables D.104 through D.107 are for variables created using the "Constituent Variables Method," and Tables D.108 and D.109 are for variables created using the "Old Method." See Chapter 9 for an explanation of the two methods.

In several instances, variable names are used without description for the purposes of brevity (see Chapter 9 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).

D.6.1 Constituent Variables Method

For the MPMN, 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 D.104 Likeness Constraints for Health Insurance, Constituent Variables Method (MPMN)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means for all missing constituent 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 constituent 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 D.105 Constraints and Portion of the Predictive Mean Vector for Health Insurance, Constituent Variables Method (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	263	None	1. HI1	1,2,7: 117 1,2: 3	1,2,7: 105 1,2: 3	1,2,7: 25	1,2,7: 10
2	Only missing MEDICARE	68	None	1. HI2	1,3,7: 29	1,3,7,8: 33	1,3,7,8: 6	No Cases
3	Missing CAIDCHIP and MEDICARE	41	None	1. HI1 2. HI2	1,2,3,7: 13 1-3: 2 1: 1	1,2,7,8: 16 1,2,8: 3 1,8: 0 Dummy: 1	1,2,7,8: 3 1,2,8: 1	1,2,7: 0 1,2: 0 1: 0 Dummy: 1
4	Only missing CHAMPUS	66	None	1. HI3	1,4,7: 32	1,4,7: 22	1,4,7: 6	1,4,7: 6
5	Missing CAIDCHIP and CHAMPUS	19	None	1. HI1 2. HI3	1,2,4,7: 10 1,2,4: 2 1,4: 1 4: 1	1,2,4,7: 4 1,2,4: 0 1,4: 0 4: 1	No Cases	No Cases
6	Missing MEDICARE and CHAMPUS	4	None	1. HI2 2. HI3	1,3,4,7: 2 1,3,4: 0 1,4: 0 4: 1	1,4,7,8: 1	No Cases	No Cases
7	Missing CAIDCHIP, MEDICARE, and CHAMPUS	15	None	1. HI1 2. HI2 3. HI3	1-4,7: 6 1-4: 0 1,4: 3 4: 2	1,2,4,7,8: 0 1,2,4,8: 1 1,4,8: 0 4,8: 1	1,2,4,7,8: 0 1,2,4,8: 0 1,4,8: 0 4,8: 2	No Cases
8	Only missing PRVHLTIN	195	None	1. HI4	1,5,7: 101 1,5: 0 1,6: 2	1,5,7: 81 1,5: 0 1,6: 0 1: 1	1,5,7: 6	1,6,7: 4
9	Missing CAIDCHIP and PRVHLTIN	76	None	1. HI1 2. HI4	1,2,5,7: 57 1,2,5: 0 1,6: 1 1: 1 Dummy: 1	1,2,5,7: 12 1,2,5: 0 1,6: 2 1: 0 Dummy: 1	No Cases	1,2,6,7: 0 1,2,6: 0 1: 0 None: 1
10	Missing MEDICARE and PRVHLTIN	7	None	1. HI2 2. HI4	1,3,5,7: 2 1,3,5: 0 1,6: 1 1: 1	1,5,7,8: 2	1,5,7,8: 1	No Cases

Table D.105 Constraints and Portion of the Predictive Mean Vector for Health Insurance, Constituent Variables Method (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	21	None	1. HI1 2. HI2 3. HI4	1-3,5,7: 4 1-3,5: 0 1,6: 0 1: 0 Dummy: 6	1,2,5,7,8: 3 1,2,5,8: 1 1,6,8: 0 1,8: 0 Dummy: 5	1,2,5,7,8: 1	1,2,6,7: 0 1,2,6: 0 1: 0 None: 1
12	Missing CHAMPUS and PRVHLTIN	23	None	1. HI3 2. HI4	1,4,5,7: 11 1,4,5: 0 1,4,6: 0 1,4: 0 4: 3	1,4,5,7: 3 1,4,5: 0 1,4,6: 0 1,4: 0 4: 2	1,4,5,7: 1 1,4,5: 0 1,4,6: 1 1,4: 0 4: 1	1,4,6,7: 1 1,4,6: 0 1,4: 0 4:0
13	Missing CAIDCHIP, CHAMPUS, and PRVHLTIN	28	None	1. HI1 2. HI3 3. HI4	1,2,4,5,7: 8 1,2,4,5: 0 1,4,6: 0 1,4: 1 4: 16	1,2,4,5,7: 0 1,2,4,5: 0 1,4,6: 0 1,4: 0 4: 2	1,2,4,5,7: 0 1,2,4,5: 0 1,4,6: 0 1,4: 0 4: 1	No Cases
14	Missing MEDICARE, CHAMPUS, and PRVHLTIN	5	None	1. HI2 2. HI3 3. HI4	1,3-5,7: 1 1,3-5: 0 1,4,6: 0 1,4: 0 1: 3	1,4,5,7,8: 0 1,4,5,8: 0 1,4,6,8: 0 1,4,8: 0 4,8: 1	No Cases	1,4,6,7: 0 1,4,6: 0 1,4: 0 4:0
15	Missing CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN	82	None	1. HI1 2. HI2 3. HI3 4. HI4	1-5: 14 1,4,6: 0 1,4: 0 4: 28	1,2,4,5,8: 8 1,4,6,8: 0 1,4,8: 0 4,8: 12	1,2,4,5,8: 3 1,4,6,8: 0 1,4,8: 0 4,8: 16	1,2,4,6: 0 1,4: 0 4: 1

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 D.106 Constraints for Health Insurance, Constituent Variables Method (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 D.107 Constraints and Portion of the Predictive Mean Vector for Health Insurance, Constituent Variables Method (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	183	None	1. HI1	1: 55	1: 108	1,2: 19 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)

D.6.2 Old Method

Table D.108 Likeness Constraints for Health Insurance, Old Method

Constraint #	Likeness Constraint
LikC1	Age of donor = Age of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Age of donor must be within 5 years of age of recipient
LikC4	Donor must not have received private health insurance (PINSUR = 0)
LikC5	Donor must not have received overall health insurance by the 1999 definition (INSUR = 0)
LikC6	Donor must have received overall health insurance by the 2001 definition (INSUR3 = 1)
LikC7	Donor must have received overall health insurance by the 1999 definition (INSUR = 1)

Table D.109 Constraints and Portion of the Predictive Mean Vector for Health Insurance, Old Method

#	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 INSUR3, no for INSUR, and no for PINSUR	62	None	1. (OVR*(1-PRV))/ (1-OVR*PRV)	1,2,4,5: 31	1,2,4,5: 28	1,2,4,5: 3	No Cases
2	Yes for INSUR3, missing INSUR, and no for PINSUR	35	None	1. (OVR*(1-PRV))/ (1-OVR*PRV)	1,2,4,6: 28	1,2,4,6: 6	1,2,4,6: 1	No Cases
3	Missing INSUR3, missing INSUR, and no for PINSUR	121	None	1. (OVR*(1-PRV))/ (1-OVR*PRV)	1,2,4: 45 1,4: 1	1,2,4: 61	1,2,4: 14	No Cases
4	Yes for INSUR3, missing INSUR, missing PINSUR	14	None	1. OVR 2. OVR*PRV	1,2,6: 12 1,6: 1	1,2,6: 1	No Cases	No Cases
5	Missing INSUR3, missing INSUR, missing PINSUR	346	None	1. OVR 2. OVR*PRV	1,2: 199 1: 1	1,2: 117 1: 1	1,2: 22 1: 4	1,2: 1 1: 1
6	Yes for INSUR3, yes for INSUR, and missing for PINSUR	77	None	1. PRV	1,2,7: 49	1,2,7: 17	1,2,7: 4 1,7: 1	1,2,7: 4 1,7: 2

[1] The predictive mean vector components are defined by the following:
1. OVR = P(Respondent received health insurance, 2001 definition)
2. PRV = P(Respondent received private health insurance | respondent received health insurance, 2001 definition)

D.7 Roster Pair Variables

Tables D.110 through D.129 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the pair variables. Tables D.110 through D.113 correspond to the first stage of imputation, where the pair relationship variable (IRPRREL) is created. Tables D.114 through D.121 are for the second stage, imputation of the multiplicity variables. Finally, Tables D.122 through D.129 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.

D.7.1 Stage One: Pair Relationship

Table D.110 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
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 D.111 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
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 D.112 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
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 D.113 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	12	None	1. SIB	1-3: 11 2: 1
2	One pair member is aged 12 to 14, the other is aged 18 to 25	13	15,20	1. SIB	1-3,10,11: 11 2,3,10,11: 1 2,3: 0 Dummy: 0 Dummy: 1
3	Both pair members are aged 15 to 17	7	10,16,21-25	1. SIB	1-5,62,101,121: 6 2-5,62,10,12: 1
4	One pair member is aged 15 to 17, the other is aged 18 to 25	22	10,16,20-25	1. SIB	1-6,10,13: 13 2-6,10,13: 8 2-6,10,12: 0 2-6: 1
5	One pair member is aged 18 to 20, the other is aged 18 to 25	48	10,17,21-24	1. SPOUSE1 2. SPOUSE2	1-5,10,13: 4 2-5,10,13: 44
6	Both pair members are aged 21 to 25	48	10,18,21-24	1. SPOUSE1 2. SPOUSE2	1-5,10,12: 17 2-5,10,12: 31
7	One pair member is aged 12 to 14, the other is aged 26 or older	3	15,21,25	1. PC	1-3,8,10,11: 3
8	One pair member is aged 15 to 17, the other is aged 26 or older	3	16,21-25	1. PC	1-5,61,8,10,14: 1 2-5,61,8,10,14: 2
9	One pair member is aged 18 to 20, the other is aged 26 or older	15	10,17,21-25	1. PC	1-5,8,10,14: 3 2-5,8,10,14: 9 2-5,8,10,11: 1 2-5,10,11: 1 Dummy: 1
10	One pair member is aged 21 or older, the other is aged 26 or older	67	10,18,21-25	1. SPOUSE1 2. SPOUSE2	1-5,7-10,12: 39 2-5,7-10,12: 28
11	Both pair members are aged 12 to 17	3	None	1. SIB	1-3: 3

[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)

D.7.2 Stage Two: Multiplicity Counts

Table D.114 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 D.115 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 D.116 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	105	None	1. PRED1	1-3,10,12,15,17,18: 49 2,3,10,12,15,17,18: 42 2,3,15,17,18: 6 2,16: 7 Dummy: 1

[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 D.117 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	0	2	1. PRED1	No Cases

[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.

Table D.118 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	0	3	1. PRED1	No Cases

[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 D.119 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	0	4	1. PRED1	No Cases

[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 D.120 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (15-17/18-25), Older Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	0	5	1. PRED1	No Cases

[1] PRED1 is the predicted mean for the number of siblings aged 15 to 17 for a respondent aged 18 to 25 who is a member of a sibling-sibling pair.

Table D.121 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (15-17/18-25), Younger Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	0	6	1. PRED1	1,3,10,12,14: 0 3,10,12,14: 0 3,14: 0 14:0

[1] PRED1 is the predicted mean for the number of siblings aged 18 to 25 for a respondent aged 15 to 17 who is a member of a sibling-sibling pair.

D.7.3 Stage Three: Household Counts

Table D.122 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 D.123 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 D.124 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
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

Table D.124 Likeness Constraints for Stage Three, Household Counts (continued)

Constraint #	Likeness Constraint
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
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 D.125 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	30	1,2	1. H1	1,3-5,8,39,41,43,44: 7 3-5,8,39,41,43,44: 3 5,8,39,41,43,44: 5	1,3-5,8,39,41,43,44: 7 3-5,8,39,41,43,44: 4 5,8,39,41,43,44: 4

[1] H1 is the predicted mean count of older siblings aged 15 to 17.

Table D.126 Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (15-17/18-25), 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	59	3,4	1. H2	1,3,4,6,10,40,42,44: 20 3,4,6,10,40,42,44: 6 6,10,40,42,44: 4 18,20,46: 2	1,3,4,6,10,40,42,44: 9 3,4,6,10,40,42,44: 10 6,10,40,42,44: 5 18,20,46: 3

[1] H2 is the predicted mean count of older siblings aged 18 to 25.

Table D.127 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	53	5	1. SPOUSE1 2. SPOUSE2	1-3,9,10,11,13-15,21,23,25,28,29: 7 2,3,9,10,11,13-15,21,23,25,28,29: 26 2,15,22,23,25,28,29: 10 2,22,24,26,29: 4	1-3,9,10,11,13-15,21,23,25,28,29: 3 2,3,9,10,11,13-15,21,23,25,28,29: 1 2,15,22,23,25,28,29: 0 2,22,24,26,29: 2

[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 D.128 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	26	6	1. SPOUSE3	1-3,9,10,11,13-15,21,23,25,28,29: 9 2,3,9,10,11,13-15,21,23,25,28,29: 8 2,15,22,23,25,28,29: 3 2,16,22,24,26,29: 3	1-3,9,10,11,13-15,21,23,25,28,29: 2 2,3,9,10,11,13-15,21,23,25,28,29: 1

[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 D.129 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	9	9,18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,32: 1 2-4,10,11,13,14,32: 6 2,4,10,11,13,14,32: 1 2,12,32: 1	No Cases
2	Missing PPCP1217	1	8,17	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,31: 1	No Cases
3	Missing PPCP1217 and PPCP1220	23	8,9,17,18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,31,32: 7 2-4,10,11,13,14,31,32: 14 2,4,10,11,13,14,31,32: 1 2,12,31,32: 1	No Cases
4	Missing PPCP1214	1	7,16	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,30: 1	No Cases
5	Missing PPCP1214 and PPCP1220	0	7,9,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
6	Missing PPCP1214 and PPCP1217	5	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: 5	No Cases
7	Missing PPCP1214, PPCP1217, and PPCP1220	41	7-9,16-18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,30-32: 4 2-4,10,11,13,14,30-32: 32 2,4,10,11,13,14,30-32: 2 2,12,30-32: 3	No Cases
8	Missing PPCC1220	19	12,15	1. PC1 2. PC2 3. H3	1-4,7,10,35: 1 2-4,7,10,35: 11 2,4,7,10,35: 3 2,7,38: 1	1-4,7,10,35: 0 2-4,7,10,35: 3
9	Missing PPCC1220 and PPCP1220	15	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: 9 2,4,7,10,11,13,14,32,35: 1 2,7,12,32,38: 1	1-4,7,10,11,13,14,32,35: 1 2-4,7,10,11,13,14,32,35: 2 2,4,7,10,11,13,14,32,35: 0 2,7,12,32,38: 1

Table D.129 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
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	0	8,15,17,18	1. PC1 2. PC2 3. H3	No Cases	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
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	1	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: 1	No Cases
16	Missing PPCC1217	1	11,14	1. PC1 2. PC2 3. H3	1-4,6,10,34: 0 2-4,6,10,34: 1	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

Table D.129 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
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
24	Missing PPCC1217 and PPCC1220	5	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: 3 2,4,6,7,10,34,35: 0 2,6,7,37,38: 1	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	0	12,14,15,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
27	Missing PPCC1217, PPCC1220, PPCP1217, and PPCP1220	19	14,15,17,18	1. PC1 2. PC2 3. H3	1-4,6,7,10,11,13,14,31,32,34,35: 1 2-4,6,7,10,11,13,14,31,32,34,35: 16 2,4,6,7,10,11,13,14,31,32,34,35: 0 2,6,7,12,31,32,37,38: 1	1-4,6,7,10,11,13,14,31,32,34,35: 1
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

Table D.129 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
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	0	7,14-18	1. PC1 2. PC2 3. H3	No Cases	No Cases
32	Missing PPCC1214	0	10,13	1. PC1 2. PC2 3. H3	No Cases	No Cases
33	Missing PPCC1214 and PPCP1220	0	9,10,13,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
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	3	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: 3	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	1	8,9,13,16-18	1. PC1 2. PC2 3. H3	1-5,10,11,13,14,30-33: 0 2-5,10,11,13,14,30-33: 1	No Cases

Table D.129 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
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
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

Table D.129 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
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
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	0	13,14,16,17	1. PC1 2. PC2 3. H3	No Cases	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	3	10-15	1. PC1 2. PC2 3. H3	1-7,10,33-35: 0 2-7,10,33-35: 1 2,4-7,10,33-35: 0 2,5-7,36-38: 2	No Cases
57	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1220	0	10,11,13-15, 18	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table D.129 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
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
60	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1214	0	11-16	1. PC1 2. PC2 3. H3	No Cases	No Cases
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	0	12-17	1. PC1 2. PC2 3. H3	No Cases	No Cases
63	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	27	13-18	1. PC1 2. PC2 3. H3	1-7,10,11,13,14,30-35: 2 2-7,10,11,13,14,30-35: 15 2,4-7,10,11,13,14,30-35: 0 2,5-7,12,30-32,36-38: 8 2,5-7,30-32: 1	1-7,10,11,13,14,30-35: 0 2-7,10,11,13,14,30-35: 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. 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 E: Quality Control Measures Used in the Imputation Procedures

Appendix E: Quality Control Measures Used in the Imputation Procedures

E.1 Introduction

For the 2011 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) weight adjustment for item nonresponse in the models, (2) predictive mean modeling, and (3) final assignment of imputed values using these predicted means. Specific checklists for the imputation of these variables were used during imputation processing for the 2011 survey and serve as formal documentation of the QC measures that were implemented.

In addition to the QC measures described in this appendix, SAS®[1] programs that were written and run by members of the imputation team were subsequently reviewed for errors by reviewers who examined messages in the SAS log file, assessed model convergence diagnostics, and identified any missing values. Imputation team members also reviewed and edited demographic variables (age, interview date, birth date, gender, race, and Hispanicity), household composition variables, and pair variables. QC measures were implemented throughout the editing process, and specific checklists were developed for the editing of demographic, roster, and pair variables. However, the QC measures that were used in the editing process are not discussed in this appendix, nor are the checks for delivering variables to other NSDUH teams or the QC checklists developed for nicotine dependence.[2]

Note that the drug use variables had an additional QC check for the random assignment of the date of first drug use. The specific checks involved in each of the PMN steps and the random assignment of age at first drug use are described in detail in the following sections.

E.2 Step 1: Weight Adjustment for Item Nonresponse

In the first step of the PMN methodology, a set of variables is defined to characterize item nonresponse. In the NSDUH, a "complete" respondent is classified as a person who responded to all the questions within a particular variable set; only complete respondents are used to build the models in Step 2. As a general practice, the weights are then adjusted so that the weights for complete respondents represent the entire domain, where "domain" is defined as the population of interest (e.g., lifetime users aged 12 to 17). This is accomplished by using an item response propensity model, a special case of the generalized exponential model (GEM),[3] which

---

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

[2] See the logical editing procedures used to create these variables in Chapters 3 and 7. For more details on other editing procedures that were performed on NSDUH data prior to imputation, see Kroutil, Handley, and Bradshaw (2013); Kroutil, Handley, Bradshaw, Chien, and Felts (2013); and Kroutil and Chien (2013).

[3] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures.

is described in detail in Appendix A in Westlake, Chen, and Gordek (2013). The following QC measures were conducted in Step 1:

- 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.
- Checks were performed on the output to see whether the GEM converged. If it did not, one or more variables were dropped. 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.
- The number of persons identified as item nonrespondents was recorded. This number was checked to ensure that it was the same as the number of persons who were excluded from the model-building process.
- 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.
- Any changes to existing programs were checked by those who ran the programs, as well as by other members of the imputation team.

E.3 Step 2: Predictive Mean Modeling

For each variable imputed using the PMN imputation method, modeling procedures were used to determine the predictive mean values for each respondent. For example, a model was used to determine the probability of lifetime usage of a given drug based on the responses to the corresponding gate question.[4] Although only item respondents contributed to the model, predictive mean values were determined regardless of whether the respondents answered the question or not. These predicted means were calculated based on Poisson regression models, failure time models, binomial and multinomial logistic models, or ordinary weighted least squares regression models with the response variable appropriately transformed. The models are discussed in detail in Chapters 3 through 10 of this report. The following QC measures were conducted in Step 2:

- Many of the independent variables were categorical variables and were subsequently converted into a set of indicator variables in an intermediate step. A list of a few observations on the dataset was printed to ensure that all of the indicator variables were created correctly.
- 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, convergence was ensured by checking the output to see whether convergence was obtained. For logistic models, the log file also was checked for "data warning" messages or other SUDAAN®-specific errors.[5] 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.
- Output was checked to verify that everything worked properly in the regression model.
- If there were two models in the drug frequency modeling programs, the convergence in both models was checked.
- For age at first use for the drug variable programs, the predicted age at first use was crossed with the respondent's age. The integer portion of the predicted age at first use could not exceed the respondent's age. Also, a subset of observations on the output dataset was carefully investigated to ensure that all of the predicted values and indicators were logical and consistent.
- A check on the predicted means from the model was created to ensure that each respondent in the domain had a valid predicted mean and was nonmissing.
- Any changes to existing programs were checked by those who ran the programs, as well as by other members of the imputation team.

[4] In the module for a given drug, the "gate question" was the first question that asked the respondent whether he or she had ever used the drug.

[5] Details can be found in the *SUDAAN® Language Manual, Release 10.0* (RTI International, 2008).

E.4 Step 3: Final Assignment of Imputed Values

E.4.1 Common Imputation Checks for PMN

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. The following QC measures were performed on both univariate and multivariate imputations:

- 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 had a neighborhood size with a donor whose predicted mean(s) was within 1 percent of the recipient's predicted mean(s) was noted.
- The number of cases that were imputed within various levels of restrictiveness of the likeness constraints (as determined by the variable SMALLFLG) was noted.[6]
- 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 changes to existing programs were checked by those who ran the programs, as well as by other members of the imputation team.

[6] Refer to Appendix D for more details about likeness restrictions and the "SMALLFLG" variable.

E.4.2 Specific Imputation Checks for UPMN

The values imputed in the imputation process by the univariate predictive mean neighborhood (UPMN) method were provisional when a multivariate predictive mean neighborhood (MPMN) method was required in the end. Otherwise, these values were final. The final univariate imputation included the following sets of variables: the Hispanic/Latino origin indicator, immigrant variables, age at first drug use, finer income variables, household composition (roster) variables, the indicator for any other health insurance, and most of the pair variables. The UPMN method used in lifetime usage of various drugs, recency and frequency of use of various drugs, and binary income variables was provisional. For these univariate imputations, the imputed values were checked for consistency against preexisting nonmissing values. Listed below are a few QC measures that were implemented to ensure consistency.

- The imputation-revised age at first use was crossed with respondent's current age to ensure that the age at first use was never greater than the respondent's age.
- If there were one or more child[7] drugs, the imputed variables of the parent drug were crossed with those of the child drug(s) to ensure consistency between them.
- For parent/child drugs, the parent drug age at first use must be less than or equal to the child drug age at first use.
- The respondent's age at first drug use must be less than the respondent's age, if the recency was "not in the past year."
- The imputed number of persons in the household younger than 18 should be within a lower and upper bound based on the value of imputed household size and the nonmissing ages in the roster.
- In binary income variable imputations, donors and recipients were required to have the same value for the variable IRFAMSKP, which indicated whether the respondent had family members in the household.
- The finer income category was checked to ensure consistency with the binary income category.
- For the immigrant age-of-entry variable, the donor's age of entry was checked to ensure that it was less than the recipient's current age.
- The edited variables were crossed with imputed variables to ensure that the imputations were conducted correctly. For example, the edited number of persons in the household aged 65 or older (HH65) was compared with the imputed number of persons in the household aged 65 or older (IRHH65) to ensure that IRHH65 had no missing values.

[7] A parent/child drug relationship occurred in modules that included subgate questions of substances that were of interest in their own right. For example, in the hallucinogens module, there was interest in the usage of LSD, PCP, and Ecstasy, which were all considered "child" drugs of the "parent" drug hallucinogen.

E.4.3 Specific Imputation Checks for MPMN

Multivariate imputations were performed on the following sets of variables: some of the demographic variables (with multinomial cells), binary income variables, health insurance variables (both the "Old Method" and the "Constituent Variables Method"), lifetime drug use, recency and frequency of drug use, and a few of the pair variables. For these multivariate imputations, the following items were checked:

- Any missing values were noted. This occurred when the program was unsuccessful in assigning a valid imputed value, such as drug recency (1, 2, 3, 4, 9), 30-day frequency (1–31, 91, 93), or 12-month frequency (1–365, 991, 993).
- Any cases where the imputed value was not consistent with preexisting nonmissing values were noted. These were cases where one or more variables were imputed, and one or more of these variables violated one or more of the following conditions:
  - The 12-month frequency must equal or exceed the 30-day frequency.
  - Past month users must have a valid 30-day frequency (not a skip code).
  - Past year users must have a valid 12-month frequency (not a skip code).
  - For alcohol, 30-day frequency must exceed or equal the "binge" drinking frequency.
  - For parent/child drugs (e.g., cocaine/crack, smokeless tobacco/snuff), the parent drug recency must occur no later than the child drug recency.
  - For cocaine and crack, the cocaine 12-month frequency must equal or exceed the crack 12-month frequency (if it existed).
  - For cocaine and crack, the cocaine 30-day frequency must equal or exceed the crack 30-day frequency (if it existed).
  - The recency and frequency-of-use variables that were imputed must be consistent with the time period between the birthday and interview date, as well as the time period between the interview date and the month that the respondent began using drugs, if that variable was available. For example, if the respondent was not a past month user, the imputed 12-month frequency of use could not exceed the maximum usage period less 30.
  - If the respondent's age was equal to the age at first use, the recency of use must be imputed to be "past month" or "past year, not past month."
  - For some drugs, the respondents were asked both the 12-month frequency and the 30-day frequency questions. For past month users, the 30-day frequency must be at least the 12-month frequency less 335, and no greater than the 12-month frequency.
  - If the edited age at first use was equal to the current age of the respondent, the imputed recency must be consistent with the time period between the birthday and the interview date, and it must be consistent with the month that the respondent began using, if available.

- For income, only persons who answered "yes" to either the welfare payments or other welfare services source-of-income questions had valid answers concerning months on welfare.
- For health insurance, respondents who indicated that they had health insurance but were missing the private health insurance indicator required donors who had some health insurance.

- The distribution of the imputed values was compared with the distribution of nonimputed values. Unusual patterns in these distributions were investigated. For example, this included the distribution of lifetime users versus nonlifetime users, the distributions of recency and frequency of use, and the age-at-first-use distributions for drugs. For income, this included the distributions of family income variables.
- 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 these respondents had either full-time or part-time status assigned to the employment status variable (EMPSTAT4), but not unemployed or other statuses.
- Each pattern of missingness was treated separately. 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 be included among the imputed values, not just past month users.
- For the recency and frequency of use, provisional imputed values were used in the process before a final vector of predicted means was created. The provisional imputed recencies were crossed with the edited and final imputed recencies by the imputation indicator. This check was established to identify whether something went wrong in the final multivariate hot-deck step.

E.5 Additional Step for Drug Variables: Assignment of the Date of First Drug Use

For the imputations of age at first drug use, an additional step was required that assigned a date of first use. The QC measures in this step had two requirements:

- The assigned date of first use must be consistent with the given birth date and the imputation-revised age at first use.
- The assigned date of first use must be consistent with the given interview date and the imputation-revised recency and frequency-of-use variables.

Respondents failing either of the two preceding checks were carefully examined. Occasionally, the error was unavoidable (e.g., when the age at first use, recency of use, and interview date were inconsistent by only 1 day), even after editing. In particular, this could occur if the birthday or interview date occurred on the 1st day of the month. It was important to ensure that all inconsistencies that appeared were of this type:

- The imputation-revised year and month of first use were crossed with the edited year and month of first use to ensure that all valid edited years and months were being transmitted to the imputation-revised year and month of first use.
- A frequency of the imputation-revised month/day/year of first use was run to ensure that all were within the acceptable numbers (i.e., month was between 1 and 12, or month was 99 for "never used").
- If there were one or more child drugs, the imputed variables of the parent drug were crossed with those of the child drug(s) to ensure consistency between them.

Sometimes, because an error was discovered further along in the process, a patch was necessary for earlier imputations. When variables were reimputed and the dataset was updated, it was crucial to compare the old (incorrect) imputation-revised variable and the new (correct) variable with the reimputed values. This was necessary to ensure that (1) the changes made were within expected limits, and (2) other cases did not inadvertently change with the correction. Cases with unanticipated changes were investigated individually. In addition, all imputation-revised variables and imputation indicators were checked to ensure that each variable label was correct and the length of the variable was acceptable. For all of the programs, any changes to existing programs were checked by those who ran the programs, as well as by other members of the imputation team.

E.6 Imputation Checklists

Most of the QC measures presented above were incorporated into specific imputation checklists for demographic, drug use, income, health insurance, nicotine dependence, roster, and pair variables. These checklists included a technician check, where the individual who ran the computer program (technician) entered his or her name and the date the check was performed. 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. In addition, for many of the checklist entries, an independent reviewer performed an additional check of the same items. This reviewer also entered his or her name and the date the check was performed. This reviewer check ensured greater quality in the imputation procedures. These checklists provided formal documentation of the QC measures that were incorporated during imputation processing. Checklists also were updated and revised to reflect the changes in the programs before each processing cycle. New checks were added to the existing checklists to ensure additional quality and to improve the process.

Checklists were developed and utilized for many imputation programs. Almost all major imputation programs and, as a result, all variable categories were covered. The specific checklists that were implemented for the 2011 NSDUH imputation programs are summarized in Table E.1.

Table E.1 Summaries of Checklists Used for Imputation Programs in the 2011 NSDUH

Program Category	Subtasks: Demographics	Drug	Income	Health Insurance	Roster	Pair	Nicotine Dependence
Editing	✓	N/A	N/A	N/A	✓	✓	N/A
Item Nonresponse Weight Adjustment	✓	✓	✓	✓	✓	✓	Regression
Predictive Mean Modeling	✓	✓	✓	✓	✓	✓	
Univariate Predictive Mean Neighborhood	✓	✓	✓	✓	✓	✓	
Multivariate Predictive Mean Neighborhood	✓	✓	✓	✓	N/A	✓	
Date of First Drug Use	N/A	✓	N/A	N/A	N/A	N/A	N/A
Delivering Variable	✓	✓	✓	✓	✓	✓	✓

NOTE: The "✓" symbol implies that a specific checklist was developed for this subtask and program category. A specific editing checklist was developed for the regression imputation method of nicotine dependence variables. The N/A abbreviation indicates that a checklist was not applicable for this subtask or program category.

Appendix F: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

Appendix F: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

F.1 Introduction

In the household roster for the 2011 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 7.

In general, two types of consistency checks were implemented in the 2011 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.

F.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 F.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 3.

In the 2002 survey, the explanations provided in Table F.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 3.

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 F.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.

F.3 Override Comments from Interviewers: Internal Consistency Check Overrides

Internal consistency checks were performed on the household roster for the 2011 NSDUH. Interviewer explanations for overrides to these internal consistency checks are provided in Table F.2. These explanations were evaluated individually to determine their legitimacy. 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, an evaluation of whether the override was considered legitimate. If the override was considered legitimate, no edit was applied to the age or relationship code of the roster member. If the override was not considered legitimate, the override was overruled, and the relationship code (and sometimes the roster member's age) was set to bad data. In this instance, a brief indication of the probable true relationship of the roster member to the respondent is provided in the "Comment" column of the table.

[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 F.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	14	13	14	response from respondent	14	Diff. ≤ 1 year	14
2	24	25	25	SR is 25 yrs old	24	Diff. ≤ 1 year	24
3	70	71	72	HAY DISCREPANCIA EN SUS RESPUESTAS	70	Diff. ≤ 1 year	70
4	19	18	18	not an error	18	***AGE was changed to equal roster age***	18
5	23	22	22	23	23	Diff. ≤ 1 year	23
6	19	20	19	earlier recording of birth date may have been in error	19	Diff. ≤ 1 year	19
7	22	23	23	dad gave wrong age	22	Diff. ≤ 1 year	22
8	31	32	32	husband made mistake on wifes age	31	Diff. ≤ 1 year	31
9	14	13	14	14 yr old continued to give me this informaion	Not in a pair	Diff. ≤ 1 year	14
10	25	24	24	25	Not in a pair	Diff. ≤ 1 year	25
11	37	36	36	computor error	37	Diff. ≤ 1 year	37
12	14	15	15	13	14	Diff. ≤ 1 year	14
13	19	18	19	CONFIRMED FROM R THAT SHE IS 18	19	Diff. ≤ 1 year	19
14	19	18	18	R is 18 y old	Not in a pair	***AGE was changed to equal roster age***	18
15	54	53	53	R gave wrong year of birth. should be 1958	54	Diff. ≤ 1 year	54
16	73	72	72	He confirmed his age is 72	Not in a pair	Diff. ≤ 1 year	73
17	46	47	46	46	46	Diff. ≤ 1 year	46
18	24	25	25	mix up on age by screening respondent	24	Diff. ≤ 1 year	24
19	32	33	33	33	Not in a pair	Diff. ≤ 1 year	32
20	43	44	42	R got her age mixed up	43	Diff. ≤ 1 year	43
21	18	17	18	emotionally respondent aswered	18	Diff. ≤ 1 year	18
22	35	34	34	male is 34	35	Diff. ≤ 1 year	35
23	17	16	17	r is 16	17	Diff. ≤ 1 year	17
24	22	23	23	person doing screening was incorrect	22	Diff. ≤ 1 year	22
25	41	40	40	sr is 40	41	Diff. ≤ 1 year	41

Table F.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
26	31	30	31	this is the information provided by the respondent	31	Diff. ≤ 1 year	31
27	23	22	22	r is23 dad said 22 - 23 is the correct	23	Diff. ≤ 1 year	23
28	41	40	40	should be 40 yrs old	Not in a pair	Diff. ≤ 1 year	41
29	34	35	35	R age is recorded as 35 on the iPAQ and on the laptop	Not in a pair	Diff. ≤ 1 year	34
30	23	22	23	she was 22 on the last birthday	Not in a pair	Diff. ≤ 1 year	23
31	20	19	19	As of the last birthday, this person is 19.	20	Diff. ≤ 1 year	20
32	15	16	16	age entered as told	15	Diff. ≤ 1 year	15
33	14	15	15	mother gave wrong age at screening	14	Diff. ≤ 1 year	14
34	14	13	13	person 5 is the interviewee	14	Diff. ≤ 1 year	14
35	23	22	23	IR confused on earlier age question	Not in a pair	Diff. ≤ 1 year	23
36	73	72	72	Today is R birthday	Not in a pair	Diff. ≤ 1 year	73
37	13	12	13	response for respondent	13	Diff. ≤ 1 year	13
38	18	19	19	r is 19	18	Diff. ≤ 1 year	18
39	24	23	23	birthday may 09 2011	24	Diff. ≤ 1 year	24
40	46	45	46	resp made a mistake earlier and said she was 46-but she is 45	46	Diff. ≤ 1 year	46
41	13	12	13	gave his age in error, he is 13	Not in a pair	Diff. ≤ 1 year	13
42	16	15	16	she is 15-adults are guardians	16	Diff. ≤ 1 year	16
43	68	67	67	R states she is still 67 will be 68 in october.	68	Diff. ≤ 1 year	68
44	35	36	35	respondent is actually 35 he has mental disabilities and does not know his age, had to do the math to determine correct age.	35	Diff. ≤ 1 year	35
45	22	23	23	22	Not in a pair	Diff. ≤ 1 year	22
46	22	21	21	17	22	Diff. ≤ 1 year	22
47	13	12	13	response from 13 yr old	13	Diff. ≤ 1 year	13
48	41	40	40	40 YO is wife, 41 YO is husband	41	Diff. ≤ 1 year	41
49	25	24	24	housemate gave incorrect information	Not in a pair	Diff. ≤ 1 year	25

Table F.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
50	13	12	12	will be 13 in May	Not in a pair	Diff. ≤ 1 year	13
51	23	22	22	father was screener	23	Diff. ≤ 1 year	23
52	22	25	25	because the r was born 09-23-1988 THAT WOULD MAKE HIM 25	22	Diff. > 1 year, other pair member supports NEWAGE	22
53	72	76	76	72	Not in a pair	Diff. > 1 year, FI does not support change leave as NEWAGE	72
54	29	24	24	age is correct	24	***AGE was changed to equal roster age***	24
55	32	24	24	The respondent is 24 and when asked at the beginning this was the age recorded as well. I'm not understanding when the age was recorded different?	24	***AGE was changed to equal roster age***	24
56	23	3	24	23	23	Diff. > 1 year, other pair member supports NEWAGE	23
57	31	994	32	31 years old	Not in a pair	Roster Age Missing, FI does not support change leave as NEWAGE	31

[1] These entries came directly from the 2011 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 F.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	Grandparent and respondent less than 30 years apart	15	31-year-old grandparent	not biological	Overrule; unsure of relationship (85)
2	Grandparent and respondent less than 30 years apart	12	24-year-old grandparent	the mother was very young when she had a baby	Overrule; unsure of relationship (85)
3	Grandchild and respondent less than 30 years apart	24	12-year-old grandchild	the mother had the child at a young age	Overrule; unsure of relationship (85)
4	Multiple spouses or partners	17	Age unknown unmarried partner	ANSWER GIVEN BY R	Overrule; unsure of relationship (85)
5	Multiple spouses or partners	17	Age unknown unmarried partner	ANSWER GIVEN BY R	Overrule; unsure of relationship (85)
6	Multiple spouses or partners	19	39-year-old unmarried partner	boyfriend	Overrule; unsure of relationship (85)
7	Multiple spouses or partners	19	37-year-old unmarried partner	boyfriend	Overrule; unsure of relationship (85)
8	Multiple spouses or partners	19	20-year-old spouse	boyfriend	Overrule; unsure of relationship (85)
9	Respondent is 16 or younger and has a spouse or partner	15	19-year-old unmarried partner	THIS IS HOW RESPONDENT considers self	Legitimate; interviewer's override stands
10	Respondent is 16 or younger and has a spouse or partner	14	42-year-old unmarried partner	the compentered 14 as cohab/I entered the 42 M as mother's co hab	Overrule; probable parent
11	Respondent is 16 or younger and has a spouse or partner	15	34-year-old unmarried partner	person 2 is the live-in partner to the house holder, the 15 year old is person #4	Overrule; probable parent
12	Respondent's daughter is older than respondent	14	14-year-old daughter	there are 2 14 yr olds 1/2 sister	Overrule; probable sibling

Table F.2 Explanations for Overrides to Internal Consistency Checks (continued)

#	Consistency Check	AGE	Roster Member's Age and Relationship to Respondent	Verbatim Explanation from Field Interviewers[1]	Comment
13	Respondent's daugther-in-law is older than respondent	28	37-year-old daughter-in-law	sister in law	Overrule; probable other nonrelative
14	Respondent's son is less than 13 years younger than respondent	15	8-year-old son	what respondent says	Overrule; unsure of relationship (85)

[1] These entries came directly from the 2011 NSDUH field interviewers. Any typographical errors or misspellings were transcribed directly and were not corrected.

Appendix G: Rules for Determining Pair Relationships

Appendix G: Rules for Determining Pair Relationships

G.1 Rules for Determining Matching Pairs, in Priority Order

The following rules are used to determine the roster member in 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 are 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 priority conditions because of their hierarchical nature, are listed in priority order in Table G.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. Some other conditions that were not evident in 2011 also are excluded from this table, provided the distribution of counts would have been unaffected by their exclusion from the code.

Table G.1 Rules for Determining Matching Pairs, in Priority Order

Priority Condition	Rule		
	Pair Member A	Pair Member B	Count
1	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, exactly one MBRSEL in right place	16,689
2	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	2,108
3	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	132
4	Age and gender match exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	245
5	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	28
6	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	3

Table G.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority Condition	Rule		
	Pair Member A	**Pair Member B**	**Count**
7	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	305
8	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	23
9	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	5
10	Age and gender match exactly, MBRSEL missing for all roster members	Age and gender match exactly, MBRSEL missing for all roster members	15
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	2
12	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	23
13	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	4
14	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	30
15	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	6
16	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender matches exactly, MBRSEL missing for all roster members	4
17	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	168
18	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	19

Table G.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority Condition	Rule		
	Pair Member A	Pair Member B	Count
19	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	4
20	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	3
21	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	2
22	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	7
21	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	2
24	Age and gender match exactly, exactly one MBRSEL in right place	Everything missing	17
25	Age within one, gender matches exactly, exactly one MBRSEL in right place	Everything missing	4
26	Age matches exactly, gender off, exactly one MBRSEL in right place	Everything missing	1
22	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	33
23	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	2

Table G.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority Condition	Rule		
	Pair Member A	Pair Member B	Count
24	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
30	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	1
31	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 selected	1
25	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	1
26	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	2
27	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	20
28	Age and gender match exactly, MBRSEL missing for all roster members	Age matches exactly, gender off, MBRSEL missing for all roster members	1
29	Age and gender match exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	48
30	Age within one, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	6
38	Age within two, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	3
39	Age and gender match exactly, MBRSEL missing for all roster members	No match at all (often paired respondent is missing from roster)	2
40	No match at all	No match at all	5

G.2 Rules for Identifying Pair Relationships among Pairs

Table G.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 10.3 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 10.6 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 G.2 Rules for Identifying Pair Relationships among Pairs

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
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				7	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				8	3
	Roommate/nonrelative	Spouse/partner	≥ 1 child both sides, equal number each side					

Table G.2 Rules for Identifying Pair Relationships among Pairs (continued)

	A-B Relationship							
Priority Condition	**B to A, according to A**	**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				10	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 G.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				7	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 G.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				11	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 G.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 G.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 G.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				5	3
	Sibling	Spouse						
55	Spouse	Sibling	One is 12-17, other is 18-25, both sides have parents or spouses				6	3
	Sibling	Spouse						

Table G.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	12-14			19	4
	Sibling	Nonmissing, not child	12-14	15-17				
59	Nonmissing, not parent	Sibling	12-14	15-17			19	4
	Sibling	Nonmissing, not parent	15-17	12-14				
60	Nonmissing, not child	Sibling	18-25	12-17			20	4
	Sibling	Nonmissing, not child	12-17	18-25				
61	Nonmissing, not parent	Sibling	12-17	18-25			20	4
	Sibling	Nonmissing, not parent	18-25	12-17				
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 G.2 Rules for Identifying Pair Relationships among Pairs (continued)

	A-B Relationship							
Priority Condition	B to A, according to A	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 H: Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members

Appendix H: Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members

H.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 both rosters did not have 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 10.2.2, it was technically possible for a respondent to have multiple spouses, but these situations were not accounted for.

H.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, we determined that this was a child-parent 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 child-parent 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.

H.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 I: Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members

Appendix I: Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members

I.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.

I.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. Among nonparent-child pairs of interest, in most cases, the counts of children in the relevant age range with parent(s) in the household (abbreviated below as children

[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.

with parent(s) 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:

- 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 had been resolved at the pair relationship stage where we would had 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. Among nonparent-child pairs of interest, 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 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

[6] For this condition, either the count for the other pair member was 0 or the count for the pair members was equal.

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).

I.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. Among pairs that were not sibling-sibling pairs of interest, 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.

I.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:

[8] In rare cases, an individual might identify two spouses in the household. As noted in Section 10.2.2, 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.

- The pair relationship was not a spouse-spouse pair, and the total household size was at least four; and
- 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.

I.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, we already excluded households without couples. We also excluded households with a possibility of two or more couples. 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 J: Priority Conditions for Creating Household-Consistent Covariates

Appendix J: Priority Conditions for Creating Household-Consistent Covariates

J.1 Household Size

In Table J.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 J.1 shows the hierarchical priority condition, with the frequency of occurrence for each priority condition in parentheses.

Table J.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 (18,823)	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 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 (9)	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 J.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 (13)	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, TOTPEOP (A) 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 (0)	Both missing or 1	Both missing or 1			SHHSIZE ≥ 2, SHHSIZE at least as large or larger than screener roster	SHHSIZE
16 (26)	Not equal, both > 1	TOTPEOP(B) = QD54 (B)	A: At least one age range variable less than min.[1] B: Age range variables all same or larger than min.			QD54B
		TOTPEOPA = QD54 (A)	B: At least one age range variable less than min. A: Age range variables all same or larger than min.			QD54A
17 (3)	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 J.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 (951)	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 (2)	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
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

Table J.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:
23 (0)	Not equal, both > 1	Not equal, both > 1	TGOODAGA > GOODAGEA, TGOODAGB > GOODAGEB, TGOODAGA = TGOODAGB	A: no B: no		TGOODAGA
24 (7)	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 (142)	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 J.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 (0)	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.

J.2 Age Variables

Table J.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 set of tables, a variable followed by "A" (either in parentheses or not) 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 10.3.1, 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 AGE011A, 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 10.2.1.2.1. 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 J.1, the first column in Table J.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 J.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 (4)[3]		AGE1217A = AGE1217B, both ≥ min.	Another count except 12-17 < min.			AGE1217A
5 (1)[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 (0)[6]					A & B: none missing A has better match measure than B	AGE1217A

Table J.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)	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)						Missing
13 (17,942)		AGE1217A = AGE1217B	All other counts equal across pair members			AGE1217A
14 (585)		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 (49)					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 (84)					A & B: none missing Age (A) ≥ Age (B)	AGE1217A
					A & B: none missing Age (B) > Age (A)	AGE1217B

Table J.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 (0)	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 J.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)[7]		AGE1217B < min. AGE1217A = min.				AGE1217A
31 (0)		AGE1217A < min. AGE1217B < min.		SAGE1217 ≥ min.		SAGE1217
32 (0)[8]		AGE1217A = AGE1217B, both ≥ min.				AGE1217A
33 (0)		AGE1217A not equal to AGE1217B	AGE1825A < min. AGE1825B ≥ min.			AGE1217B
34 (1)[9]			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 J.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)[10]	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 (644)		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 J.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 (451)		Priority conditions 43-55 not met				AGE1217B

Table J.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 (175)	TGOODAGA = HHSIZE					TAG1217A
	TGOODAGB = HHSIZE					TAG1217B
58 (35)[11]	SHHSIZE = HHSIZE	AGE1217A, AGE1217B $\leq$ SAGE1217		AGE1217A & B $\leq$ SAGE1217		SAGE1217
59 (0)[12]	SHHSIZE = HHSIZE, HHSIZE = EXC1217	AGE1217 missing	Other counts not missing, AGE011A equals AGE011B			MIN1217
60 (2)[13]	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 J.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)[14]	Previous priority conditions for HHSIZE, TOTPEOP, GOODAGE, not met, SHHSIZE = HHSIZE	AGE1217 missing	At least 3 of the other counts missing			SAGE1217
99 (0)	All prior conditions were not met.					Missing

[1] The following frequencies were observed for priority condition #1: AGE1825 and AGE2634 = 1; and AGE011 = 5.

[2] The following frequencies were observed for priority condition #2: AGE2634 = 1; and AGE3549 = 2.

[3] The following frequencies were observed for priority condition #4: AGE011 = 0; AGE3549 = 1; AGE2634 = 2; and AGE1825 = 3.

[4] The frequency of priority condition #5 for AGE011, AGE1825, AGE2634, and AGE50p was 0.

[5] The frequency of priority condition #6 for AGE3549 and AGE50p was 1.

[6] The frequency of priority condition #9 for AGE1825 and AGE2634 was 1.

[7] The frequency of priority condition #30 for AGE1825 was 1.

[8] The frequency of priority condition #32 for AGE2634 and AGE50p was 1.

[9] The frequency of priority condition #34 for AGE011, AGE1825, AGE2634, AGE3549, and AGE50p was 1.

[10] The frequency of priority condition #39 for AGE3549 was 1.

[11] The following frequencies were observed for priority condition #58: AGE011 = 21; AGE2634 and AGE15p = 30; AGE1835 = 31; AGE1220 = 32; and AGE3549 = 34.

[12] The following frequencies were observed for priority condition #59: AGE1825 and AGE3549 = 1; AGE2634 = 4; and AGE011 = 6.

[13] The following frequencies were observed for priority condition #60: AGE011 and AGE3549 = 0; and AGE1825, AGE2634, and AGE15p = 1.

[14] The frequency of priority condition #61 for AGE15p was 0.

Appendix K: Creation of Household-Level and Person-Level Files

Appendix K: Creation of Household-Level and Person-Level Files

K.1 Introduction

For the 2011 administration of the National Survey on Drug Use and Health (NSDUH), a person was randomly selected for an interview through a four-stage sample selection process. States were first stratified into a total of 900 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. Within sampled census tracts, adjacent census blocks were combined to form the second-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 dwelling units (DUs) within the sample segment boundaries. These lists served as the frames for the third stage of sample selection. After the DUs were selected within each segment, an interviewer visited each selected DU to obtain a roster of all persons aged 12 or older. This roster information was then used to select zero, one, or two persons from the household at the fourth 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 K.3 for the definition of a completed case.

K.2 Dwelling Unit-Level Eligibility and Completeness Criteria

Before proceeding with the fourth 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 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.

[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 2011 NSDUH sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2012) contains more information regarding the sample design.

Therefore, DU information from the U.S. Census Bureau for these areas was used in the household-level file.

During the second 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.

K.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 K.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 K.1. The number of eligible, selected, and interviewed persons also is summarized in the table.

Table K.1 NSDUH Household, Person Eligibility, and Completed Interview Counts: 2011

	Selected Dwelling Units	Eligible Dwelling Units	Completed Screenings (Dwelling Units)	Eligible Persons	Selected Persons	Inter-viewed Persons	Completed Cases (Interviews)
CAI	216,521	179,293	156,048	328,395	88,536	70,170	70,109

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 recorded in the binary variable PRCOMP, which had a value of 1 if the respondent was a

[2] For more details on editing rules regarding the drug usage gate questions, refer to the 2011 NSDUH editing and coding reports (Kroutil, Handley, Bradshaw, Chien, & Felts, 2013; Kroutil, Handley, & Bradshaw, 2013).

completed case, and 0 if not. For a summary of the number of completed cases in the 2011 survey, see Table K.1.

K.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 persons 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 persons 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 persons completed the interviews from the five age groups within a household.

As described in the 2011 NSDUH sample design report (Morton et al., 2012), 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-digit segment identifier, which uniquely identified each 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 variable PDEN2 classified 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 the modified 2000 census data, which was adjusted to more recent data from Claritas, Inc.[3] This variable had five levels: segment in core-based statistical area (CBSA)[4] with 1 million or more persons; segment in CBSA with 250,000 to 999,999 persons; segment in CBSA with fewer than 250,000 persons; segment in urban area but not in CBSA; and segment in rural area (not in CBSA and not in urban area).

The variable PLACNAME was the census place name associated with each segment. According to the census documentation, this variable was defined as places, for the reporting of decennial census data, which includes census-designated places, consolidated cities, and incorporated places. If duplicate place names existed within the same county, the places were

[3] Claritas, Inc., is a market research firm headquartered in San Diego, California.

[4] 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.

distinguished by their legal description (e.g., "city" or "village"). However, because the variable PLACNAME was used to help the field interviewers locate the segment and was limited by the number of characters printed on the map, identifiers like "city" or "village" have been removed from the place name. 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 RURORURB is derived from the 2000 census block-level designation of rural or urban.[5] In the NSDUH sample, if 100 percent of the blocks are rural, the segment is defined as rural (RURORURB = 1). If one or more of the blocks within a segment is urban, the segment is defined as urban (RURORURB = 2). Other sampling variables such as DIVISION, SSREGION, GQTYPE, ID, STNAME, STUSAB, and QUARTER[6] also were included in the household-level and person-level files.

[5] 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 persons per square mile and surrounding census blocks that have an overall density of at least 500 persons per square mile. In addition, under certain conditions, less densely settled territory may be part of each UA or UC.

[6] For more details on these sampling variables, refer to the 2011 NSDUH sample design report (Morton et al., 2012).